The Dissertation Committee for Kristina Kichler Holder

Certifies that this is the approved version of the following dissertation:

Dynamics of Adaptive Evolution in Two Experimental Viral Systems

Committee:

James J. Bull, Supervisor

David M. Hillis, Co-Supervisor

Mark Kirkpatrick

Ian J. Molineux

Holly A. Wichman

Dynamics of Adaptive Evolution in Two Experimental Viral Systems

by

Kristina Kichler Holder, B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2001

ACKNOWLEDGMENTS

There are many people who have made my experience as a student at the University of Texas so enjoyable and worthwhile. Among them, I would like to specifically thank my committee members for all of their help in guiding the work that is described in this dissertation. Particularly I would like to thank Jim Bull who, in spite of hardly saying a word to me during my first two years as a student here, turned out to be incredibly generous with his time and willingness to talk with students about their projects and is genuinely interested in and appreciative of a broad range of science. I would also like to thank the members of the Hillis, Bull and Cannatella labs, past and present, that I have had the opportunity to know. I can't imagine a group of people that I would have enjoyed being around more. Leaving such good friends makes the move to Connecticut very bittersweet. I would like to thank my family for being so supportive and encouraging during these many years that I've been in grad school. Most importantly, I would like to thank Mark. Only he knows how much he has done for me, and I can't begin to express how much it has meant.

Dynamics of Adaptive Evolution in Two Experimental Viral Systems

Publication No. _____

Kristina Kichler Holder, Ph.D.
The University of Texas at Austin, 2001

Supervisor: James J. Bull
Co-supervisor: David M. Hillis

Two species of coliphage were adapted to a novel thermal environment (44°) by serial passaging in a homogenous environment at a low multiplicity of infection. Genomewide nucleotide changes were identified by whole genome sequencing. Both phage achieved tremendous gains in fitness in the new environment during the course of the experiments with little or no loss of fitness in the ancestral environment. Although the number of mutations underlying the fitness gains varied greatly between the two systems, early large-effect mutations were common to both phage. Five missense mutations were observed during the adaptation of ϕX174; the first two substitutions were responsible for 85% of the total fitness gain. Adaptation to a less extreme environment (41.5°) was necessary before G4 could be adapted to 44°. Seventeen mutations were observed: 14 missense, two silent, and one intergenic. The first three missense substitutions accounted for over half the ultimate fitness increase.

The order in which mutations occurred was determined by probing archived isolates with radiolabeled oligonucleotides. This revealed the surprising result of extensive polymorphism in both phage systems. While there was evidence of epistasis, many of the mutations were beneficial in a variety of genetic backgrounds. Although many adaptive pathways were available, other factors, such as mutational rate heterogeneity, limited the number of pathways that were likely to be followed. This study provides the most complete descriptions to date of the adaptive evolution of a microbial system.

TABLE OF CONTENTS

1. INTRODUCTION

A species evolves to become better adapted to its environment through a process of selection in which genotypes in the population, arising via mutation or existing as standing variation, differ in fitness. The next generation will have a higher mean fitness than the previous generation. In a closed environment, the species will continue to evolve higher fitness until reaching some optimum. How the species arrives at that optimum and what that final fitness will be depend on a number of factors, such as the spectrum of mutations, the mutation rate, the order in which mutations arise and their epistatic interactions. These factors were the subject of theoretical studies in the early half of the twentieth century, and their relative importance has been revisited by theorists in the past few years in light of data generated from studies of experimental evolution and the history of adaptation in natural populations.

Theoretical Treatments of Adaptive Evolution

One of the classic models is adaptation by periodic selection. Asexually reproducing populations under weak selection are expected to increase fitness by the periodic sweep to fixation of newly arising beneficial mutations (Fisher 1930; Muller 1932; Atwood *et al.* 1951). Variability in the population slowly increases, primarily as the result of the accumulation of neutral mutations, until a beneficial mutation occurs, eliminating that variability as it displaces other genotypes. Under some

conditions, such as a low beneficial mutation rate or small population size, this model may be quite applicable. However a mutant's sweep to fixation will be hindered if another beneficial mutant arises (Muller 1932). When the beneficial mutation rate is high, a model of periodic sweeps may not adequately capture the important characteristics of adaptation.

Pathogens often evolve under strong selection to develop resistance to drug therapies or evade detection by the immune system. In these situations, the distribution of mutational effects may include a high frequency of beneficial mutants. Gerrish and Lenski (1998) have modeled the effects of clonal interference (a variation of the effect of linkage on selection noted by Hill and Robertson [1966]) on adaptive evolution in asexual populations. Because of the expectation of multiple beneficial mutations in competition in the population at any given time during adaptation, they predict high frequency polymorphisms and suggest that the rate of evolution of viral populations may be close to a limit imposed by clonal interference rather than by a lack of beneficial mutants for selection to act on (a conclusion disputed by Orr [2000a]). They emphasize the importance of large-effect mutations in that larger selection coefficients will be necessary for mutations to fix as the number of competitor mutations increases with population size. Studies of fitness profiles have been consistent with hypothesized clonal interference in experimental populations of *E. coli* (de Visser *et al.* 1999) and an RNA virus (Miralles *et al.* 1999; Miralles *et al.* 2000).

Fisher's (1930) view of adaptation emphasized a single optimal phenotypic state achieved through the accumulation of numerous mutations of very small effect. He proposed a geometric model in which a complex organism can be thought of as being composed of traits arranged along n-orthogonal axes; a population is a point on the surface of an n-dimensional hypersphere where the origin represents the optimal combination of trait values for a particular environment. With this model he could demonstrate mathematically that the probability of a mutation being beneficial rapidly decreases as the size of the phenotypic effect increases relative to the distance to the optimum. Biologically this can be thought of as the cost imposed by the inherent tradeoffs in optimizing interacting traits, such that only mutations with very small phenotypic effects have a realistic chance of having a net beneficial effect on fitness.

Orr (1998) revisited Fisher's geometric model to generate a distribution for the phenotypic effects of mutations fixed during the "adaptive walk", the sum of the substitutions that occur as a species (specifically a haploid species with a large population size) moves from some starting point to the optimum. The results of his simulations gave an exponential distribution, independent of both the fitness function and the distribution of mutational effects at each step in the walk. By considering the expected substitutions for an entire walk, the probability of fixing a mutation of large effect was greatly increased. Even for complex organisms, the maximum rate of adaptive evolution is obtained with mutations of intermediate size (Orr 2000b),

mutations that have the highest probability of both being beneficial and escaping stochastic loss (Kimura 1983), not the very small mutations proposed by Fisher.

Wright (1932, 1988) focused on the possibility of multiple solutions, of varying quality, to the challenge of adapting to an environment. His model places more emphasis on chance events, such as the first mutation to fix during adaptation, and epistasis, selection coefficients of mutations being dependent on the background in which they arise. He suggested the image of an adaptive landscape with multiple selective peaks. Without becoming explicitly mathematical, this image captures the dominant features of adaptation when a species, or population, becomes divided into subpopulations (Korona *et al*. 1994).

These different theoretical treatments of adaptive evolution are generally not mutually exclusive. They place greater or lesser importance on different parameters and may be most appropriate under certain conditions or at certain times during the process of adaptation. However, it is difficult to determine which model or combination of models is most appropriate for describing or predicting adaptation of actual species. In order to refine existing models, the range of parameter values that are of biological interest need to be determined. The factors that are most important during the stage of adaptation relevant for answering a particular question, e.g. the expected time for a virus to evolve resistance to a new therapy, need to be identified. This will not be easy, and the answers derived for some species will surely not be applicable to all species or all situations. Microbial systems offer the possibility of

examining adaptive evolution in detail. In addition to being tractable organisms for laboratory experimentation, microbes are of inherent interest because of their role as pathogens.

Experimental Studies of Adaptive Evolution

This is hardly the first experimental study of the adaptation of a microbial system. There have been extensive studies of bacterial populations (Korona *et al.* 1994; Korona 1996; Helling *et al.* 1986; Rosenzweig *et al.* 1994; Bennett *et al.* 1992; Mongold *et al.* 1999) including a long-term study of *E. coli* (Lenski *et al.* 1991; Lenski and Travisano 1994), yeast (Adams and Oeller 1986), and both RNA(Burch and Chao 1999; Escarmis *et al.* 1999; Elena *et al.* 1997; Miralles *et al.* 2000) and DNA viruses (Wichman *et al.* 1999; Crill *et al.* 2000).

Directly assaying for the genetic changes in bacterial and yeast systems is not possible and can even be difficult for organisms with small genomes such as RNA viruses because of their high mutation rates (Burch and Chao 1999). Step function analyses of fitness trajectories have been used to infer underlying mutations (Lenski *et al.* 1991; Elena *et al.* 1996; Burch and Chao 1999) on the basis that the selective sweeps of periodic selection are predicted to cause step-like increases in mean fitness (Lenski *et al.* 1991). Diversity at sites responsible for fitness increases has been estimated for a subset of mutations (Notley-McRobb and Ferenci 2000; Riehle *et al.* 2001) and indirectly estimated by analyzing genomes for restriction fragment length polymorphisms (Adams and Oeller 1986; Papadopoulos *et al.* 1999) or by testing

correlated phenotypic traits (Korona 1996). The level of polymorphism revealed by these studies indicates that adaptive evolution and the experiments designed to study it are not as simple as previously assumed. Inferring changes, even if the model on which those inferences are based is correct, still leaves out important details of the adaptive process.

Wichman *et al*. (1999) were able to identify genetic changes by whole genome sequencing and monitor the frequency of those mutations in the populations over time. In this way they were able to show that replicate populations adapted via many of the same mutations, although the order in which mutations fixed varied. They relied on the dynamics of the frequency changes to infer the adaptive value of the mutations.

In this study I adapted two species of bacteriophage to a novel thermal environment. I was able to assay directly for genomewide changes by DNA sequencing. Using techniques developed by Wichman *et al*. (1999), I observed frequency changes during adaptation. Additionally, by selecting the phage in an environment that could easily be replicated in fitness assays, I was able to determine the fitness of individual genotypes both in the selected environment and in alternative environments. This allowed me to provide a much more detailed picture of adaptation than previous studies. By identifying potential causes for trends that had been observed in other experiments, I was able to further test particular aspects of

evolution by reviving populations for additional selection and by using site-directed mutagenesis to examine the effects of mutations in different backgrounds.

Bacteriophage ϕX174 and G4

The viruses used in this study are two isometric coliphage, ϕX174 and G4. Both phages have small single-stranded DNA genomes (5386 bases in ϕX174, sequenced by Sanger *et al*. [1977]; 5577 bases in G4, sequenced by Godson *et al*. [1978]). Although approximately 33% divergent at the DNA sequence level, the genome organization of these two phages is the same; both code for the same set of eleven proteins (Godson *et al*. 1978). In multiply-infected cells, recombination mediated by host proteins can occur (Tessman and Tessman 1959). In addition to differences in host range, G4 is capable of growing at a lower temperature than ϕX174 (Godson 1974).

The infection process for these phage is known in detail (summarized from Hayashi *et al*. 1988). The single-stranded phage genome and a phage protein, gpH, enter the host cell after the virion binds to a lipopolysaccharide receptor on the cell surface. Host proteins replicate the phage DNA (stage I replication), producing a double-stranded replicative form (RF) of the genome. More RF genomes are made with host proteins and the phage protein gpA (stage II replication). Host replication is shut off by gpA*. Most of the RF genomes function solely in transcription. Synthesis of single-stranded progeny phage genomes (stage III replication) is concurrent with the packaging of those genomes into capsids. A capsid precursor, the

procapsid, is made from gpB, gpD, gpF, gpG and gpH. The switch to stage III replication is controlled by gpC. DNA for progeny phage is generated by rolling-circle replication, again requiring gpA in addition to a number of host proteins. As the new strand of DNA is formed, it is packaged into the procapsid with gpJ while gpB is removed. A stable, mature virion is produced with the removal of gpD. The mature capsid consists of 60 gpF (major capsid protein), 60 gpG (major spike protein), 12 gpH (minor spike protein) and 60 gpJ (internal protein). Once progeny phage are formed, gpE lyses the cell. Only the function of gpK is unknown.

The ease of propagating large phage populations and their rapid response to selection allowed me to carry out experimental lines for hundreds of phage generations, generating detailed profiles of adaptation. Storing whole lysate and individual plaque isolates allowed me to test the fitness of different populations and genotypes under a variety of conditions, assay mutation frequencies, sequence interesting isolates, revive lines for further selection and generate new lines from individual genotypes. I was able to verify the transient nature of the polymorphisms I observed and explore the short term effect of that variability on the response to selection in a new environment. By sequencing whole genomes, I identified the mutations conferring increased fitness and tested the fitness effects of some of those mutations in an alternative genetic background. Knowledge of the infection process made it possible to model population dynamics in the experimental passages in simulations that tested realistic mutation rates, determining a range that was

compatible with the empirical results, and the efficiency of selection in the experiments.

2. PROFILES OF THE ADAPTIVE EVOLUTION OF TWO VIRAL SPECIES

INTRODUCTION

While previous studies of experimental evolution have provided fitness profiles for a variety of microbial systems (e.g. Lenski *et al.* 1991; Korona *et al.* 1994; Burch and Chao 1999; Adams *et al.* 1992), the goals for the experiments described in this chapter were to detail the underlying genetic changes and their effects on fitness and to begin to identify the factors that are most relevant in determining the outcome of adaptation.

Populations of ϕX174 and G4, each started from a single ancestral genotype, were propagated by serial transfer at an elevated temperature to select for maximal phage growth rate. Fitness was assayed under the selection conditions, and genetic changes were mapped to the fitness changes. Assays for changes in genotype frequencies revealed unexpected levels of polymorphism while results of fitness assays at alternative temperatures indicated that even quite large increases in fitness in the selected environment could be achieved without a significant fitness cost in other environments.

MATERIALS AND METHODS

Phage and host: Single isolates of ϕX174 and G4 were used as the ancestors for these studies. The sequence of the ancestor ϕX174 isolate used in this study differs at

five positions from the published sequence (Sanger *et al*. (1977), GenBank V01128 for the published sequence; Bull *et al*. (1997), GenBank AF176027 for our ancestor sequence).

Phage from a single G4 isolate, whose sequence differed from the published sequence of Godson *et al*. (1978; GenBank NC001420) at 20 positions (sequence submitted to GenBank), were propagated according to the G4 passaging protocol given below for 40 passages at 37°. A single isolate from the final passage was selected at random and used to begin the high temperature adaptation experiment. I will refer to this isolate as the G4 ancestor. During the 40 passages at 37°, it accumulated two missense substitutions (3740G, a K380E change in gene F, and 4530G, a T171A change in gene G). Both phage ϕX174 and G4 were propagated on the host *Escherichia coli* C.

Selective conditions: The goal was to select the phage in a constant environment that could be easily replicated in the fitness assays. This was done to ensure that fitness assays accurately reflected the fitness of the phage in the selective conditions. Passages and fitness assays were conducted in water baths that held temperature to within ±0.1°. Each passage was begun with a naïve host to prevent co-evolution of the bacteria with the phage. Phage were grown in an excess of bacteria, keeping the multiplicity of infection (MOI) below one throughout each passage. This reduces intracellular competition which can lead to genotypes whose fitness is frequency-dependent (e.g., Turner and Chao 1999).

ϕX174 passages: ϕX174 was serially propagated at 44° for a total of 50 passages. Passages were carried out in 1 L flasks containing 100 mL LB supplemented with 2 mM $CaCl_2$. Aliquots of *E. coli*, from stocks stored at –80°, were thawed and grown for 1.5 to 2 hours in the flasks at the selection temperature with shaking until reaching a density of 2-5 x 10^8 /mL. Approximately 10^7 phage were added to the cells and allowed to grow until the population had increased to 10^{10}, a 1000-fold increase in population size. Aliquots were collected and treated with chloroform to kill remaining bacteria. The aliquot was used to initiate the subsequent passage; the remainder was stored in an equal volume of borate EDTA with14% DMSO at –80°.

Because of the low initial fitness of our ancestral ϕX174 at 44°, the initial passage was allowed to proceed for two hours. Because the number of phage declined during the first passage, 9 mL of lysate were needed to initiate the second passage. After the second passage, the fitness was sufficiently high for the population to increase 1000-fold in just over an hour, and the above passaging protocol was followed.

G4 passages: Passages were carried out in a manner similar to those of ϕX174 with the following differences. Smaller flasks (125 mL with 10 mL of LB) were used for convenience. Host cells were grown for one hour at the selection temperature to a density of 2-5 x 10^8 /mL. Passages were initiated with 10^6 phage, and a 10,000-fold increase in population size was permitted before the passage was terminated. When the fitness of the population at the selected temperature was too low for such an

increase in population size to occur in 1.5 hours, the initial population size was increased up to ~10^8 phage.

Even after adapting G4 to 37°, the fitness at 44° was too low to maintain the phage during serial passage. Therefore, there were an initial 50 passages of selection at 41.5°. A single representative plaque isolate from the population evolved at 41.5° was selected and propagated for an additional 80 passages at 44°.

Sequencing: Population consensus genotypes were obtained by sequencing PCR product amplified from lysate. Sequence of lysate containing multiple genotypes gives the most common base in the population at each polymorphic site. It is not possible to reliably distinguish underlying peaks that are the result of polymorphism from those that are simply the result of background noise. All sequences were obtained by PCR amplification of the whole genome in two overlapping segments followed by chain-termination sequencing reactions. For both ϕX174 and G4, overlapping fragments covering the entire genome were created by sequencing with twelve primers, all hybridizing to the same strand. The reactions were analyzed on an automated sequencer (ABI377). Sequences were aligned and analyzed in SeqMan II (DNASTAR, Inc. 1997-1998).

For ϕX174, a consensus sequence was obtained from the 50^{th} passage at 44°. Five individual isolates from that time point were also sequenced to check for polymorphisms. For G4, sequence data were obtained for the 41.5° isolate that served as the ancestor for the 44° passages and for the 10^{th}, 30^{th}, 50^{th} and 80^{th} passages

at 44°. During the selection experiment, a single isolate was sequenced from the 50^{th} passage to monitor molecular changes, and five isolates were sequenced from the 80^{th} passage to check for polymorphisms. Consensus sequences were used for the 10^{th} and 30^{th} passages in order to correlate molecular and fitness changes at those time points.

Fitness assays: Fitness assays were carried out at 37° and 44° for both ϕX174 and G4; additional assays were conducted at 41.5° for G4. Assays were repeated four times for each temperature and phage combination. Fitness was calculated as the $\log_2$ increase in phage concentration per hour, as described in Bull *et al*. (2000). Assay conditions mimicked passage conditions. Cells were grown in LB for one hour to a density of ~2×10^8 /mL. At this point (t_0), 10^2-10^5 phage were added, and the mixture was grown for 40 minutes (t_{40}). Phage titers of the initial and end point cultures were determined by plating.

An outlier test was used to remove the value of one fitness assay, growth at 37° of G4 from the final passage selected at 44° (Snedecor and Cochoran 1980). Removal of this data point, thought to be caused by experimental error, did not qualitatively change the results. Results of the remaining fitness assays were compared using t-tests with Bonferroni corrections.

Oligo probing: Plaque isolates from selected passages were archived in microtiter plates for analysis by hybridization with radiolabeled oligonucleotides (Wichman *et al*. 1999; Crill *et al*. 2000). Probing provides a way to determine the order and frequency dynamics of the substitutions. Twelve isolates from every fourth passage of

the ϕX174 adaptation and eight isolates from every third passage of the G4 41.5° adaptation were screened for single nucleotide substitutions identified in consensus sequences. Additional isolates were selected for sequencing on the basis of the probing results.

RESULTS

ϕX174 fitness evolution: Figure 1 shows the fitness of phage populations at various times. From an initial fitness of approximately -2.8, indicating a decline in phage concentration, the population fitness increased to a final fitness of ~22.5 doublings/hour at 44°, a 4 x 10^7 increase in the number of descendants/hour. Most of this increase occurred rapidly after the second passage.

ϕX174 molecular evolution: Consensus sequence from the final passage revealed three missense mutations in two genes (Table 1). The 1727T mutation, responsible for the initial leap in fitness, had been observed in other high temperature adaptation studies (Bull *et al*. 2000); the other changes were not observed in previous experiments. Oligonucleotides were designed to probe for these changes.

The mutations 1727T and 1614A appeared to quickly sweep to fixation. Because only a small number of isolates were probed for each time point, "fixation" is used loosely here to indicate that all subsequent isolates probed or sequenced had the evolved nucleotide. Surprisingly, the 927T mutation, which was first detected at passage 20 and rapidly increased to high frequency, did not fix during the subsequent

Figure 1. ϕX174 population fitness at the selected temperature, 44°, average number of mutations per genome ("Mutations"), and polymorphism level ("Polymorphism"; $\Sigma p_i(1-p_i)$ over all sites assayed). Fitness in this and all other graphs is given with error bars indicating ± one standard error. The frequencies of individual mutations are shown below the graph, where white indicates the proportion of isolates with the ancestral state, black the proportion with the evolved state. Mutation frequencies are based on oligo hybridization results (twelve isolates probed for each time point) for the closest passage assayed (0, 4, 16, 32, 48). After being detected in sequence data, the changes at 1727, 1614, and 927 were assayed by oligo screening with probes specific to the wildtype and mutant sequences. The change at 2903 was assayed by sequencing (12 isolates at passage 4, 9 isolates at passage 8, 5 isolates at passage 50); it was present only in 5 of the passage-4 isolates, so its absence between passages 8 and 50 is inferred. The change at 926 was first detected in the sequence of an isolate at passage 44 that failed to hybridize to the 927T oligo and only weakly hybridized to the wildtype oligo. Of 16 additional isolates in passages 28-50 exhibiting a similar hybridization pattern, sequences of 5 isolates confirmed 926T in all. The presence of 926T in the other 11 is inferred from the hybridization pattern.

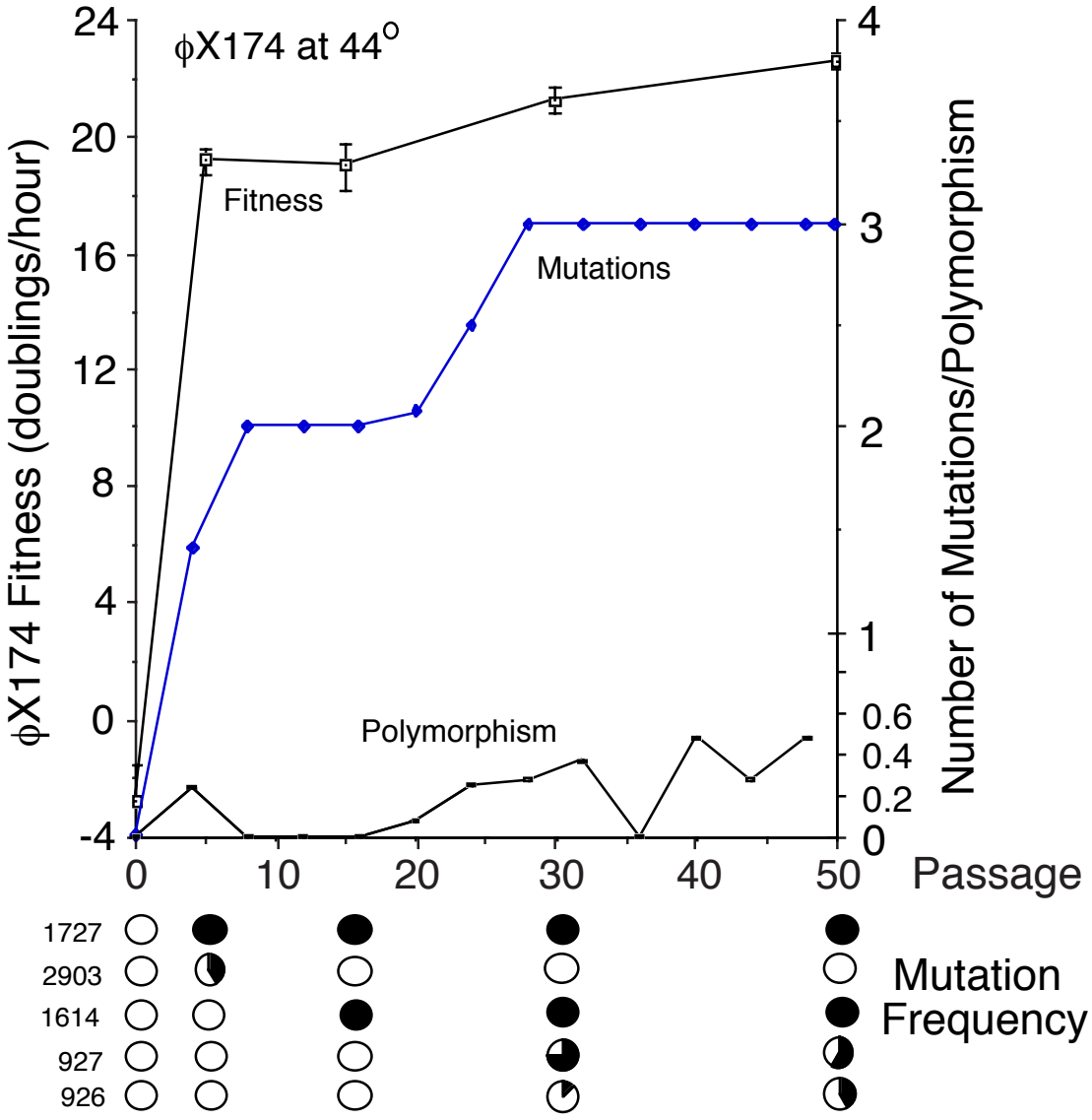

ϕX174 at 44°
Fitness
Mutations
Polymorphism
ϕX174 Fitness (doublings/hour)
Number of Mutations/Polymorphism
Passage
24
20
16
12
8
4
0
-4
4
3
2
1
0.6
0.4
0.2
0
0
10
20
30
40
50
1727
2903
1614
927
926
Mutation Frequency

Table 1. Nucleotide mutations observed during the high temperature adaptation of ϕX174. The passage at which the mutations were first detected by oligo probing of archived isolates is indicated.

Mutation	Gene and Codon	AA Change	1st Detected	Final Status
1727 C→T	F242	L→F	4	fixed
2903 G→T	G169	C→F	4	lost
1614 C→A	F204	T→N	8	fixed
927 G→T	J26	G→V	20	polymorphic
926 G→T	J26	G→C	28	polymorphic

passages.

Sequencing of isolates that lacked 927T revealed the mutation 926T. Both the 926 and 927 mutations are in codon 26 of gene J but encode different amino acids (cysteine and valine, respectively). The frequency of isolates with the 927T mutation remained around 0.75 during the final 22 passages. The persistence of polymorphism at these two sites could be actively maintained by frequency dependence, or it could be the result of the two genotypes being selectively neutral relative to each other, i.e. having equal and constant fitness in the experimental environment. Fitness assays failed to detect significant fitness differences between the two genotypes ($p>0.73$). Despite coding for very different amino acids, both changes appear to confer similar fitness advantages.

The estimated fitness for the population with only the substitution at 1727 was higher than a previous estimate for that genotype (Bull *et al*. 2000), suggesting the presence of another mutation. Sequencing revealed a transient polymorphism, 2903G/T. Although the 2903T mutation was found in five of 12 isolates from the fourth passage, it was not observed in any other sequenced isolates. Apparently this genotype was displaced when the 1614A mutation went to fixation.

G4 fitness evolution: G4 population fitness at 44° increased from approximately -11 doublings/hour to 13 doublings/hour, a 1.6×10^7 increase in the number of descendants/hour, but this increase was achieved by an initial adaptation to a less extreme environment of 41.5° prior to selection at 44°. Attempts to adapt G4 directly

to 44° were unsuccessful because the phage could not be maintained during serial transfer. The ancestral G4 fitness at 44° was much lower than that of ϕX174 (-11 vs. -2.5 doublings/hour); population size dropped rapidly during passaging. By the third passage, the population had dropped too low for continued propagation, and, in contrast to ϕX174, fitness had not noticeably increased, remaining far below the fitness level needed to offset the diluting effects of the passaging protocol. The highest temperature at which serial passaging was possible was 41.5°. After 50 passages at that temperature, G4 fitness at 44° had increased to ~3 doublings/hour; selection was then continued for an additional 80 passages at 44°.

G4 molecular evolution and polymorphism at 41.5°C: Three mutations (two missense and one silent) appeared to fix during selection of G4 at 41.5°; at least five additional missense mutations were polymorphic in the population when a single isolate was chosen to begin the selection at 44° (Table 2, Figure 2; 153T is not listed in the figure as it was identified from the sequence of a single isolate, and its frequency in the population was never determined with probing). That isolate differed from the 37°-adapted ancestor at five sites located in three genes (Table 2). Its genotype was not present among the eight isolates probed from the final 41.5° passage.

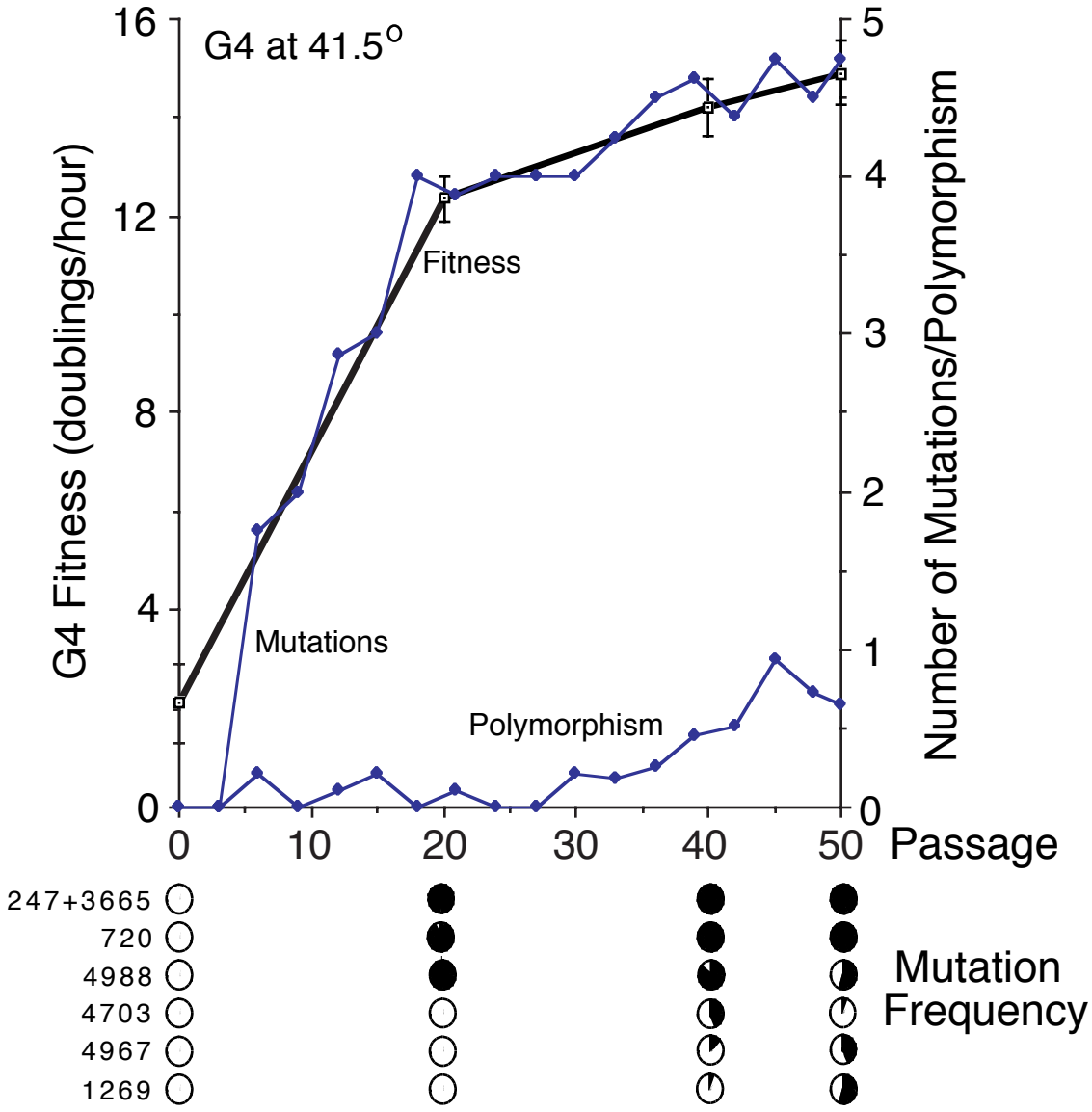
Mutations either fixed rapidly or remained polymorphic for many passages. The substitutions 247T and 3665T were first detected as polymorphisms at passage 6 and were present in all subsequent isolates assayed. The substitution 720T appeared

Table 2. G4 mutations observed during 50 passages of selection at 41.5°. The "Final Status" classification is based on probing and sequences of 3 isolates at passage 50, before initiating the 44° selection.

Mutation	Gene and Codon	AA Change	1st Detected	Final Status
247 C→T	A63	silent	6	fixed
3665 C→T	F355	P→S	6	fixed
4988 G→A	H142	G→D	12	polymorphic
[a]4967 G→A	H135	G→D	15	polymorphic
720 C→T	A221/A*8	T→I	18	fixed
4703 C→T	H47	A→V	33	polymorphic
[a]1269 A→G	A404/A*191	N→S	42	polymorphic
[a]153 C→T	A32	S→F	50	polymorphic

[a]Mutations not present in the isolate used to initiate the 44° selection.

Figure 2. G4 population fitness at 41.5° with substitution frequencies. White in a circle indicates the proportion of isolates with the ancestral state, black the proportion with the evolved state. Fitness, average number of substitutions ("Mutations") and polymorphism levels are shown across the entire 50 passages of adaptation at 41.5°. Mutation frequencies are based on oligo hybridization results (8-16 isolates probed; two time points were combined if they were within two of the indicated passage number). The change at 153, found in the sequence of a single isolate from passage 50, is not included because the other isolates have not been probed for the mutation. The "Polymorphism" curve is $\Sigma p_i(1-p_i)$ over all sites assayed (see legend to Figure 1). The "Mutations" curve highlights the average number of mutations per genome at each time point.


G4 at 41.5°
G4 Fitness (doublings/hour)
Number of Mutations/Polymorphism
Fitness
Mutations
Polymorphism
Passage
0
4
8
12
16
0
1
2
3
4
5
0
10
20
30
40
50
247+3665
720
4988
4703
4967
1269
Mutation
Frequency

to fix by passage 24 after initially being detected at passage 18. The mutation 4988A remained polymorphic. It was detected by passage 12 and was maintained at high frequency (present in 7-8 of 8 isolates assayed per time point) for at least 30 passages. The mutation 4967A remained polymorphic for at least 36 passages, 4703T for at least 18 passages, and 1269G for at least the last nine passages at 41.5°. What is particularly striking is that, from passage 39 through passage 50, at least three genotypes were detected in the eight isolates probed from each time point. This level of persistent polymorphism was unexpected in this system in which even small differences in fitness were expected to result in rapid fixation of a beneficial mutation.

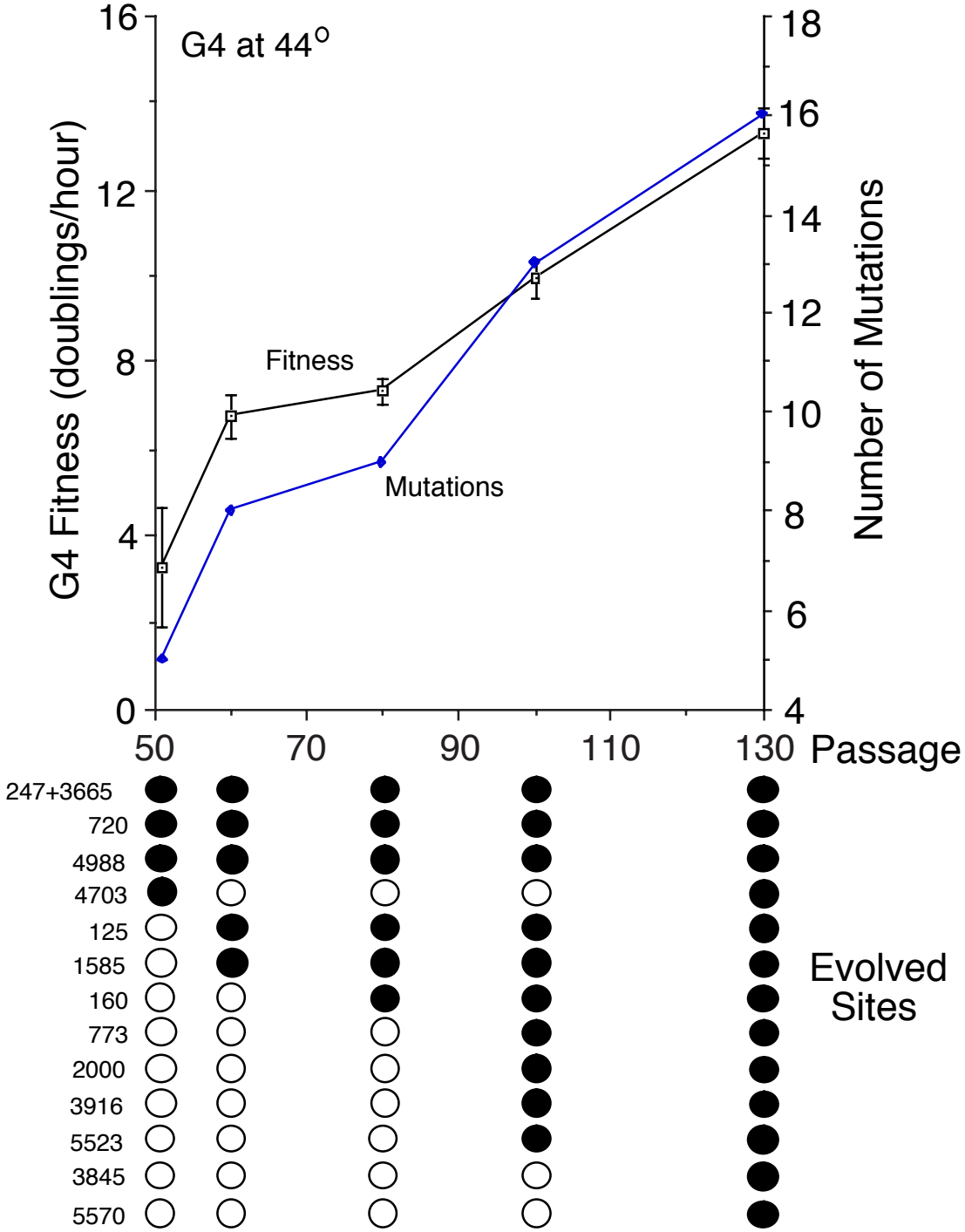
G4 molecular evolution at 44°: Figure 3 shows the increase in fitness from ~3 doublings/hour to 13 doublings/hour during 80 passages of selection at 44°. Nine nucleotide substitutions were observed in addition to the five mutations present in the initial population (Table 3). These nine substitutions were distributed among five genes and an intergenic region. Of the eight located in coding regions of the genome, seven resulted in amino acid substitutions. One (1585G) was a missense mutation only in gene B although it was located in a region of overlap for genes A, A* and B.

Site 4703, which was fixed for the evolved nucleotide at the beginning of the 44° selection, had reverted to the ancestral state by passage 10. The evolved state was not observed again in any sequences prior to the final passage. The silent substitution 160T was the only change observed between the sequences from the 60th and 80th

Table 3. G4 nucleotide substitutions observed after selection at 44° (passages 51-130), based on consensus and isolate sequences at passages 60, 80, 100, and 130. The mutation 1585G is in a region of overlap of genes B, A and A* but was missense only in B. The status "fixed" means that all 5 isolates sequenced at passage 130 exhibited the change.

Mutation	Gene and Codon	AA Change	1st Detected	Final Status
125 A→G	A23	T→A	60	fixed
1585 A→G	B104	T→A	60	fixed
160 C→T	A34	silent	80	fixed
773 G→A	A239/A*27	V→I	100	fixed
2000 G→T	D9	V→L	100	fixed
3916 A→G	intergenic	---	100	fixed
5523 C→A	H320	N→K	100	fixed
3845 C→T	F415	H→Y	130	fixed
5570 C→A	H336	S→Y	130	fixed

Figure 3. G4 population fitness at 41.5° with substitution frequencies. White in a circle indicates the proportion of isolates with the ancestral state, black the proportion with the evolved state. Fitness and average number of substitutions ("Mutations") are shown across the 80 passages of adaptation to 44° (passages 51-130). Presence or absence of a substitution was determined from the consensus sequence. Only five mutations seen in the population at 41.5° were present in the isolate used to initiate the 44° selection. The "Mutations" curve highlights the number of substitutions observed in the consensus sequence at each passage that was assayed.


G4 at 44°
G4 Fitness (doublings/hour)
Number of Mutations
Fitness
Mutations
Passage
0
4
8
12
16
4
6
8
10
12
14
16
18
50
70
90
110
130
247+3665
720
4988
4703
125
1585
160
773
2000
3916
5523
3845
5570
Evolved Sites

passages (the 10^{th} and 30^{th} passages at 44°), indicating that it fixed independently of the other changes.

Fitness Similarity of G4 Mutations: Mutations 4988A and 4967A together were able to displace the ancestor genotype that was wildtype for both sites, but they were unable to displace each other over the course of at least 35 passages. One possible explanation for the persistence of these polymorphisms is that the fitness effects of the two mutations were very similar. A likelihood analysis was performed to determine the range of differences in selection coefficients for genotypes with the 4967A or 4988A mutations that are compatible with the observed data.

The likelihood of observing the experimental data for a range of initial frequencies and selection coefficients (compared to the 4988A genotype) for 4967A genotype was calculated as:

$$\ln L_x = t_x \ln(f_x) + r_x \ln(1-f_x),$$

where f_x is the frequency of the 4967A genotype, t_x is the number of observations of 4967A and r_x is the number of observations of 4988A at sample point x; ln likelihood values for the sample points were summed. The change in frequency of 4967A per generation was calculated as:

$$f_x = \frac{f_{x-1}(1+s)}{(f_{x-1})s+1} .$$

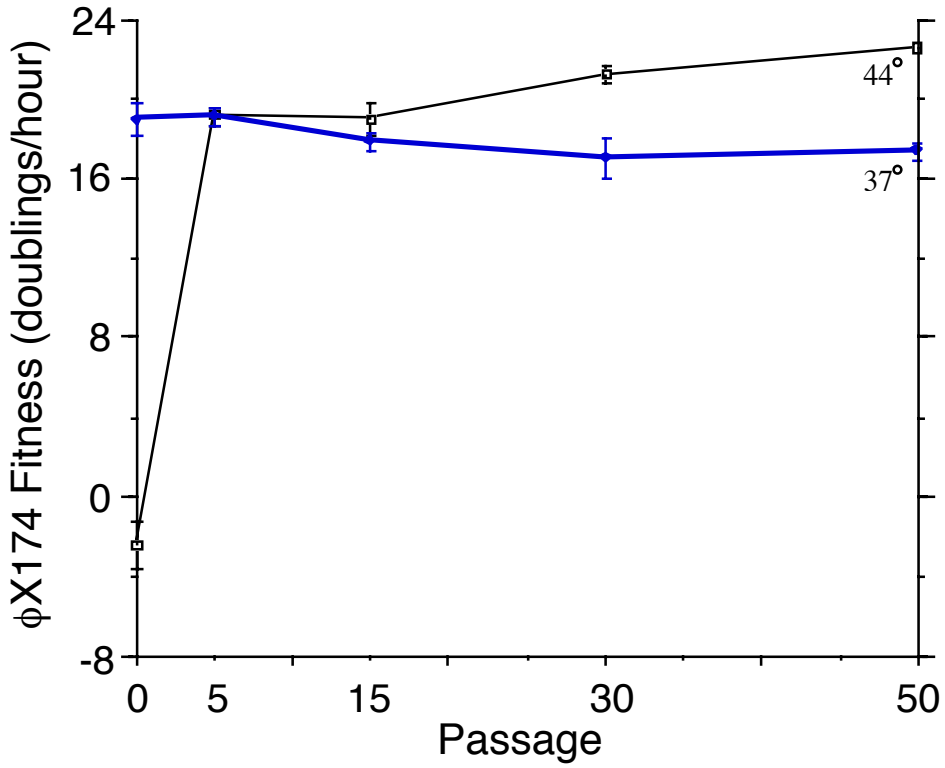
Based on data from burst experiments described in Chapter 3, I assumed that there were three generations per passage. Stochastic variation in the frequency of the genotypes was ignored because of the large population sizes used in the experiment.

Analysis was complicated by the fact that additional mutations accumulated after the initial appearance of these two mutations and was therefore limited to observations of 4967A and 4988A genotypes in samples from passages 15 through 39. Since the mutation 720T occurred in both backgrounds during this time (appearing to fix in the population by passage 24) without causing a detectable change in the frequencies of 4967A and 4988A, its effect on fitness was ignored (i.e., it was assumed to have the same fitness effect regardless of the genetic background in which it was found). Because the mutation 4703T appeared only in the 4988A background and increased in frequency during these passages, isolates with the 4703T mutation were not included as observations of the 4988A genotype. Thus at passage 36, for example, where four of the eight isolates assayed had both the 4703T and 4988A mutations while the remaining four isolates had only the 4988A mutation, the data from that passage counted as four observations of the 4988A genotype and zero observations of the 4967A genotype.

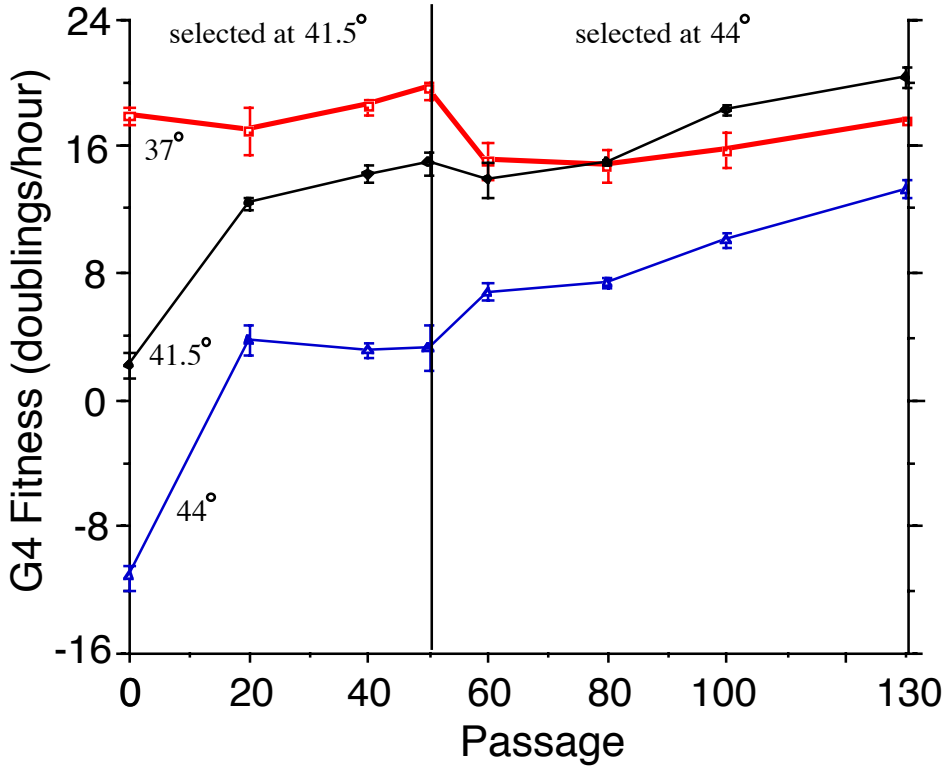
The maximum likelihood value was obtained for an s value of 0.0197 for the 4967A genotype; i.e., the preferred hypothesis for explaining the observed data was that the 4967A genotype had a slightly higher fitness than the 4988A genotype. Using

Figure 4. Fitness at multiple temperatures. (a)ϕX174 at 37º and 44º. (b) G4 at 37º, 41.5º, and 44º.

(a)


24
16
8
0
-8
ϕX174 Fitness (doublings/hour)
0
5
15
30
50
Passage
44°
37°

(b)


24
16
8
0
-8
-16
G4 Fitness (doublings/hour)
0
20
40
60
80
100
130
Passage
selected at 41.5°
selected at 44°
37°
41.5°
44°

a likelihood ratio test, values of s for the 4967A genotype, compared to the 4988A genotype, ranging from –0.0358 to 0.0838 could not be rejected.

Fitness tradeoffs: Large increases in fitness in the selected environment did not have to come at the expense of fitness in the ancestral environment. While the fitness at 37° of the ϕX174 population, after 30 passages at 44°, had fallen significantly below that of the ancestor (Figure 4a), the fitness at 37° fitness of G4 evolved for 130 passages at elevated temperatures was not significantly different than the fitness of its ancestor ($p>0.2$; Figure 4b).

DISCUSSION

This study provides the most precise mapping to date between fitness changes and genetic changes over an entire genome. Growth of ϕX174 and G4 at high temperature resulted in rapid and significant changes in fitness. Yet, despite the general similarity of the two phages, the amount of genetic change required to adapt to high temperature was quite different. ϕX174 was able to adapt directly to 44º with an average of just three substitutions that affected two genes. G4 required an initial adaptation to an intermediate environment before successfully adapting to 44º. A total of 14 substitutions, in five genes and an intergenic region, had accumulated by the end of the G4 selection. In both systems, there was the expected pattern of rapid fixation of some mutations while other sites remained polymorphic for many passages.

Fitness Profiles

There may be general patterns that characterize adaptive evolution of microbial systems. Identifying these patterns could enable scientists to use short-term responses to selection to make predictions about rates and long-term outcomes of evolution. Such predictions would facilitate the design of experiments to generate attenuated viruses that are less likely to revert to a virulent form and the development of protocols to maximize the effectiveness of drug therapies.

The data generated in this study provide information about the magnitude of the fitness effects of mutations, the fate of those mutations in the evolving populations and their effects in other environments. The comparison of two phage systems in this study make it possible to look past the specifics of the adaptations, which varied considerably, and focus on general outcomes.

Large-effect mutations

There is debate over the role of large and small effect mutations in adaptive evolution. While beneficial mutations of small effect are thought to be more common (Fisher 1930), there are theoretical reasons to expect large-effect mutations to be important in adaptation (Orr 1998; Orr 2000b; Gerrish and Lenski 1998). Burch and Chao (1999) found support for adaptation by small steps, as proposed by Fisher (1930). In their study, they first induced a deleterious mutation in the RNA phage ϕ6 and then allowed the phage, maintained at varying population sizes, to recover fitness. Although a reversion of the presumed deleterious mutation would have

recovered the fitness in a single step, this rarely occurred. Most populations increased in fitness by multiple small improvements. Bull *et al*. (2000) used growth inhibition at high temperature to identify several big-benefit mutations for ϕX174, including the mutation 1727T that was observed in this study. Because they were assaying specifically for large-effect mutations, their study verifies the existence of such mutations without providing information on their relative frequencies or their role in adaptive evolution.

In this study, both ϕX174 and G4 adapted to elevated temperatures by way of large-effect mutations. The 1727T mutation alone accounts for over 60% of the total gain in fitness of ϕX174 at 44° (based on a fitness estimate for 1727T from Bull *et al*. 2000). Because of the short duration of the experiments, there was not enough time to accumulate many small effect mutations. Had the selection been continued, small effect mutations might have dominated the remaining adaptation. However, after just three mutations per genome, the fitness of ϕX174 seemed to be approaching a biological limit; it was close to the fitness of ϕX174 that had been propagated for 66 passages at 37° (Bull *et al*. 2000). Regardless of the final number of beneficial mutations, the bulk of the fitness increase would be due to the initial three large-effect mutations. While the fitness gains in G4 were not as dramatic, the first two substitutions accounted for ~30% of the total increase in fitness at 44° (data not shown). This pattern, in which a major adaptive shift is brought about by a small number of substitutions of large-effect, has been identified in comparisons of species

differences in a variety of enzymes (reviewed by Golding and Dean 1998).

While the experimental design used in this study does not provide a very precise test of the relative frequencies of small and large effect mutations, the rapid fitness increases in both phage show that adaptation is certainly not limited to the accumulation of small effect mutations. Large-effect mutations are sufficiently common as to be readily accessible for the evolution of populations that could be considered moderate in size in comparison to natural populations of microbial organisms.

Selective Sweeps

In asexual populations, beneficial mutations are expected to fix in a series of selective sweeps (the periodic selection of Atwood *et al.* [1951]), with low levels of polymorphism accumulating between sweeps. This pattern of evolution has been observed in chemostats (Wichman *et al.* 1999). Other studies have observed stepwise increases in fitness and assumed evolution by selective sweeps to use these steps to infer the number of mutations and the distribution of mutational effects from fitness changes in lieu of direct observation of molecular changes (Lenski *et al.* 1991; Elena *et al.* 1996; Burch and Chao 1999).

□ Of the beneficial mutations whose frequencies were monitored during the evolution of ϕX174 and G4, several arising early in the adaptation swept to fixation. However, although selective sweeps were observed, generally characterizing the adaptation of the phage in this manner would be misleading as both populations were

highly polymorphic during many of the passages.

Frequent Polymorphism

Mutations that reached high frequency without fixing during the duration of the experiment or before being replaced by alternate mutations were observed in both phage systems. One explanation for these patterns of polymorphism is clonal interference (Gerrish and Lenski 1998). When multiple beneficial mutations are present at the same time in an asexual population, the mutation resulting in the most fit genotype will ultimately displace the other mutants as it reaches fixation. The time it takes to reach fixation can be greatly lengthened by the presence of the co-occurring beneficial mutants; during this slow periodic sweep, high levels of polymorphism may be observed. Mutations may reach high frequency but will disappear if a mutation with higher fitness arises before they fix. The same patterns can be observed in sexual populations as a result of epistasis. Although beneficial mutations can reach fixation simultaneously if they remain beneficial when recombined into the same genome, epistatic interactions may cause mutations that are individually beneficial to be deleterious when recombined. Epistasis, rather than clonal interference, could be the mechanism underlying the pattern of polymorphism in these phage. The persistent polymorphisms often involved mutations occurring in the same gene and even in the same codon. The pattern of evolution at 4703 in G4, in which the mutation was gained, lost, then regained, suggests epistasis. It is difficult to assess the importance of recombination in these phage systems. Although the phage

were capable of recombining, multiply-infected cells are necessary to provide the opportunity for recombination. Because the initial MOI was kept low, these opportunities should have been quite rare early in each passage, allowing the phage to evolve as asexuals. The greatest opportunity for recombination would have been at the end of each passage when the frequency of multiply-infected cells was highest. However, the probability of stochastic loss was also highest at this stage in the passage because of the lack of time for a new genotype to increase in frequency before the population was put through a bottleneck to begin the next passage.

Previous studies have uncovered apparently stable polymorphisms in other experimental populations. Studies of *S. cerevisiae* (Adams *et al.* 1992) and *E. coli* (Helling *et al.* 1987; Rosenzweig *et al.* 1994; Turner *et al.* 1996; Papadopoulos *et al.* 1999) have shown that stable polymorphisms can be established in environments consisting of a single limiting resource. In these systems, stable polymorphism appears to be the result of multiple niches forming (reviewed by Rainey *et al.* 2000) and requires complicated, frequency-dependent cross-feeding interactions. Elena *et al.* (1997) demonstrated a stable polymorphism in an RNA virus, vesicular stomatitis virus, by frequency-dependent selection. Although they did not identify the mechanism by which the frequency dependence was generated, they believed it could be caused either by the existence of multiple niches, due to heterogeneity in the host cells, or by intracellular interactions between the viral genotypes.

The polymorphisms in this study persisted in the absence of conditions that

could be considered permissive for frequency-dependent selection. The selective environment was kept constant during the course of the individual passages and across passages. Phage growth was terminated while there was an excess of host, minimizing coinfection and intracellular competition. A homogenous environment was maintained by using a single host (cells within a passage were derived from a single colony) growing at log phase in shaken flasks. While it is difficult to imagine a biological scenario that would generate frequency dependence in these phage systems, eliminating the possibility of frequency-dependent selection would require passaging experiments in which genotypes were started at various initial frequencies and large numbers of isolates assayed at the end to accurately determine changes in genotype frequencies. This would be impractical for the G4 polymorphisms as genotypes continued to acquire additional mutations.

With such limited opportunities for frequency-dependent interactions of genotypes, the persistent polymorphism in these systems suggests the possibility of multiple genotypes having similar absolute fitness. There are three lines of evidence that support the plausability of this explanation. Fitness of two ϕX174 genotypes, with either the 926T or 927T mutation, were indistinguishable by direct estimation. Likelihood analysis of persistent polymorphism in G4 showed that the fitness differences of the genotypes did not have to be negligible to be consistent with this explanation; the range of compatible differences in selection coefficients included values as high as 0.08. Furthermore, the G4 mutations, 4967A and 4988A, were

located in close proximity to each other in the same gene while the ϕX174 926T and 927T mutations affected the same amino acid residue. It is not unreasonable to think that they were equally acceptable solutions to the same underlying problem. The difficulty with this explanation is the timing of the occurrence of the mutations. For example, if the ϕX174 926T and 927T mutations resulted in genotypes of identical fitness, the mutations would have had to arise virtually simultaneously (during the same infection cycle) to reach the observed frequencies of the two genotypes. More accurate estimates of the frequencies of the genotypes over the course of multiple passages are needed to determine how similar the fitness of the genotypes are in order to put bounds on the window of opportunity during which the mutations had to occur.

Regardless of the mechanism by which it was maintained, the level of polymorphism observed at the end of the 41.5° selection of G4 could have had a significant effect on the response to selection at 44° had a single isolate not been used to initiate that selection. It has been shown that even low frequency genotypes, maintained at a mutation-selection balance in pretreatment conditions, can greatly affect the time to resistance to HIV-1 therapies (e.g. Frost *et al*. 2000). Additional experiments were conducted to examine the prevalence and impact of polymorphism in this system and are discussed in detail in Chapter 3.

Correlated Fitness Effects

A common generality of adaptive evolution is that an increase in fitness in one environment will often come at a cost in fitness in alternate environments (Huey and

Hertz 1984; Bell 1997). It is the underlying assumption for trying to attenuate viruses by selecting them in different environments (Bull 1994), and it is often the basis for strategies to avoid the evolution of resistance to antibiotics and pesticides. However, both phage obtained tremendous fitness gains at 44° with only a minor loss (ϕX174) or no detectable loss (G4) of fitness at 37°.

There are other examples of adaptation without a concomitant reduction in fitness in another environment. Attempts at viral attenuation by adaptation to culture conditions have not always been successful (Fenner and Cairns 1959). Bacteria have reduced and even eliminated the initial cost of antibiotic resistance through compensatory mutations (Lenski 1998; Schrag and Perrot 1996).

In this study, it is possible that the lack of tradeoffs were the result of many of the changes being adaptive to aspects of the culture conditions other than temperature. There is evidence that neither of the ancestor phages were fully adapted to the culture conditions at 37°. ϕX174 that had been passaged at 37° had a higher fitness at 37° (Bull *et al*. 2000) than my ϕX174 ancestor, and the initial adaptation of G4 to 41.5° increased fitness at all three of the temperatures assayed. It is only when a species is at a fitness optimum that any mutation is expected to be detrimental or, at best, neutral. However, in studies of high temperature adaptation of *E. coli* that were already well-adapted to the culture conditions, correlated changes at non-selected temperatures were found to be mostly positive (Bennett *et al*. 1990; Bennett *et al*. 1992).

My results still serve to highlight the problem of assuming tradeoffs are an inevitable consequence of adaptive evolution. The ability of viruses with 5.5kb genomes to evolve drastic increases in fitness while experiencing a general lack of significant tradeoffs at non-selected temperatures demonstrates the potential difficulty in using adaptive evolution to attenuate viruses.

3. EFFECTS OF POLYMORPHISM AND EPISTATIC INTERACTIONS ON ADAPTATION

INTRODUCTION

After 50 passages at 41.5°, the G4 population was highly polymorphic, with four genotypes identified among nine isolates examined. Mutations arising as early as passage 12 increased to high frequency but failed to fix. Two possible explanations for this persistent polymorphism can be tested by extending the passaging at 41.5°. Continued polymorphism would support the stability of the polymorphism and suggest experiments to determine if the stability is due to frequency-dependence or other interactions of the phage genotypes. Loss of polymorphism would indicate that the polymorphism was transient and the process of fixation prolonged by the presence of competing genotypes.

Persistent polymorphism was observed in both phage systems, implying this may be a common phenomenon of adaptation. The high level of standing variation in a polymorphic population may enable it to respond more quickly to a change in environment than a homogenous population. The importance of polymorphism as a feature of adaptation in a stable environment and in determining the rate of evolution after a change in the environment was tested by passaging polymorphic populations and populations started from single genotypes.

The order in which mutations are observed to fix during adaptation depends on the order in which they arise in the population, their relative fitness effects when multiple beneficial mutations are present in the population simultaneously and epistatic interactions. A general pattern observed during the evolution of the two phage systems in Chapter 2 was that the earliest substitutions had the largest fitness effects. Did the late substitutions appear so late in the adaptation because they were unable to outcompete the earlier substitutions, or were their beneficial effects conditional upon the presence of the earlier substitutions? Because only a single line of each phage was adapted to high temperature, there are insufficient data to identify the factors that determined the order of substitutions. However, the reversion of 4703T to the ancestral base after the population was fixed for the mutant and the subsequent regain of the derived base much later in the selection suggests epistatic interactions between mutations. To examine the importance of epistasis, mutations observed in the final passages of the adaptation of G4 to high temperature were placed in the ancestor background to test the dependence of their fitness effects on the presence of earlier substitutions.

MATERIALS AND METHODS

Polymorphism

Phage: During the adaptation of G4 to 41.5° described in Chapter 2, the lysate from each passage was frozen in borate EDTA, and clonal isolates from every third

passage were archived. A polymorphic population of phage was obtained from frozen lysate from the final passage at 41.5° (passage 50). The lysate was used to initiate a single passage at 41.5°, producing a lysate free of borate EDTA. I refer to this lysate as 50*. A consensus sequence of the 50* lysate had the derived states, 247T, 720T and 3665T (these sites appeared fixed in the population at passage 50) and 1269G and 4967A (these sites were polymorphic at passage 50). Additional sites that were identified as polymorphic at passage 50 had the ancestral states, 4703T and 4988A, in the consensus sequence.

Extension of Selection of G4 at 41.5°: To determine the nature of the polymorphism in the population of G4 after 50 passages at 41.5°, the experiment was extended for an additional 20 passages. The 50* lysate was used to begin the first passage. The passaging protocol for all of the experiments in this chapter is described in Chapter 2 under G4 passages. Lysate from the final passage of this experiment, the 70th passage of G4 at 41.5°, was sequenced. To check for polymorphism, 96 isolates from this passage were archived and probed with radio-labeled oligonucleotides.

Passaging of Individual G4 Genotypes: Stocks of monomorphic populations were made from isolates with three different genotypes observed during the adaptation of G4 to 41.5°. Genotypes of these populations were verified by sequencing. Each was then individually propagated for 20 passages. The original stock and lysate from the final passage were sequenced. Genotypes of multiple isolates from the final passage

for each line were determined by sequencing or probing with oligonucleotides to check for polymorphism.

Adaptation of a Polymorphic Population at 44°: In the experiment described in Chapter 2, a single G4 isolate was selected from passage 50 at 41.5° to initiate the selection at 44°. To examine the effect of variability in the initial population on adaptation to high temperature, the 50* lysate was used to establish additional lines for passaging at 44°. For the first experiment, the phage were simply propagated for ten passages at the elevated temperature. For the second experiment, the 50* lysate was used to initiate a single high MOI passage at 41.5° to promote recombination. The resulting phage were then propagated for ten passages at 44°. Lysate from the final passage of each experiment was sequenced and 96 isolates probed to check for polymorphism.

Epistasis

Fitness of Late Mutations in an Ancestral Background: To examine the possibility of epistatic interactions, I placed mutations observed during the last 30 passages of G4 at 44° in the 37°-adapted ancestral background by site-directed mutagenesis, using a protocol modified from Sambrook *et al*. (1989). This was done by generating PCR products with known mutations that were then denatured in the presence of ancestor DNA to form a heteroduplex. The heteroduplex DNA was then electroporated into host cells that will convert the phage DNA to RF. Some fraction of repairs or replication of heteroduplex genomes will generate mutant G4 genomes.

To generate PCR products with known mutations, I used PCR amplified fragments of the G4 genome from lysate from passages 100 and 130 (the 50th and 80th passages at 44°) as the DNA template for additional reactions with internal primers. This resulted in PCR products containing two to three mutations. To obtain ancestor DNA, I plated 37°-adapted G4 ancestor (used to start the selection at 41.5°) at 37°, collected multiple isolates and extracted DNA by means of a phenol/chloroform extraction. PCR product and ancestor DNA were added to reactions containing 2µl 10X T4 polymerase buffer, 4µl 100µM dNTPs, 2µl 10X BSA and enough sterile water to give a total volume of 20µl. PCR product was annealed to ancestor DNA by slowly cooling the reactions to 37° after denaturing at 94°. T4 polymerase was then added and the reaction allowed to proceed at 37° for an hour to extend the length of the PCR fragments. The DNA was then co-precipitated with 0.5µl yeast tRNA as a carrier by the addition of ethanol. After washing the precipitate in 70% ethanol, it was air dried and resuspended in 10µl sterile water. Using 3µl of the solution, I electroporated competent *E. coli* C and plated the transfected cells at 37°. Once plaques were visible, 1-2mL LB were added and the top agar scraped into microcentrifuge tubes. Chloroform was added, and the tubes were centrifuged to pellet the debris. The liquid layer, containing resuspended phage from the plaques, was saved, providing a stock consisting of the G4 ancestor and an unknown frequency of mutagenized phage.

To determine if the mutagenized phage were more fit than the ancestor, two lines, begun from stocks containing different mutant genotypes, were serially propagated for five passages at 41.5°. Changes in mutation frequency were then tested by probing archived isolates. I screened 48 isolates from the stock used to initiate the passages and 48 isolates from the final passage for each experiment. Based on those results, fresh stocks were made from isolates of varying genotypes and assayed for fitness at 41.5°, following my standard fitness assay protocol (described in Chapter 2). The stocks were sequenced to verify genotypes.

RESULTS

Polymorphism

Extension of Selection of G4 at 41.5°: The original selection experiment at 41.5° was extended for an additional 20 passages (Table 4). The consensus sequence of the 50* lysate used to initiate the first of the additional passages had the derived states 1269G and 4967A and the ancestral states 4703C and 4988G; these sites were polymorphic at the end of the original 50 passages of selection at 41.5°. After 20 passages, the consensus sequence now showed the derived states 4703T and 4988A as well as 125G, a mutation that had previously been observed during selection at 44°. Sites 1269 and 4967 had been replaced by the ancestral states. Probing of 96 isolates from the final passage (passage 70 at 41.5°) indicated that the mutations 4988A and 4703T appeared fixed, and 125 and 1269 were polymorphic. Site 4967

Table 4. Fate of polymorphic sites after an additional 20 passages at 41.5°.

Nucleotide	50* Consensus	Frequency[a] of the Derived Base (Passage 50)	Passage 70 Concensus	Frequency[b] of the Derived Base (Passage 70)
1269A→G	G	0.75	A	0.1
4967G→A	A	0.625	G	0.0
4703C→T	C	0.11[c]	T	1.0
4988G→A	G	0.375	A	1.0
125A→G	A	0.0	G	0.8

[a]Frequency of the derived base in 8 isolates assayed.
[b]Frequency of the derived base in 96 isolates assayed.
[c]Frequency based on presence in a ninth isolate selected for sequencing.

was fixed for the ancestral base. Three genotypes were observed among the 96 isolates from passage 70, all with the mutations 247T, 720T, 3665T, 4703T and 4988A (compared to 37°-adapted ancestor). One genotype had the additional mutation 125G (frequency of 0.8), a second genotype had the mutation 1269G (frequency of 0.1), and the third genotype had the ancestral base at both sites (125A and 1269A; frequency of 0.1).

Passaging of Individual G4 Genotypes: Monomorphic populations were generated from archived isolates for additional passaging at 41.5° to examine the prevalence of polymorphism in adapting populations. The three populations will be referred to hereafter as 247/3665, 4703/4988 and 1269/4967.

Sequencing the initial population of 247/3665 verified that it differed from the 37°-adapted ancestor at only those two sites. The consensus sequence of lysate from the 20th passage had acquired the mutations 720T and 4988A. These mutations were present in all 96 isolates probed from the 20th passage. The 4703T mutation was observed at low frequency (0.02). This is very similar to the result of the original adaptation of G4 to 41.5°, described in Chapter 2 (Table 5). During that experiment, the 247/3665 genotype appeared fixed in the population by passage 9. The majority genotype at passage 30 (the genotype of all eight isolates assayed) had the mutations at 720 and 4988. The 720T mutation was first observed at passage 18 and appeared to fix in the original experiment while 4988A remained polymorphic for 38 passages. The 4703T mutation was first observed at passage 33 and remained polymorphic for

Table 5. Changes in an initially monomorphic 247/3665 population after 20 passages of selection at 41.5°. These results are compared to passage 30 of the original selection of G4 at 41.5° because the 247/3665 genotype appeared fixed in the population at passage 9 of that experiment.

Mutation	Frequency at the end of 20 Passages[a]	Frequency at Passage 30 of the Original Selection[b]
720T	1.0	1.0
4988A	1.0	1.0
4703T	0.02	0.0[c]

[a]96 isolates assayed.
[b]8 isolates assayed.
[c]This mutation was observed in one of eight isolates assayed from passage 33.

the subsequent 16 passages.

Two additional populations were, separately, selected at 41.5° for 20 passages with no apparent molecular changes arising. The 4703/4988 population differed from the ancestor at five sites: 247T, 720T, 3665T, 4703T and 4988A. The consensus sequence of the lysate from the 20th passage showed no additional changes. Ten isolates from that passage were sequenced, and none contained any additional changes. The 1269/4967 population also differed from the ancestor at five sites: 247T, 720T, 3665T, and 1269G and 4967A. After 20 passages no additional changes were detected in the consensus sequence. Sequencing of ten isolates from the final passage also failed to detect any mutations.

Simulations of Mutation Rate and Selection Efficiency: The lack of additional mutations after 20 passages of selection on populations derived from single isolates was surprising. It is known from the other adaptations that G4 is capable of higher fitness at 41.5°; the 125G mutation is able to increase significantly the fitness of the 4703/4988 genotype. To assess the effect of mutation rate on the probability of observing mutations of known fitness effect and to determine the efficiency of selection in the passaging experiments, experiments were simulated using software written by Mark Holder.

Estimating burst time and burst size: In order to model the dynamics of the changes in the phage population during a passage, experiments were conducted to estimate the time from addition of phage until first burst and burst size at 41.5° for three G4

genotypes: the 37°-adapted ancestor, genotype 247/3665, and the dominant genotype from passage 70 at 41.5° (I will refer to it as the 4703/4988/125 genotype; this genotype differed from the 37°-adapted ancestor by six mutations: 247T, 720T, 3665T, 4703T, 4988A, and 125G).

Cells were grown according to the passaging protocol and 10^3 to 3 x 10^4 phage added, depending on the estimated fitness of the phage genotype. After five minutes, the culture was distributed into tubes which could then be removed at different times with minimal disturbance to the remaining tubes. At various times the mixture of infected cells and free phage was plated and phage density estimated by plaque formation. The culture was not treated with chloroform so that multiple mature phage within an unlysed cell would form a single plaque when plated. Estimated phage density increased only once the phage began lysing cells.

Early mutations were sufficient to produce a burst time similar to that of more evolved genotypes; subsequent mutations increased fitness primarily by increasing burst size (see Table 6 for an example of data collected in one of the trials). In both trials using the 37°-adapted ancestor, phage density did not increase over the course of 40 minutes. In the two trials with genotype 247/3665, which had fixed by passage 9 during the original 41.5° selection, the burst time was approximately 20 minutes with a burst size of ten. Genotype 4703/4988/125, taken from passage 70 at 41.5°, had a burst time of approximately 20 minutes while the burst size had increased to 40 phage, based on three trials.

Table 6. Change in phage density of genotype 4703/4988/125 during one trial. Elapsed time is the time after phage were added to the cells.

Elapsed Time (min.)	Phage/ml
0	161.2
15	180
18	400
20	4800
22	5250
24	5650
26	5650
35	5800
Estimated Burst Size	35 phage/burst

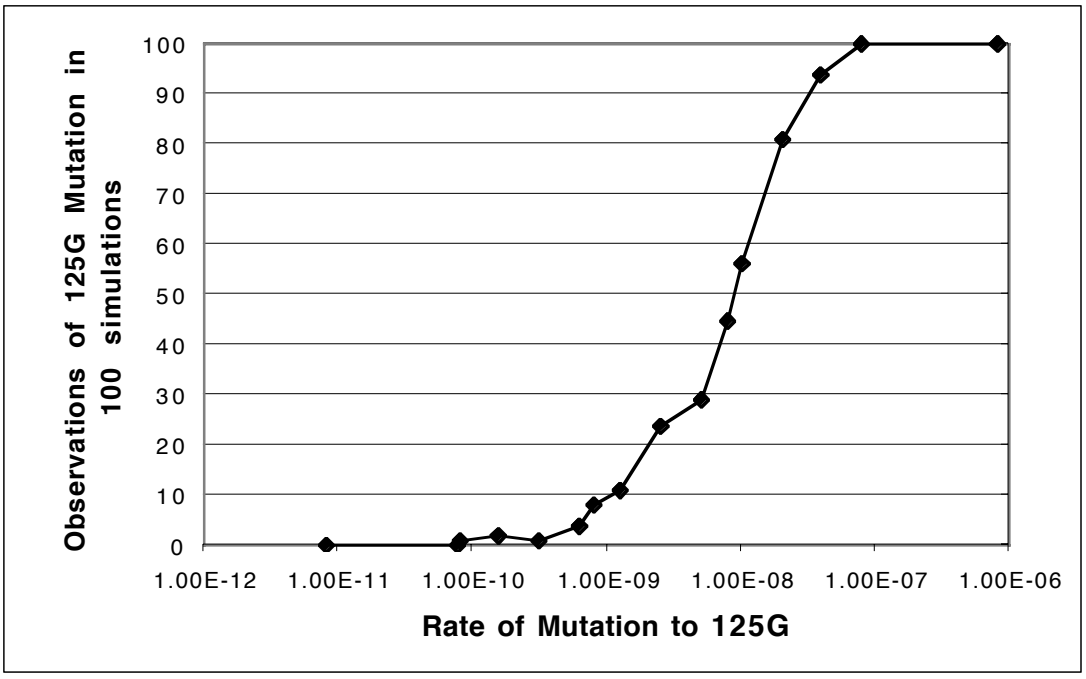
Simulation model: Based on the experimental data, burst time was held constant and burst sizes varied according to genotype fitness. In each passage, the phage population started at 10^6 and increased to 10^{10}, following the design of the empirical study. For computational ease, bursts and adsorption/reinfection were assumed to be synchronized; thus an hour long passage would permit three bursts (equivalent to three phage generations). Stochastic variation was incorporated into the simulations by using the mutation rate and a random binomial distribution to generate the number of mutant genotypes produced in each burst. The rate of mutation to a deleterious genotype, with an assumed fitness of zero, was held constant in all simulations at 0.004, the estimated genomewide mutation rate (Drake *et al*. 1998). Drift was incorporated into the simulations by using a random number generator to sample the population at the end of one passage to generate the starting population for the subsequent passage. If the frequency of a genotype was less than 0.0001, sampling was done without replacement to avoid oversampling. For genotypes with frequencies greater than 0.0001, sampling was done with replacement to speed up the calculations.

Mutation Rate Simulations: Despite its large fitness effect, the mutation 125G was not detected in ten isolates sequenced from the 4703/4988 population that had been propagated for 20 passages at 41.5°. To estimate a mutation rate compatible with this observation, the passaging experiment was simulated for a range of mutation rates and replicated 100 times for each rate.

In each simulation, the first passage began with a monomorphic population. The burst size for this ancestral genotype was 21.5, based on the estimated fitness of the 4703/4988 genotype. During each passage, the population underwent three bursts (each burst is equal to a phage generation). For each burst, the number of mutant genotypes generated was calculated as described above. For each set of simulations, the mutation rate to the 125G mutant was decreased by an order of magnitude, starting at a rate of 8×10^{-7}. That rate was based on the genomewide mutation rate of 0.004 and a genome size of 5000 bases (not the actual 5577). The burst size of the evolved genotype was 40, based on the results of the burst experiments for that genotype, giving it an s value of 0.86 (calculated as [burst size of evolved genotype/burst size of ancestor genotype] –1). If the 125G mutation arose during a passage, the population at the end of the passage would contain two genotypes, the original ancestor genotype and the 125G evolved genotype. This polymorphic population would then be sampled, as described above, to generate the starting population for the subsequent passage.

At the end of each simulation, a sample of ten individuals from the population was randomly generated, and the mutation was scored as found in the simulation if it was present in any of the ten individuals. The mutation was found in 100% of the replicates for the initial mutation rate and a rate 10-fold lower (Figure 5). The frequency with which the mutation was found rapidly declined with subsequent drops in the mutation rate. When the mutation rate had been lowered 100-fold (to a rate of 8

Figure 5. Results of mutation rate simulations. The frequency with which the 125G mutation was observed in simulations of an initially monomorphic 4703/4988 population after 20 passages of selection at 41.5° is given for a range of mutation rates.


Observations of 125G Mutation in 100 simulations
100
90
80
70
60
50
40
30
20
10
0
1.00E-12
1.00E-11
1.00E-10
1.00E-09
1.00E-08
1.00E-07
1.00E-06
Rate of Mutation to 125G

X 10^{-9}), the frequency with which the mutation was found was 45% and was no longer being detected in the population when the mutation rate reached 8 X 10^{-12}.

Efficiency of Selection Simulations: Similar simulations were performed in which the fitness of the beneficial genotype varied to determine the minimum s value needed for a mutation to become detectable in a sample size of ten after 20 passages. The burst size for the ancestral genotype was 20.15 (based on the estimated fitness of the 1269/4967 genotype); the burst size for the evolved genotype was increased by 0.1 for each set of simulations. The mutation rate to the evolved genotype was 6.5 X 10^{-7} (the estimated transition rate based on the genomewide mutation rate of 0.004 and the genome size of 5577 bases). The population was seeded with one individual with the mutant genotype to start the first passage (based on the estimated frequency of the evolved genotype in a plaque isolate grown from an ancestor phage). Simulations were replicated 100 times for each evolved burst size tested. Again, a mutation was scored as found in the simulation if it was present in any of the ten randomly sampled individuals from the population after 20 passages. The beneficial mutant was not found consistently in at least one of the 100 simulation replicates until it had an s value of 0.06. It was found at a frequency of 90% when s = 0.17 and 99% when s = 0.19.

Adaptation of a Polymorphic Population at 44°: Two lines were started from the 50* lysate for propagation at 44°. The difference in the two lines was the addition of a high MOI passage, to allow recombination, before beginning selection at 44° for one

of the lines. After ten passages at 44°, the results for both lines were essentially indistinguishable from each other and the original adaptation experiment. Of four sites that were polymorphic at the start of these selections, consensus sequences of the 44°-selected lines showed the derived state 4988A and the ancestral states 1269A, 4703C and 4967G. Additional mutations, 125G and 1585G, were present. Probing of 96 isolates from each of the two lines at 4988, 4703, 1585 and 125 indicated the populations were essentially fixed at all four sites (the frequency of the evolved nucleotide at 1585 was 0.99 in one of the experiments).

Epistasis

Increase in Frequency of Site-Directed Mutations: Mutations detected during the final passages of G4 at 44° were placed into the 37°-adapted ancestral background to determine if the beneficial effect of those mutations was dependent on the presence of substitutions that had accumulated earlier during adaptation to high temperature. Mutations 720T, 773A and 2000T were added in experiment 3.A. Mutations 5523A and 5570A were added in experiment 3.B. Site-directed mutagenesis resulted in a low frequency of mutagenized phage; the frequency of the mutations had increased dramatically after five passages at 41.5° (Table 6). This result demonstrates that the beneficial effect of the mutations was not specific to the genetic background in which they were originally observed nor to the more elevated temperature of 44°. However sequencing of isolates from the final passage of each experimental line (hereafter referred to as evolved phage) revealed all contained additional missense mutations.

Table 6. Results of site-directed mutagenesis. Initial frequency is the frequency of the mutations in the stock used to begin selection at 41.5°; final frequency is the frequency after five passages. Frequencies are based on probing of 48 isolates with the exception of mutation 720T.

Experimental Line	Site-Directed Mutations	Initial Frequency	Final Frequency	Additional Mutations
3.A	720T	unknown	0.5[a]	missense: 3665, 4988
	773A	0.021	0.458	silent: 1831, 2818
	2000T	0.042	0.604	
3.B	5523A	<0.02	0.958	missense: 3665
	5570A	<0.02	0.896	silent: 5481

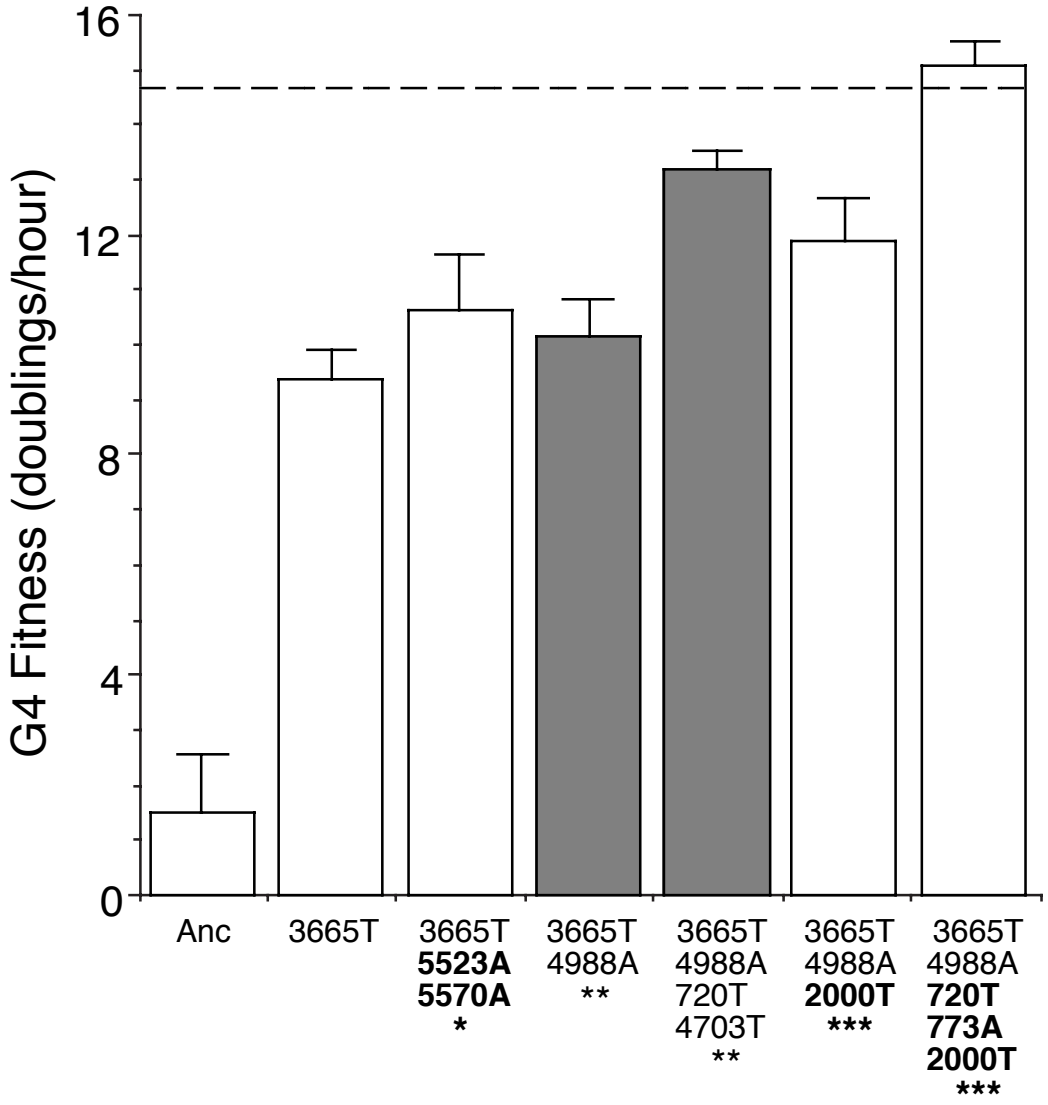
[a] Frequency based on presence in one of two isolates sequenced.

Isolates with site-directed mutations also had silent mutations.

Additional Mutations: All of the evolved phage had gained the mutation 3665T, the first missense substitution observed during adaptation of G4 to 41.5°. Contamination seems unlikely as this change was observed in isolates from both experiments and was not present in an ancestral stock made from an isolate from the initial phage population of experiment 3.B. As most of the fitness increase of the evolved phage can be explained with just that one mutation (Figure 4), it is plausible that the mutation arose *de novo* in both experiments and rapidly increased in frequency. An additional missense mutation (4988A, observed by passage 12 during adaptation of G4 to 41.5°) was observed in evolved phage of experiment 3.A. Silent mutations 1831C and 2818T were observed in both sequenced isolates of evolved phage from experiment 3.A. A silent mutation, 5481C, was observed in one of the two evolved isolates sequenced from experiment 3.B.

Fitness Effects of Site-Directed Mutations: Fitness assays of the evolved phage showed that genotypes that included late mutations, mutations originally observed during the final 30 passages of the selection of G4 at 44°, had significantly higher fitness at 41.5° when compared to the corresponding genotypes lacking those late mutations (Figure 6). There appeared to be multiple ways of achieving high fitness at 41.5° in a comparable number of mutational steps. The results indicate that the mutation 3665T is a large-effect mutation that will be nearly unavoidable given my passaging protocol.

Figure 6. Fitness of G4 genotypes at 41.5°. Anc is the 37°-adapted ancestor genotype. Missense mutations of the other genotypes are listed with site-directed mutations in **bold**. With the exception of the mutation 720T, all site-directed mutations were originally observed near the end of selection of G4 at 44°. Silent mutations are indicated with symbols: * is 5481C, ** is 247T and *** are 1831C and 2818T. They are omitted in the figure to facilitate comparison of the other mutations across genotypes. White columns are used for the fitness of isolates from the epistasis experiments. The first three genotypes are of isolates from experiment 3.B; the last two are from experiment 3.A. Gray columns are used for isolates archived during the original adaptation of G4 to 41.5°. The dashed line indicates the fitness of the polymorphic G4 population after 50 passages at 41.5°.


16
12
8
4
0
G4 Fitness (doublings/hour)
Anc
3665T
3665T
5523A
5570A
*
3665T
4988A
**
3665T
4988A
720T
4703T
**
3665T
4988A
2000T
***
3665T
4988A
720T
773A
2000T
***

DISCUSSION

Persistent polymorphism was a surprising outcome of the adaptation of ϕX174 and G4 to high temperature. Both the observed polymorphism and the results of the epistasis experiments, which did not indicate strong epistatic interactions between mutations, suggest that G4 can adapt to high temperature by a variety of pathways. However the results of the polymorphism experiments indicated that, despite the many possible pathways, there may be limitations to the number of adaptive pathways that a population is actually likely to follow.

Multiple Pathways of Adaptation

There are several lines of evidence that support the conclusion that there are multiple adaptive pathways available for G4 evolution that include intermediate genotypes of similar fitness. During the original adaptation to 41.5°, the population divided along different genetic paths. Mutations 4988A and 4967A arose and were unable to displace each other. Both lines acquired the 720T mutation without a noticeable change in their frequencies. The 4988A line further divided into at least three lines: one with mutation 4703T, one with mutation 1269G, the third with 153T. The 4967A line persisted with the addition of the 1269G mutation. The presence of these genotypes was detected in samples of eight isolates and from multiple times, indicating that this was not just chance sampling of rare genotypes. In the absence of any biologically realistic frequency-dependent selection, maintenance of these genotypes at these frequencies suggests that their fitnesses were quite similar. The

fitnesses of three genotypes from passage 50 at 41.5°, each with five mutations and differing from each other by one or two mutational steps, were indistinguishable. The introduction of mutations originally observed during selection at 44° into the ancestral background by site-directed mutagenesis provided more examples of alternate adaptive pathways as all of the mutations caused significant improvements in fitness at 41.5°.

The suggestion that there are multiple pathways of adaptive evolution is unsurprising. Evidence of multiple pathways has been documented in other studies including the fitness recovery of an RNA virus (Escarmis *et al*. 1999) and the adaptation of bacteria in a structured habitat (Korona *et al*. 1994). It is expected that stochastic processes will cause small populations to evolve along different paths (Wright 1932). What is surprising here is that these paths can have such similar fitnesses that a single large population was able to divide into multiple lineages that persisted for many passages with the accumulation of additional mutations. However a similar trend was observed in the bacterium *Comamonas* when adapted to a novel nutrient (Korona 1996). Replicate lineages selected in a homogenous environment evolved similar fitness but appeared to have followed different adaptive pathways, inferred by variability in isolates, both within and between lineages, at phenotypic traits correlated to the fitness increase.

Repeatability of Evolution

Another trend from the experiments in this chapter is that, despite multiple possible pathways, replicate lineages appeared to follow the same path. The substitutions during 20 passages at 41.5°, started from a population of 247/3665, generally repeated those of the original adaptation experiment. Two repetitions of the first ten passages at 44°, started from a polymorphic population, yielded identical results.

High rates of parallel evolution have been observed in other experimental phage lines (Bull *et al.* 1997; Crill *et al.* 2000; Wichman *et al.* 1999; Wichman *et al.* 2000). This could be explained simply by the fact that these experiments involve strong selection of large populations of species with a high mutation rate. Such populations should evolve in a fairly predictable manner because of their ability to effectively explore the mutational neighborhood. If all neighboring genotypes are explored at each step of adaptation, the mutation with the greatest fitness effect in that genetic background will be expected to fix. Replicate populations exposed to the same selective environment should evolve along the same path. In the G4 system, this is complicated by the fact that there were several mutations that resulted in similar fitness gains. I think that my data suggest another factor that may be important in determining the repeatability of evolution.

Mutation Rate Heterogeneity

The sample of neighboring genotypes that are explored during a given generation is not simply determined by the population size and the genomewide

mutation rate. There is evidence of considerable rate variation both among general classes of mutations, such as transitions and transversions, and among sites (Ronen and Rahat 1976). Results from experiments in this chapter suggest heterogeneity in the mutation rate in G4. During the extension of the passages at 41.5°, the 125G mutation occurred in a background that included the mutations 4703T and 4988A, increasing the fitness from ~13 to ~16 doublings/hour. Yet the 125G mutation was not detected after 20 passages of selection at 41.5° in a population derived from that same background genotype. The simulation results indicate that the mutation rate to that particular mutant had to be at least 100-fold lower than expected from the estimated genomewide mutation rate. In a study of ϕX174, Bull *et al*. (2000) found considerable variance in the frequency with which three large-effect mutations arose. During the original adaptations to high temperature, the first substitutions for both ϕX174 and G4 not only had large fitness effects but were C→T mutations, a common mutation for single-stranded genomes.

Under my passaging protocol, the population generally had time for three to four bursts before undergoing a bottleneck to begin the next passage. In order for a mutation to successfully fix in this experimental system, it not only had to be able to outcompete co-occurring beneficial mutations, it had to arise early enough in the passage that it could increase to a frequency sufficient to be among the genotypes sampled to begin the subsequent passage. In practice, this meant that, even for mutations with large selection coefficients, the mutation had to arise during the first

burst of a passage to have a reasonable chance of avoiding stochastic loss. Among the pool of mutations with similar fitness effects, heterogeneity in mutation rate can greatly influence which mutant is likely to fix.

Epistasis

The order in which mutations fix during adaptation may be affected by interactions between mutations. The fitness effect of a mutation can be dependent upon the context of the genotype in which it is observed (Bull *et al*. 2000). A mutation may not be beneficial except in the presence of another mutation. The loss and later reacquisition of the 4703T mutation during selection at 44° is indirect evidence of this type of interaction. To test the strength of epistasis in this system, I specifically introduced mutations that were observed at the end of adaptation at 44° (late mutations) into the ancestor background.

All of the introduced mutations increased in frequency in response to selection at 41.5°, but sequencing showed evolved phage had also picked up one or two of the mutations that had originally been observed during the first 12 passages of adaptation of G4 at 41.5° (early mutations). Without assaying fitness of the late mutations in the absence of those early mutations, it is impossible to determine if the presence of the early mutations in those isolates is simply due to their large fitness effects or if the beneficial effects of the late mutations were conditional upon the presence of the early mutations. It is clear from the results that the late mutations were able to significantly improve fitness at the lower temperature and in a very different genetic

background than the one in which they arose during selection at 44°. In a study of ϕX174, Wichman *et al*. (1999) reached a similar conclusion, that the parallel adaptive mutations in their chemostats were beneficial across a range of backgrounds, on the basis of the variability in order in which those substitutions occurred.

Most of the epistatic interactions in my system appear to be weak. The beneficial effects of the late mutations were not dependent upon the presence of the earlier substitutions. Furthermore, the mutations 720T and 1269G were observed in multiple genotypes during the selection of G4 to 41.5° which implies that those mutations also function to improve fitness in a variety of genetic backgrounds. What seems likely is that the order in which substitutions arose was dependent upon a combination of the magnitude of their fitness effects and the frequency with which the mutations occurred. All eight of the mutations observed during selection at 41.5° were transitions, five C→T; the only transversions were among substitutions observed during the last 30 passages at 44°.

Evolvability

On three occasions the same genotype in the polymorphic population at passage 50 at 41.5° was the ancestor of the genotype that appeared to fix, or to be approaching fixation, in the population. The 4703/4988 genotype was the ancestor of the genotype that appeared to be successfully fixing at 41.5° after an additional 20 passages. It was also the ancestor of the genotype that had fixed in the population after ten passages in both of the replicate lineages selected at 44°. Does this mean

that, among the genotypes present at passage 50, 4703/4988 had the greatest evolvability?

Evolvability is the ability to increase fitness by mutation (Ancel and Fontana 2000). It is generally assumed that an organism with a high mutation rate has high evolvability. Burch and Chao (2000) proposed that, even for organisms with such high mutation rates as RNA viruses, evolvability can be limited by a genotype's mutational neighborhood. In their study, they evolved replicate lineages of $\phi 6$ founded from clones from two populations of similar fitness but different genotypes. After 100 generations of additional selection, the lineages derived from one population experienced a mean increase in fitness while the lineages derived from the second population experienced a mean decrease in fitness. They interpreted these results as indicating that there were multiple advantageous mutations accessible to the first population whereas all genotypes only one or a few mutational steps away from the genotype of the second population were deleterious. Despite the high mutation rate, the second population was stuck on a suboptimal fitness peak.

A population started from genotype 1269/4967 and selected for 20 passages at 41.5° showed no increase in fitness, and no mutations were observed in ten sequenced isolates. This is insufficient to prove that the genotype's evolvability was limited by its mutational neighborhood. An alternative explanation of the apparent evolutionary stasis of the population is that a beneficial mutation simply did not arise quickly enough to reach a detectable frequency after 20 passages. The simulations indicated

that a mutation present at the start of the first passage would need an s of approximately 0.2 in order to be detected reliably in a sample of ten isolates after that period of selection. My data also show that the phage were only effectively exploring a subset of the mutational space that is theoretically accessible each generation given their population sizes and mutation rate. There were no additional substitutions in the population started from genotype 4703/4988 after selection for 20 passages at 41.5° even though a known genotype of significantly higher fitness was a single mutation away.

Adaptation of genotypes during these experiments is limited not only by the existence of more fit genotypes only one mutational step away but by the frequency with which those mutations occur. It may be that rate of adaptation should be considered a component of the evolvability of a genotype. However caution should be used when interpreting the results of experiments like these as tests of evolvability.

4. CONCLUSIONS

Microbial systems are attractive for the study of adaptive evolution because they are easily maintained in the lab where the environment can be carefully controlled, and adaptation occurs on a timescale of hours to months not hundreds to thousands of years. Viral systems, such as the ones I used, have the added appeal of genomes so small that numerous isolates can be completely sequenced and the actual genetic mutations causing the changes in fitness identified. While it is the very simplicity of these systems that makes them tractable for testing theories of adaptation, it also makes it difficult to generalize the conclusions to the evolution of more complex organisms. However, understanding the adaptive evolution of these systems is valuable in its own right, as microbes play important roles as human and plant pathogens and as possible solutions to the problem of these pathogens. As studies of microbial systems accumulate, trends that appear over a range of systems may indicate relevant aspects of adaptation that should be incorporated into theory.

Since Fisher introduced his geometric model, there has been an expectation that adaptive evolution will proceed by the accumulation of numerous mutations of small effect. However, mutations with very large fitness effects were observed during the adaptations of both ϕX174 and G4. It may be argued that substitutions with large effect are more likely to occur in these systems because of the experimental design or the simplicity of the study organisms, but large-effect mutations are not limited to

high temperature adaptation in bacteriophage. Large-effect mutations have also been observed in the evolution of antibiotic resistance (Knox 1995) and insecticide resistance (Newcomb *et al*. 1997; Campbell *et al*. 1998). Analyses of quantitative trait loci (QTL) have found numerous examples of QTL of large effect (Doebley and Stec 1991, 1993; Tanksley 1991).

Evidence supporting the occurrence of large-effect mutations in a variety of systems have led theoreticians to reconsider the probable distribution of fitness effects among adaptive substitutions (Orr 1998, 2000b; Gerrish and Lenski 1998). Experimental systems may overemphasize the general importance of large-effect mutations by exposing large populations to strong selection, which should greatly favor large-effect mutations; however, they are addressing a primary concern in efforts to combat microbial pathogens. It is increasingly apparent that a much higher probability of large-effect mutations providing quick solutions to challenges, such as antibiotics or host switches, should be incorporated into our thinking about how microbial systems are likely to evolve. Small-effect mutations may still dominate in the long run but the probability of large short term gains will determine the effectiveness of strategies to control pathogens.

Another important trend is mutational rate heterogeneity. The estimated genomewide mutation rate for bacteriophage is 0.004, but the mutation rate in G4 from A→G at base 125 had to be at least 100-fold lower than estimate implies in order to be consistent with the observed data. Ronen and Rahat (1976) found the

mutation rate of a base substitution can vary by orders of magnitude among sites in the bacteriophage T4. The 1727 C→T mutation was six times more common than the 1565 A→G mutation and 60 times more common than the 28 G→A mutation in ϕX174 (Bull *et al.* 2000). This rate heterogeneity helps to explain, in part, the repeatability of evolution that was observed in this study. Mutational bias in HIV-1 reverse transcriptase gives rise to an ordered appearance of mutations in the evolution of resistance to lamivudine (Keulen *et al.* 1997).

This represents an important violation of most theories of adaptation. Theoretical models generally rely on standard statistical distributions of the frequency of mutations of varying fitness effects. There may be many fewer large-effect than small-effect mutations, but, if one of those large-effect mutations is favored by mutational bias, it will be overrepresented in the population and have a decreased probability of stochastic loss. This could allow large-effect mutations to play a part in the adaptation of much smaller populations than expected given the small number of such mutations compared to the pool of small-effect mutations. It is also a reason why these systems may not generate the distributions of fitness effects predicted by theory.

A third trend that was observed in this study was frequent polymorphism. The theory of adaptation by periodic sweeps predicts that appreciable polymorphism will only occur during the waiting period for a new beneficial mutation. This model was already being called into question by the results of studies in yeast (Adams and Oeller 1986) and bacteria (Helling *et al.* 1987; Rosenzweig *et al.* 1994; Korona 1996).

Gerrish and Lenski (1998) emphasized the role of clonal interference in the evolution of asexual populations to provide a theoretical basis for expecting these higher levels of polymorphism. Much of the polymorphism that has been observed in these other systems has been the result of spatial or temporal heterogeneity in the selective environment or interactions between genotypes that promote frequency-dependent selection. This study provides a strong case for the persistence of polymorphisms as the result of similarity in absolute fitness of genotypes. The experimental system was designed to provide a homogenous selective environment with minimal opportunities for viral interactions. Monitoring of the spread of mutations and assaying of individual genotypes demonstrated that the mutations at the polymorphic sites were beneficial in the passaging environment.

The polymorphism in G4 at 41.5° was inherently transient as the population was not geographically subdivided and genotypes of higher fitness were still accessible. If one were to only look at the fitness profile and the end point of adaptation, it might appear to resemble the Fisherian model of adaptation in which the population steadily works toward a single fitness optimum. However, that description would fail to capture the actual nature of the adaptive process. This is proving to be the case for bacterial systems as well (Papadopoulos *et al*. 1999; Notley-McRobb and Ferenci 2000; Riehle *et al*. 2001).

The strength of this study was that the phage were evolved in a carefully defined environment in which the selective forces were easily recognizable. It

allowed me to present the most complete picture to date of the adaptation of organisms to a novel environment. The simplicity of the phage systems made it possible to begin to explain some of the unexpected complexities that arose. Repetition of experiments such as these does not just give the trite result that fitness increases during adaptation; it makes it possible to identify trends that are important in our understanding of the evolution of microbial populations in particular and highlight areas worth pursuing in the interest of understanding adaptive evolution in general.

REFERENCES

ADAMS, J., and P. W. OELLER, 1986 Structure of evolving populations of *Saccharomyces cerevisiae*: adpative changes are frequently associated with sequence alterations involving mobile elements belonging to the Ty family. Proc. Natl. Acad. Sci. USA **83:** 7124-7127.

ADAMS, J., S. PUSKAS-ROZSA, J. SIMLAR and C. M. WILKE, 1992 Adaptation and major chromosomal changes in populations of *Saccharomyces cerevisiae*. Curr. Genet. **22:** 13-19.

ANCEL, L. W. and W. FONTANA, 2000 Plasticity, evolvability, and modularity in RNA. J. Exp. Zool. **288:** 242-283.

ATWOOD, K. C., L. K. SCHNEIDER and F. J. RYAN, 1951 Selective mechanisms in bacteria. Cold Spring Harbor Symp. Quant. Biol. **16:** 146-155.

BELL, G., 1997 *Selection: the mechanism of evolution*. Chapman and Hall, New York.

BENNETT, A. F., K. M. DAO and R. E. LENSKI, 1990 Rapid evolution in response to high-temperature selection. Nature **346:** 79-81.

BENNETT, A. F., R. E. LENSKI and J. E. MITTLER, 1992 Evolutionary adaptation to temperature. I. Fitness responses of *Escherichia coli* to changes in its thermal environment. Evolution **46:** 16-30.

BULL, J. J., 1994 Virulence. Evolution **48:** 1423-1437.

BULL, J. J., M. R. BADGETT, and H. A. WICHMAN, 2000 Big-benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. **71:** 942-950.

BULL, J. J., M. R. BADGETT, H. A. WICHMAN, J. P. HUELSENBECK, D. M. HILLIS, *et al.*, 1997 Exceptional convergent evolution in a virus. Genetics **147:** 1497-1507.

BURCH, C. L., and L. CHAO, 1999 Evolution by small steps and rugged landscapes in the RNA virus ϕ6. Genetics **151:** 921-927.

BURCH, C. L., and L. CHAO, 2000 Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature **406:** 625-628.

CAMPBELL, P. M., R. D. NEWCOMB, R. J. RUSSELL and J. G. OAKESHOTT, 1998 Two different amino acid substitutions in aliesterase, E_3, confer alternative types of organophosphorus insecticide resistance in the sheep blowfly, *Lucilia cuprina*. Insect Biochem. Molec. Biol. **28:** 139-150.

CRILL, W. D., H. A. WICHMAN and J. J. BULL, 2000 Evolutionary reversals during viral adaptation to alternating hosts. Genetics **154:** 24-37.

DE VISSER, J. A. G. M., C. W. ZEYL, P. J. GERRISH, J. L. BLANCHARD and R. E. LENKSI, 1999 Diminishing returns from mutation supply rate in asexual populations. Science **283:** 404-406.

DOEBLEY, J. and A. STEC, 1991 Genetic analysis of the morphological differences between maize and teosinte. Genetics **129:** 285-295.

DOEBLEY, J. and A. STEC, 1993 Inheritance of the morphological differences

between maize and teosinte: comparison of results for two F2 populations. Genetics **134:** 559-570.

ELENA, S. F., V. S. COOPER and R. E. LENSKI, 1996 Punctuated evolution caused by selection of rare beneficial mutations. Science **272:** 1802-1804.

ELENA, S. F., R. MIRALLES and A. MOYA, 1997 Frequency-dependent selection in a mammalian RNA virus. Evolution **51:** 984-987.

ESCARMIS, C., M. DAVILA and E. DOMINGO, 1999 Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. J. Mol. Biol. **285:** 495-505.

FENNER, F. and J. CAIRNS, 1959 Variation in virulence in relation to adaptation to new hosts, in *The Viruses*, *Vol. 3, Animal Viruses*, edited by F. M. BURNET and W. M. STANLEY. Academic Press, New York.

FISHER, R. A., 1930 *The genetical theory of natural selection*. Oxford University Press, Oxford, England.

FROST, S. D. W., M. NIJHUIS, R. SCHUURMAN, C. A. B. BOUCHER AND A. J. LEIGH BROWN, 2000 Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. J. Virol. **74:** 6262-6268.

GERRISH, P. J., and R. E. LENSKI, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102/103:** 127-144.

GODSON, G. N., 1974 Evolution of ϕX174. Isolation of four new ϕX-like phages and comparison with ϕX174. Virol. **58:** 272-289.

GODSON, G. N., B. G. BARRELL, R. STADEN and J. C. FIDDES, 1978 Nucleotide sequence of bacteriophage G4 DNA. Nature **276:** 236-247.

GOLDING, G. B. and A. M. DEAN, 1998 The structural basis of molecular adaptation. Mol. Bio. Evol. **15:** 355-369.

HELLING, R. B., C. N. VARGAS and J. ADAMS, 1987 Evolution of Eschirichia coli during growth in a constant environment. Genetics **116:** 349-358.

HILL, W. G. and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269-294.

HUEY, R. B., and P. E. HERTZ, 1984 Is a jack-of-all-temperatures a master of none? Evolution **38:** 441-444.

HAYASHI, M. A. AOYAMA, D. L. RICHARDSON, JR. AND M. N. HAYASHI, 1988 Biology of the bacteriophage ϕX174. In: *The Bacteriophages*, Vol. 2, edited by R. CALENDER. Plenum Publishing.

KEULEN, W., N. K. T. BACK, A. VAN WIJK, C. A. B. BOUCHER and B. BERKHOUT, 1997 Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational bias of human immunodeficiency virus type 1 reverse transcriptase. J. Virol. **71:** 3346-3350.

KIMURA, M., 1983 The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge, U.K.

KNOX, J. R., 1995 Extended-spectrum and inhibitor-resistant TEM-type beta-lactamases- mutations, specificity, and 3-dimensional structure. Antimicrob. Agents Chemother. **39:** 2593-2601.

KORONA, R., 1996 Genetic divergence and fitness convergence under uniform selection in experimental populations of bacteria. Genetics **143:** 637-644.

KORONA, R., C. H. NAKATSU, L. F. FORNEY AND R. E. LENSKI, 1994 Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. Proc. Natl. Acad. Sci. USA **91:** 9037-9041.

LENSKI, R. E., 1998 Bacterial evolution and the cost of antibiotic resistance. Internatl. Microbiol. **1:** 265-270.

LENSKI, R. E., M. R. ROSE, S. C. SIMPSON and S. C. TADLER, 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. **138:** 1315-1341.

LENSKI, R. E., and M TRAVISANO, 1994 Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. USA **91:** 6808-6814.

MIRALLES, R., P. J. GERRISH, A. MOYA and S. F. ELENA, 1999 Clonal interference and the evolution of RNA viruses. Science **285:** 1745-1747.

MIRALLES, R. , A. MOYA and S. F. ELENA, 2000 Diminishing returns of population size in the rate of RNA virus adaptation. J. Virol. **74:** 3566-3571.

MONGOLD, J. A., A. F. BENNETT and R. E. LENSKI, 1999 Evolutionary adaptation to

temperature. VII. Extension of the upper thermal limit of *Escherichia coli*. Evolution **53:** 386-394.

MULLER, H. J., 1932 Some genetic aspects of sex. Am. Nat. **66:** 118-138.

NEWCOMB, R. D., P. M. CAMPBELL, D. L. OLLIS, E. CHEAH, R. J. RUSSELL and J. G. OAKESHOTT, 1997 A single amino acid substitution converts a carboxylesterase organophosphorus hydrolase and confers insecticide resistance on a blowfly. Proc. Natl. Acad. Sci. USA **94:** 7464-7468.

NOTLEY-MCROBB, L. and T. FERENCI, 2000 Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. Genetics **156:** 1493-1501.

ORR, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52:** 935-949.

ORR, H. A., 2000a The rate of adaptation in asexuals. Genetics **155:** 961-968.

ORR, H. A., 2000b Adaptation and the cost of complexity. Evolution **54:**13-20.

PAPADOPOULOS, D., D. SCHNEIDER, J. MEIER-EISS, W. ARBER, R. E. LENSKI and M. BLOT, 1999 Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. USA **96:** 3807-3812.

RAINEY, P. B., A. BUCKLING, R. KASSEN and M. TRAVISANO, 2000 The emergence and maintenance of diversity: insights from experimental bacterial populations. TREE **15:** 243-247.

RIEHLE, M. M., A. F. BENNETT and A. D. LONG, 2001 Genetic architecture of

thermal adaptation in *Eschirichia coli*. PNAS **98:** 525-530.

RONEN, A. and A. RAHAT, 1976 Mutagen specificity and position effects on mutation in T4rII nonsense sites. Mutat. Res. **34:** 21-34.

ROSENZWEIG, R. F., R. R. SHARP, D. S. TREVES and J. ADAMS, 1994 Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics **137:** 903-917.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, *Molecular Cloning*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SANGER, F., G. M. AIR, B. G. BARRELL, N. L. BROWN, A. R. COULSON *et al.*, 1977 Nucleotide sequence of bacteriophage ϕX174 DNA. Nature **265:** 687-692.

SCHRAG, S. J. and V. PERROT, 1996 Reducing antibiotic resistance. Nature **381:** 120-121.

SNEDECOR, G. W. and W. G. COCHRAN, 1980 *Statistical methods*. 7th ed. Iowa State University Press, Ames, Iowa.

TANKSLEY, S. D., 1993 Mapping polygenes. Ann. Rev. Genet. **27:** 205-233.

TESSMAN, E. S. and I. TESSMAN, 1959 Genetic recombination in phage S13. Virol. **7:** 465-467.

TURNER, P. E. and L. CHAO, 1999 Prisoner's dilemma in an RNA virus. Nature **398:** 441-443.

TURNER, P. E., V. SOUZA and R. E. LENSKI, 1996 Tests of ecological mechanisms

promoting the stable coexistence of two bacterial genotypes. Ecology **77:** 2119-2129.

WICHMAN, H. A., M. R. BADGETT, L. A. SCOTT, C. M. BOULIANNE and J. J. BULL, 1999 Different trajectories of parallel evolution during viral adaptation. Science **285:** 422-424.

WICHMAN, H. A., L. A. SCOTT, C. D. YARBER and J. J. BULL, 2000 Experimental evolution recapitulates natural evolution. Phil. Trans. R. Soc. Lond. B **355:** 1677-1684.

WRIGHT, S., 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. 6th Int. Congr. Genet. **1:** 356-366.

WRIGHT, S., 1988 Surfaces of selective value revisited. Am. Nat. **131:** 115-123.

VITA

Kristina Kichler Holder was born on Thanksgiving Day, November 25, 1971, in Mobile, AL. Her parents are David and Kathryn Kichler. She graduated from McGill-Toolen High School in May, 1990, and began studying at Spring Hill College that fall. She graduated with a B. S. in marine biology in May, 1993. She entered the lab of Dr. David Owens at Texas A&M University that fall, receiving her M. S. in zoology in December, 1996. Her master's thesis was on multiple paternity in Kemp's ridley sea turtles; the research was published in Molecular Ecology. While a student at Texas A&M, she met Mark Holder. They were married on June 1, 1996. She enrolled at the University of Texas that fall. Upon completion of the requirements for her PhD, she plans to begin preparing for the arrival of Baby Holder, due in mid-October. Most of the work in this dissertation is included in a manuscript that has been accepted for publication in Genetics.

Permanent Address: 603 Wormwood Hill Rd., Mansfield, CT 06250

This dissertation was typed by the author.