INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

EXPERIMENTAL EVOLUTION AND MOLECULAR BASIS OF HOST-SPECIFIC VIRAL ADAPTATION

Approved by
Dissertation Committee:

James J. Bull, Co-Supervisor

David M. Hillis, Co-Supervisor

Ian J. Molineux

Paul A. Krieg

Nigel Atkinson

EXPERIMENTAL EVOLUTION AND MOLECULAR BASIS OF HOST-SPECIFIC VIRAL ADAPTATION

by

WAYNE DOUGLASS CRILL, B. S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

December, 1998

Dedication

To my parents.

Acknowledgements

I thank my surpervising professors, Jim Bull and David Hillis for all their attention over the years and encouragement and support. J. Bull, D. Hillis, P. Krieg, N. Atkinson, I. Molineux, H. Wichman, A. Graybeal, M. Brauer, M. Holder, and K. Kichler offered invaluable comments on the manuscript. Additionally, I thank R. Huey, D. Hillis, J. Bull, I. Molineux, R. Lenski, B. Levin., D. Begun, D. Cannatella, M. Kirkpatrick, H. Wichman, J. Bernardo, K. Crandall, C. Cunningham, A. Graybeal, D. Brumbaugh, B. Mable, T. Reeder, P. Chippendale, C. Austin, K. Halanych, J. McGuire, M. Brauer, M. Holder, K. Kichler, L. Dries, K. Agnew, and M. Badgett for helpful comments and encouragement either towards this project specifically or in general. J. Bull, H. Wichman, M. Badgett, and A. Gulati all directly helped to differing degrees with the research proper.

This research was funded by grants DEB-9411950 and DEB 95-08987 from the National Science Foundation and by grant MCB 95-06954 from the national institutes of health.

EXPERIMENTAL EVOLUTION AND MOLECULAR BASIS OF HOST-SPECIFIC VIRAL ADAPTATION

Publication No._____

Wayne Douglass Crill, Ph.D.

The University of Texas at Austin, 1998

Supervisors: James J. Bull and David M. Hillis

Experiments were designed to examine the molecular basis of host-specific viral adaptation. Replicate bacteriophage lineages were evolved in liquid culture, alternating between either of two bacterial hosts, *Escherichia coli* C and *Salmonella typhimurium*. There was a distinct asymmetry in the pattern of correlated responses to host-specific adaptation. Adaptation to *S. typhimurium* consistently resulted in the loss of all previous fitness gains on *E. coli*, whereas adaptation to *E. coli* did not affect fitness on the *S. typhimurium*. Hence, there was a strong one-way antagonistic pleiotropic fitness response.

Although numerous nucleotide substitutions accumulated throughout the experiment, relatively few substitutions in one gene, the major capsid protein gene, explained the majority of the host-specific fitness adaptation and the correlated fitness response. These five host-specific changes in the viral capsid gene consistently evolved to one nucleotide state associated with adaptation to *S.*

typhimurium, and then reverted back to their ancestral or previous state when adapted to *E. coli*. Analysis of the temporal dynamics of fitness and nucleotide changes in the evolving viral populations suggested that major changes in fitness resulted from a few of these host-specific substitutions. One of these host-specific substitutions appears to be necessary for host-specific adaptation, but site-directed mutagenesis demonstrated that it alone is not sufficient to explain the observed patterns of fitness evolution. Moreover, the fitness effects of this particular substitution differ dramatically depending upon the genetic background in which it is found. Two or three of these host-specific substitutions in the capsid gene repeatably cause most of the observed fitness evolution, implying that there are a limited number of pathways available for efficient host-specific adaptation. All five of these host-specific substitutions result in amino acid replacements that are located on the surface of both the capsid protein and the virion itself, and therefore are presumably available to interact with the cell surface. These results emphasize the importance having molecular data to make accurate inferences about the genetic basis of phenotypic responses. I discuss how the results relate to the evolution of host-specificity, correlated responses to selection, and to eukaryotic viral systems including attenuated viral vaccines.

Table of Contents

INTRODUCTION

Motivation-

Though little is known of overall viral diversity there clearly exits a large natural variance in viral host range. Some viruses have narrow host ranges, infecting a single species, or a group of closely related species, whereas other viruses seem to readily infect a wide range of distantly related hosts and hence have broad host ranges (Morse, 1993; Truyen *et al.*, 1995, 1996). The historic expansion and restriction of viral host range must have played an important role in generating present viral diversity. Important host range expansions still occur today with newly "emerging" viruses. Unfortunately the molecular mechanisms controlling such variance in host range are largely unknown (Fenner and Cairns, 1959; Novella *et al.*, 1995; but see Westrop *et al.*, 1989; but see Truyen *et al.*, 1995, 1996). Understanding the molecular bases of viral host specific adaptation and correlated responses to adaptation are interesting not only because of their broad applications within population biology proper, but also because of their direct relevance to the understanding, prevention, and treatment of many human viral diseases.

Many human viral diseases appear to be due to recent host shifts of viruses historically adapted to other hosts, newly "emerging" into the human host. Two examples are HIV (Myers *et al.* 1992) and influenza virus (Webster *et al.*, 1992). Other human viral diseases results not from the acquisition of humans as a novel host *per se*, but from dead end or "accidental" human infections. These cases are

"dead ends" because they do not result in transmission of the virus to another host, yet they often result in causing disease in the accidental host. Two examples of such dead end human disease viruses are hantaviruses such as the N. American "four corners" virus (Nichol *et al.*, 1993); and Ebola and other hemmoragic fever viruses (Peters *et al.*, 1993). The pathogenic results of such infections are not due to viral adaptation (since the virus has no history of adaptation in this host) but specifically result from correlated fitness responses to viral adaptation in the traditional host. What molecular mechanisms control such viral host shifts, and correlated responses to adaptation? Are there general molecular patterns involved? For example, are viral host shifts generally controlled by a few or by many molecular changes? What types of molecular interactions are involved? Which viral genes tend to be important for viral host specific adaptation? How do the answers to these questions affect correlated responses to host specific adaptation? Answering these questions will help us to understand how these viral diseases evolved, allow us to make predictions about the emergence of new viral diseases, and develop treatments to combat the vast array human viral diseases.

An understanding of the molecular basis of viral host specific adaptation and correlated responses to adaptation has important relevance to live attenuated viral vaccines. Vaccines function by initiating an immune response without causing disease, thereby leaving the vaccinated individual "immune" to future infection. Many vaccines utilize live "attenuated" or non-pathogenic viruses. Such attenuated viruses have historically been obtained by adapting the virus to a novel host, tissue, or temperature and looking for correlated responses to that adaptation which reduce pathogenicity back in the original host, tissue, or

temperature (Fenner and Cairns, 1959; Lwoff, 1961; Sabin and Boulger, 1973; Westrop *et al.*, 1989; Novella *et al*, 1995). Viral attenuation is therefore a correlated response to adaptation known as an antagonistic pleiotropic fitness constraint (Lenski, 1988; Bull 1994). Because attenuated viral vaccines are alive they can evolve to overcome attenuation and revert back to full pathogenicity in the host (Bull, 1994; Novella *et al.*, 1995; and Elena *et al.*, 1998). Though this rarely results in disease for the vaccinated individual they can readily infect and cause disease in non-vaccinated individuals Sabin and Boulger, 1973; Bull 1995). Additionally, pathogenic viruses can evolve to escape the immune response of vaccinated individuals and then cause disease. These three processes, viral attenuation, reversion to pathogenicity, and escape from vaccination are evolutionary phenomena based on the processes of viral adaptation, and correlated responses to adaptation. A mechanistic understanding of the molecular bases of viral host-specific adaptation and correlated responses to adaptation are therefore critical to effectively combating viral disease with live attenuated viral vaccines.

ϕX174 Background-

Bacteria and bacteriophage have been classical organisms for genetic studies since the 1940's when research on these organisms led to such important advances in genetics as the identification of DNA as the hereditary molecule (Hershey and Chase, 1952), the discovery of the genetic transformation of DNA (Avery *et al.*, 1944), and the identification that recombination occurs within genes

and creates novel genotypes (Benzer, 1955). This groundbreaking research established bacteriophage as model genetic systems.

Many of the same properties that make bacteria and bacteriophage well suited for genetics also makes them ideal for experimental evolutionary studies. Bacteriophage are simple genetic organisms. For example, bacteriophage ϕX174, the virus used in this research, is 5386 bases in length and contains nine essential genes (Hayashi *et al.*, 1988). Automated sequencing allows rapid nucleotide sequencing of the complete genome, thereby ensuring the identification of all mutations that evolved during selection. The rapid reproductive rate of bacteria and bacteriophage allow the examination of relatively "long term" evolutionary phenomena such as adaptation to novel environments (Lenski, 1988; Elena and Lenski, 1997; Turner and Chao, 1998). The very large population size and high mutation rates of many microbes rapidly generate large amounts of genetic variation upon which selection can act. This is particularly useful when starting from individual isolates or clonal populations. Moreover, the large population size provides the opportunity for selection to act not only on changes of large effect but also on changes with minor fitness effects. Finally, the ease with which microbes such as ϕX174 can be manipulated in the lab provides additional benefits for evolutionary studies. For example, the long term storage of phage isolates allows direct side by side fitness comparisons of ancestors and their evolved descendants, something which is not possible in most animal systems.

ϕX174 is a small icosahedral lytic bacteriophage (fig 1). Because of its small genome and relatively simple protein structure it has become a model system for examining DNA replication, gene expression, morphogenesis, and DNA-protein and protein-protein interactions (Hayashi *et al.*, 1988). ϕX174 is part of a group of related bacteriophages defined by their isometrical capsid containing a circular single stranded DNA molecule. In ϕX174 the genome is 5386 bases in length and contains nine essential genes which code for at least 11 different gene products, some of which are encoded by overlapping open reading frames (fig 2).

Molecular Basis of Host-Specific Viral Adaptation-

This research was inspired by the work of Bull *et al.* (1997) who examined the role of molecular convergence in independently evolved lineages of ϕX174. Though it was not formally addressed at the time, the data indicated that there were fitness and molecular differences associated with adaptation to each of the two hosts. Here I present the results of an experiment specifically designed to examine the molecular basis of host-specific viral adaptation. Replicate lineages of bacteriophage ϕX174 were alternately evolved on either one of two bacterial hosts. The two hosts used were *Escherichia coli* C, the typical laboratory host for ϕX174, and *Salmonella typhimurium*, a novel host. These two bacteria are closely related yet evolutionary quite distinct lineages. Although the occasional

importance of horizontal gene transfer between these bacteria cannot be discounted, numerous genetic and phylogenetic analyses support the predominately distinct evolutionary histories of these two bacterial lineages (Ochman and Wilson, 1987; Dubose *et al.*, 1988; Rayssiguier *et al.*, 1989; Sharp, 1991; Li *et al.*, 1995; Selander *et al.*, 1996). These two lineages diverged from a common ancestor roughly 120-160 million years ago, at a time concordant with the origin of the mammals, the predominant host lineage of *Escherichia coli* (Li *et al.*, 1995). Since their divergence Escherichia has evolved predominantly as a gut commensal of warm blooded vertebrates, whereas Salmonella has evolved primarily as a specialized intracellular parasite of reptilian lineages (Ochman and Wilson, 1987).

Replicate lineages were grown for approximately 1000 generations on each host, first on *S. typhimurium*, then for 1000 generations on *E. coli*, then back on *S. typhimurium* and finally again on *E. coli*. I sampled individual plaque isolates at the end of each period of adaptation, sequenced them in their entirety and measured their fitness on both hosts. The fitness of isolates from each of the endpoints was high (relative to ancestor) when measured on the host most recently selected on. However, correlated responses to host-specific adaptation differed dramatically between the two hosts. Adaptation to *S. typhimurium* resulted in the loss of all previous fitness gains on *E. coli*, whereas adaptation to *E. coli* did not affect fitness back on *S. typhimurium*.

Although nucleotide substitutions were observed at many sites throughout the experiment, relatively few of them, in one gene (the major capsid protein), explained the majority of the host specific fitness adaptation and of the correlated

fitness response. Furthermore, these host-specific changes in the viral capsid protein gene reverted back and forth between two states each associated with adaptation to one host. One particularly important substitution appears necessary, yet it alone is not sufficient to explain the observed patterns of fitness evolution. All host-specific changes in gene F result in amino acid replacements that are located on the outer surface of the virion.

This research represents a novel approach to both experimental evolutionary studies and to basic genetics. This is the first evolutionary study of its kind to identify the molecular basis of an adaptive fitness response, follow adaptive mutations as they sweep through evolving populations, and then examine the mechanistic fitness and morphological effects of these same adaptive mutations. This novel evolutionary approach has led to the discovery of amino acid positions which are important for host specific adaptation yet have not previously been identified through classic genetic studies as a part of the host receptor binding site, or even being involved in host interactions. The use of microbes such as ϕX174 in this case, then leads to new insights in the areas of molecular adaptation and molecular evolution, and the experimental evolutionary approach results in new insights into basic viral biology and genetics

MATERIALS AND METHODS

Phage and host lineages:

ϕX174 is a small icosahedral and lytic bacteriophage (fig. 1). The capsid contains a circular single stranded DNA genome 5386 bases long which encodes 9 essential and two non-essential genes (fig 2., Tessman and Tessman, 1978).

A wild-type isolate of ϕX174 was used as the ancestor for this study. The complete sequence of this isolate was found to differ from the published sequence (Sanger *et al.*, 1977; GenBank V01128) at five positions: 4784, I observed a T, 4518 an A, 2811 a T, 1650 a G, and at 833 an A ; Table 1). The two hosts utilized in this study were *Escherichia coli* C (restrictionless strain) and Type I restrictionless (hsd), ϕX174^{S} *Salmonella enterica* serovar Typhimurium, LT2 strain IJ750 (hereafter refered to as *S. typhimurium*)[*xyl-404 metA22 metE551 galE719 trpD2 ilv-452 hsdLT6 hsdSA29 hsdSB121 fla-66 rpsL120 H1-b H2-e nix*] [provided by M. M. Susskind to Ian Molineux (as MS3849)].

Chemostat:

Phage populations were evolved in a continuous culture, 2-chambered chemostat (fig. 3). Media (LB: 10g NaCl, 10g Bacto Tryptone, and 5g Yeast extract per liter, + 0.005% antifoam B, Sigma A5757) was pumped continuously into the first chamber which contained cells only. Excess suspension from this

cell chamber was drained continuously and unidirectionally into the 2nd chamber, which contained phage. Excess suspension from this phage chamber was drained off continuously. Aeration of both chambers was achieved with bubbling of filtered air through chamber one, drawn with negative pressure through chamber two and exhausted with excess liquid from this chamber. Both chambers consisted of 100 x 20 mm glass test tubes with stoppered tops.

The chemostat was sterilized and re-inoculated with cells from a frozen stock every two to three days. The most recent phage sample treated with chloroform (which kills cells but not phage) was used to inoculate the phage tube. Thus, any bacteria adapted to the chemostat or the phage were discarded but phage adaptation accumulated.

A flow rate of 6-10 ml/hr resulted in total turnover of the cells and phage in the phage tube of about 80-100 times per day. Phage samples were collected from the chemostat, mixed with chloroform, and saved for future analyses and/or used to inoculate the chemostat after sterilizing and changing tubes, media, and cells.

Experimental Design:

Phage populations, usually 10^7-10^8, were grown in the chemostat at high temperature (43.5°C) on either of two hosts, *Escherichia coli* C or *Salmonella typhimurium* (abbreviated C or S respectively, fig. 4). I used the relatively high temperature of 43.5°C because Bull *et al.* (1997), evolved the initial phage populations to this temperature; as it was (and is still) unknown whether the adaptive phenomena observed at 43.5°C occur at 37°C, I chose to continue the

experimental lineages at 43.5°C so that host was the only treatment variable. After a period of selection on a given host (11 days, except in one case noted below) a single plaque isolate was chosen for fitness and molecular analyses and utilized to inoculate the chemostat for the next period of selection on the alternate host. I therefore imposed a severe population bottleneck of N=1 at each host switching event. For each sub-lineage plaque isolate I measured fitness on both hosts and sequenced the entire genome (see below and fig. 4) . Starting with ancestor (A) on the S host, the phage population (generally 10^6-10^8 phage/ml) was gradually adapted to high temperature as described in Bull *et al.* (1997). After this initial period of selection on S, the phage culture was plated and a single plaque isolate (S1) was selected for analysis and used to inoculate two new replicate lineages propagated on host C (fig 4). After this selection on C, one plaque isolate was chosen from each replicate lineage (SC1 and SC2), saved for analyses, and used to seed the next round of selection back on S. After this final selection on S, plaque isolates (SCS1 and SCS2) were selected and utilized to inoculate the final period of selection back on host C, which resulted in isolates SCSC1 and SCSC2. Phage samples were removed daily from the chemostat during the final selections on S and on C for the population level fitness and molecular dynamics analyses (see below).

The experimental design created two replicate lineages sharing the initial selection period on Salmonella (A->S1) and followed by three independently derived periods of selection producing three independent sub-lineage isolates for each replicate (S1->SC1->SCS1->SCSC1, and S1->SC2->SCS2->SCSC2). Both

replicate lineages therefore share the two ancestral sub-lineage isolates, A and S1, and have three remaining independent sub-lineage isolates, SC, SCS and SCSC. All selection periods on a single host lasted 11 days except the S1->SC2 period of selection on C which lasted 22 days. The historical and temporal relationships, design, sampling regime, and additional details are depicted in Fig. 4.

Fitness assays:

Fitness assays were performed as phage growth rates (GR) over one hour, assayed on one host species in small (2.5 ml) volumes in 15 ml glass tubes. Host cells were grown in LB broth batch cultures at 43°C with shaking, to a density of ~ 2 x 10^8 cells / ml (measured with a Klett). 0.2 ml of these cells were added to individual tubes containing 2.5 mls LB. Phage were added to a final concentration of ~5 x 10^2/ ml (moi≈10^{-5}), thereby avoiding multiple infection of individual cells throughout most of the assay. 0.5ml of the phage infected inoculum was immediately removed from the tube, vortexed with chloroform, and titered for a t=0 value. The culture was then shaken for 60 min. in a 43°C water bath and a second (t=60) sample was taken and titered over chloroform.

Fitnesses were quantified as $\log_2$ of the ratio of phage concentration at t=60 divided by the phage concentration at t=0. For example, GR=10 is an increase of 2^{10} or a 1,000 fold phage increase per hour, whereas GR=20 is approximately a 1,000,000 fold phage increase per hour. Fitnesses were assayed both on populations derived directly from isolated plaques (individual fitness) and on populations derived from whole culture chemostat samples (population

fitness). These fitness estimates are simple measures of the growth capacity of an organism, they can be thought of as the number of phage doublings per hour. This growth rate component of fitness is affected by other fitness components such as attachment rate, rate of DNA entry, replicative rate within host cells, and final phage burst size. I do not attempt to disentangle these fitness components but instead prefer the population level "growth rate" approach which can be affected by changes in any of these other fitness components.

Some GR assays were performed with 0.25 mM $CaCl_2$ added to the LB. This increases GR somewhat but all GR comparisons were made within a single treatment so this does not affect any fitness comparisons made in this paper. Calculations show that as long as most host cells undergo a single binary division during the course of the assay uninfected cells will remain at the end of the GR assay in all isolates except for those with the highest growth rates achieved on *E. coli* C (e.g., SC2 and SCSC2). Nonetheless, infecting all cells in the culture prior to the end of the assay can only result in an underestimate of the actual GR of that isolate.

The total magnitude of fitness improvements differed between the two hosts. Growth rates on *E. coli* C (GR_C) often exceeded 20, whereas on growth rates on*S. typhimurium* (GR_S) rarely exceeded 13. This difference in GR between the two hosts appears to be primarily due to absolute differences in growth rates of the two hosts, though this difference was not quantifed. Nevertheless, all quantitative fitness comparisons are made between isolates measured on the same host, only qualitative comparisons are made between isolates measured on different hosts.

Sequence analysis:

Nucleotide sequences were determined by the dideoxy chain termination method using either single-stranded viral DNA or PCR products of viral DNA as template. Sequences of the A, S1 and SC isolates have previously been reported (Bull *et al.*, 1997) Any corrections from these previously reported sequences are reported in Table 1. Sequences were obtained from a core facility either at the University of Texas, or at the University of Idaho using ABI 377 automated sequencers. Sequences were compared to the published viral strand or its complement sequence (Sanger *et al.* 1977, GenBank V01128). Any deviations from this published sequence were verified for accuracy. Ambiguities were resequenced, if necessary utilizing a different primer or DNA prep, or the opposite strand for PCR products.

Oligo Screening Assays:

Population frequencies of substitutions were analyzed from samples of viral isolates and screened with radiolabeled probes to detect mismatches. Daily whole culture samples were plated at low density, and individual plaques were picked into 12 x 8 well microtiter plates. 16 plaques were chosen per day, allowing examination of up to five days on a single 96 well microtiter plate.

Phage from the microtiter plate were replica-plated onto 15 cm LB agar plates and grown 6-8 hours at 43°C. These plates were then blotted overnight

onto a nylon membrane in NaOH. After blotting, phage DNA was UV crosslinked to the membrane.

16-mer oligos were designed complementary to either the ancestral or evolved state for the sites of interest. Using TMAC buffer (3M tetramethylammonium chloride, 5X Denhardts solution, 0.1% SDS, 1mM EDTA, and 0.1M $NaPO_4$) a single base mismatch near the middle of the oligo suffeciently reduces binding that plaques can be scored unambiguously as matching the probe or not (Wichman, *et al.*, in prep.) For each hybridization reaction, one of the oligos was radiolableld and the other (the "competitor" oligo) was unlabeled. Oligos were end-labeled with $\gamma - {}^{32}P$ ATP in a polynucleotide kinase reaction at 37°C for 30 minutes (Sambrook *et al.*, 1989). Blots were pre-hybridized with 5% salmon sperm for 30 min. in 6M tetramethylammonium chloride (TMAC) buffer at 48 °C. Pre-hybridization buffer was removed and hybridization buffer (3M TMAC) was added. A 100:1 ratio of competitor probe to target probe (labeled) was added and the hybridization was incubated at 48°C for 2-8 hrs. All analyzed isolates were screened with both "evolved" and "ancestral" oligonucleotide probes to ensure correct genotyping. Resulting hybridized blots were washed in 2X SSC solution, wrapped in plastic food wrap, and exposed to film. Resulting autoradiographs were developed and analyzed.

Site Directed Mutagenesis

Site-directed mutants were generated by hybridizing mismatched mutagenic oligos (25mers) to single stranded ϕX174 DNA and extending this duplex with T4 polymerase (a protocol similar to Sambrook *et al.*, 1989; Fane, pers. com.). Resultant mixtures of single-stranded background, single-stranded mutagenic, and heteroduplex DNA's were transfected into electrocompetent cells via electroporation. Resulting transfectants were screened utilizing the same oligo screening assays described above. Positively probed site-directed mutant plaques were purified via replating, a second round of probing, and verification by sequencing the entire genome.

Location of Changes in Protein Structure:

I analyzed protein structure and location of host specific substitutions in the ϕX174 F and G protein unit, and in whole phage using RasMol v2.5 software and the published coordinates for the ϕX174 pentameric unit containing the F, G, and J proteins (Sayle, 1994; McKenna *et al.*, 1994).

RESULTS

Individual Analyses: Fitness and Correlated Molecular Changes

Individual fitness -

In contrast to the ancestor, A, all sub-lineage isolates evolved high fitness when measured on the host to which the isolate was most recently evolved (fig. 5). The initial fitness of the ancestral isolate was very low on either host at 43°C (fig. 5, GR_C= -0.28, GR_S= -0.77). Negative growth rates can result from phage infecting host cells without packaging phage by the end of the assay, resulting in a net phage decrease. After the initial period of selection on the *S. typhimurium* host (S), GR of isolate S1 measured on S had improved approximately 1000 fold (GR_S =9.38). All subsequent isolates (regardless of host selected on, see below) maintained at least this magnitude of GR_S or greater relative to ancestor (fig. 5). Plaque isolates of lineages most recently selected on the *E. coli* C host (SC1, SC2, SCSC1, and SCSC2) improved up to 950,000 fold in GR_C relative to both the initial ancestor, A and relative to their most recent S selected ancestors, S1 and SCS1 &SCS2 ($GR_C \approx$16-20, details in fig. 5).

Correlated fitness responses (GR_C of an isolate selected on S, and GR_S of an isolate selected on C) were highly host dependent. Selection on host S consistently resulted in the dramatic loss of fitness on host C (fig. 5). However, selection on C did not appreciably affect GR_S (details in fig. 5). This

asymmetrical host-dependent pattern of correlated fitness responses was repeatable across host shifts within a single replicate and between each replicate.

Despite the similarity between the two replicates of the pattern of fitness evolution, the absolute magnitude of fitness improvements did diverge significantly between the two replicates by the end of the experiment. Isolate SCSC2 had higher fitness on both hosts (GR_S=13.0±0.34 and GR_C=19.6±1.1) than did isolate SCSC1 (GR_S=10.1±0.2 and GR_C=15.5±0.59; t-test: $p<0.001$).

Molecular Analysis of Isolates-

I observed numerous nucleotide changes from ancestor in the sub-lineage plaque isolates. Table 1 lists all observed mutations from ancestor in the seven sub-lineage isolates. Summed across all isolates, 56 nucleotide sites evolved, consisting of greater than 1% of the ϕX174 genome. These sites occur in all ϕX174 genes except the small non-essential gene K. I also observed one insertion, one deletion, and relatively large magnitudes of parallelism, and reversions (tables 1 and 2).

The rate of molecular evolution was more rapid during the beginning of the experiment than later. The three earliest sub-lineage isolates S1, SC1, and SC2 all had rates greater than 1.2 nucleotide substitutions per day (S1=1.4, SC1=1.2, and SC2=1.3 substitutions per day). Conversely, the last four sub-lineage isolates all had substitution rates less than 0.75 nucleotide substitutions per day (SCS1=0.55, SCS2=0.36, SCSC1=0.73, and SCSC2=0.27 substitutions per day; t-test:p=0.002).

Host Correlated Nucleotide Reversions-

With the exception of the initial period of selection on S, the design manipulates host as the only experimental variable. Thus, I focus here primarily on nucleotide reversions correlated with host switches. I observed 11 nucleotide sites which, in at least one replicate, evolved to a new nucleotide on one host and then reverted back to the ancestral or previous nucleotide when subsequently grown on the alternative host (table 2). Some sites (e.g. 1305 in both replicates or 2009 in replicate #2) showed a perfect correlation between the two alternative nucleotides and selection on either host, evolving at all possible host switching opportunities (table 2). Other sites only showed this correlation at one or more of the six total reversion opportunities. Molecular patterns such as these host correlated reversions suggest that these substitutions are important for host-specific adaptation. Moreover the consistent involvement of particular sites (e.g., 1305) suggests their necessity for host adaptation.

There was a significant difference between the two replicate lineages in the total number of host specific molecular reversions. Ten such sites evolved in replicate 1, whereas only 4 sites evolved in replicate 2. Half of the host-specific reversions in replicate 1 and three of the four reversions in replicate 2 are found in gene F, the major capsid protein. For this reason, and because the capsid protein is known to have important interactions with the host cell membrane and the lippopolysaccharide (LPS) receptor (McKenna *et al.*, 1994), I focus the following molecular analyses primarily on these host-specific reversions in gene F.

Fitness Consequences of Selective Sweeps of Host Specific Substitutions

I will focus first on analyses of replicate #2, the results of which seem simple. These results will then be compared and contrasted with the more complicated results from replicate #1 which nonetheless support the general conclusions from replicate #2.

Molecular dynamics and basis of fitness evolution on E. coli C, Replicate #2-

Dramatic changes in population fitness on host C (GR_C) were observed during selection on S and during selection on C and were consistently correlated with selective sweeps of one or more of the host specific reversions in gene F (fig. 6) Between days 4 and 8 of selection on host S, selective sweeps of evolved nucleotides at sites 1305 and 2009 are correlated with a 92,000 fold reduction in GR_C (GR_C= 16.5±2.28 on day 4 to GR_C= 0.91±2.1 on day 8; t- test, $p<<0.001$; fig. 6).

There was a rapid initial fitness increase during selection back on *E. coli.* This initial 21,000 fold increase in GR_C (GR_C=1.4±1.3 in isolate SCS2 vs. GR_C= 14.4±0.91 on day 1; t-test, $p<0.001$) was correlated with the selective sweep of the back mutation to the ancestral nucleotide at site 1305. The second GR_C improvement during selection on C occurred between days 10 and 11 (GR_C=16.3±0.75 on day 10 vs. GR_C= 21.5±0.33 on day 11; t-test, $p<<0.001$) and is correlated with the sweep of the back mutation at site 2009 (fig. 6). Because all selections were derived from single plaque isolates, these nucleotide reversions

are true *de novo* back mutations, not simply changes initiated from an existing polymorphism.

These patterns clearly demonstrate the importance of changes at sites 1305 and 2009 for controlling both host specific adaptation to *E. coli* and the correlated response of fitness loss on *E. coli* during *S. typhimurium* adaptation. These results are not likely to be confounded by other molecular changes I did not assay for, since isolate SCS2 differed from isolate SC2 by only the four substitutions shown in fig. 6. SCSC2 also differed by only four changes from its previous ancestor SCS2, the two host-specific reversions at sites 1305 and 2009, and two additional changes not shown in fig. 6, a silent substitution at position 551 and a 2 base insertion into an intergenic region (table 1). Neither of these changes were present (at discernible frequencies) in the population prior to day 10 of selection on *E. coli.*

Molecular dynamics and basis of fitness evolution on E. coli C, Replicate #1-

Replicate #1 evolved twice as many molecular changes as replicate #2 during the period of selection for this dynamics analysis (table 1). This results in a more complex pattern of molecular evolution in this replicate. I choose only to examine the dynamics of the host-specific reversion sites from gene F (fig. 7).

During selection on S, a selective sweep at site 1305 is again correlated with the initial 280 thousand fold reduction in GR_C (GR_C=18.1±0.97 in isolate SC1 vs. GR_C= 11.7±0.98 on day 3 ; t-test, $p<<0.001$; fig. 7). The fixation of a second host specific gene F reversion site again appears important for final reductions in GR_C; in this replicate that second site is 1460 instead of site 2009 as

in replicate #2 (fig. 7). The change at site 1460 sweeps through the population between days 3 and 5 and is correlated with the final 3000 fold drop in GR_C to the level previously observed in the SCS1 plaque isolate (GR_C=11.7±0.98 on day 3 vs. GR_C= 0.372±0.53 on day 5; t-test, $p<0.001$).

Again as in replicate #2, there was a rapid initial increase in GR_C during selection back on *E. coli*. Selection back on host C results in a 2500 fold increase in GR_C (GR_C=0.089±1.7 in isolate SCS1 vs. GR_C= 11.1±0.46 on day 2; t-test, $p<0.001$), correlated with the selective sweep of the back mutation at site 1305, just as in replicate #2. The sweep of the host-specific gene F reversion at site 2093 may also play a role in this initial fitness increase, as this site had reverted in the population as a whole by day 4 of selection on C. However, the exact nature of the timing and rate of the selective sweep of the site 2093 reversion is unknown (fig. 7). The reversion at site 1460 sweeps back through the C selected population between days 7 and 9, but I did not observe and significant changes in GR_C associated with this sweep.

Fitness effects of host specific mutations: Individually and Together

Fitness effects of individual gene F host switching mutations were assayed in phage genotypes created by site-directed mutagenesis or in some cases in clonal phage isolates obtained during selection which I isolated from the evolving populations (all verified by complete genome sequencing, see methods). Site 1305 from gene F was chosen as the primary focus of this work because of its perfect correlation of reversions with host switching (table 2), and the tight

correlation between selective sweeps at this site and major changes in GR_C (figs. 6 and 7).

Fitness Effects of Changing only Site 1305-

In spite of site 1305's reverting at every host switching opportunity and its tight correlation with evolutionary changes in GR_C, the changes at site 1305 alone are clearly not sufficient to explain the entire pattern of fitness evolution on *E. coli* C (fig. 8). The substitution on the ancestral "G" nucleotide at site 1305 for the evolved or *S. typhimurium* specific "A" nucleotide (resulting in a Glycine to Aspartic Acid replacement) into the *E. coli* isolates of SC1 and SC2 resulted in reductions of GR_C from 21.2±0.39 to 15.8±0.12 in replicate #1 and from 22.5±0.24 to 12.17±0.29 in replicate #2 (2.3 million and 5.8 million fold reductions respectively, fig 8).

Substituting the evolved "A" at 1305 with the ancestral or *E. coli* specific "G" into the *S. typhimurium* adapted isolates of SCS1 and SCS2 increased GR_C variably, depending upon which replicate genetic background the change was inserted into (fig. 8). Changing site 1305 to the *E. coli* evolved state in the genetic background of isolate SCS1 resulted in an increase in GR_C from -0.29±0.32 to 4.68±0.58 (only a 30 fold increase in GR_C; t-test, $p<0.01$; fig 8). However, inserting the same change into the genetic background of replicate #2 resulted in the much greater increase in GR_C from 2.25±0.98 to 16.1±0.46 (a 70,000 fold increase in GR_C; t-test, $p<0.001$; fig. 8). Since the replicates were initiated from the same S1 isolate, the difference between site 1305's effect on

GR_C in the two replicates can only be due to epistatic interactions between site 1305 and other sites which diverged during the course of this study.

Fitness Effects of site 1305 plus others in Replicate #2-

Isolate SCS2 contained only four substitutions from its ancestor SC2 (C->T at position 593, A->G at 921, G->A at 1305, and T->A at 2009, table 2), yet it exhibited the typical correlated fitness response to selection on S: its GR_C plummeted over 1 million fold during selection on S. Figure 9 shows the effects of these substitutions on GR_C. Adding the evolved "T" at site 593 and the A->G reversion of site 921 into the SC2 genetic background reduces GR_C from 22.5±0.24 to 16.7±0.45 . Adding just the evolved "G" at site 1305 into the same SC2 genetic background reduces GR_C to 12.17±0.29. Together these three substitutions inserted into the SC2 background appear to act additively reducing GR_C to 6.71±0.16. Finally addition of the evolved T->A change at site 2009 reduces GR_C to the observed levels in plaque isolate SCS2 (GR_C=2.25±0.98, fig. 9).

Isolate SCSC2 also contained 4 changes from its SCS2 ancestor (C->T at 551, the A->G back mutation at 1305, the A->T back mutation at 2009, and a GT insertion into an intergenic region between gene H and gene A, table 1). Together these four mutations (only the two gene F changes result in amino acid replacements) increased GR_C over 9 million fold from 2.25±0.98 to 23.2±0.39 (fig. 9).

Effects of site 1305 plus others in Replicate #1 -

Replicate #1 appeared qualitatively similar to replicate #2 though 1305 has less of an overall effect on GR_C in this genetic background (fig. 8 and fig. 10). Six nucleotide substitutions differentiate isolate SC1 from SCS1 (table 1). Adding the evolved site 1305 into the SC1 background reduced GR_C from 21.2±0.39 to 15.8±0.12. Adding the evolved changes at sites 1460 (a gene F host-specific reversion site) and 815 in addition to 1305 into that same SC1 background reduced GR_C to 1.59±0.96 (fig. 10), greater than a 2 million fold reduction in GR_C compared to SC1 and only a two fold greater value of GR_C than in isolate SCS1 (GR_C = -0.29±0.32).

Eight substitutions differentiate isolate SCSC1 from SCS1 (table 1). We previously saw that the effect of adding the ancestral 1305 site into the SCS1 background increased GR_C only 30 fold (fig 8). Yet changing 1305 together with a second gene F reversion site, 2093 improves GR_C over 89 thousand fold (GR_C= 16.1±0.57; fig. 10). These results are consistent with the findings from replicate #2 where two amino acid replacements in the F protein (caused by substitution and subsequent reversion at sites 1305 and 2009) cause most (in SCS2) if not all (in SCSC2) of the changes in GR_C(Fig. 9) In replicate #1 amino acid changes in the F protein resulting from host-specific substitutions at sites 1305 and 1460 appear to be important for reducing GR_C during selection on *S. typhimurium*, and reversions at sites 1305 and 2093 together explain most of the recovery of GR_C during adaptation to *E. coli* (fig. 10).

Protein Structure-

The above results collectively demonstrate the importance of the five gene F host-specific reversions at positions 1305, 1460, 2009, 2093, and 2167, corresponding to major capsid protein (gene F) residues 101, 153, 336, 364, and 388 respectively.

The ϕX174 capsid is composed of 60 copies each of the major capsid protein (gpF) and major spike protein (gpG), and twelve copies of the minor spike protein (gpH). F protein and G protein form pentameric units surrounding each icosahedral vertex. G protein sits atop F protein and forms a "collar" at each vertex, with the H protein sitting inside this collar (fig. 11, McKenna *et al.*, 1994). Hence much of the outer "surface" of the F protein is actually covered by the G and/or H proteins. Nonetheless, all five gene F host-specific reverting residues are located on the major capsid protein's outer surface and none are covered by the placement of protein G (fig. 12). Hence, all gene F host-specific residues are presumably exposed to the host cell surface.

Interestingly, none of these gene F changes occur in the putative host receptor binding site (McKenna *et al.*, 1994). Only one of these five gene F reversion sites is involved in known intra- or inter-molecular interactions, this is site 2093 which plays a minor role in a 5-fold H_20 mediated polar F-F interactions (McKenna *et al.*, 1994). The general lack of these host-specific gene F sites involvement with intra- and inter-molecular phage interactions is consistent with their predominant involvement in phage-host.

DISCUSSION

This study represents a novel approach for experimental evolutionary research. I have elucidated the molecular basis of a complex host-specific fitness adaptation and of the correlated fitness response to this adaptation. In spite of relatively large numbers of nucleotide substitutions that accumulated during the experiment, the majority of the adaptive fitness response can be explained by a few nucleotide substitutions in the major capsid protein gene. These host-specific substitutions show a dramatic pattern of evolution and subsequent reversion correlated with adaptation to the two hosts (table 2). Furthermore, selective sweeps at these host-specific nucleotide sites during adaptation are correlated with evolutionary changes in population fitness. One of these host-specific substitutions appears necessary for adaptation, yet it alone is not sufficient to explain the observed patterns of fitness evolution. Finally, I show that these host-specific residues are located on the surface of the capsid protein and hence are presumably exposed to the host cell surface.

Host-Specific Adaptation and Correlated Responses to Adaptation

Host-Specific Adaptation-

All end-point plaque isolates evolved high fitness relative to ancestor when measured on the host the isolates were most recently selected on. This improvement in fitness can be attributed to the nucleotide substitutions identified, since the entire genome of all isolates was sequenced. Yet the exact fitness

contribution of most substitutions are unknown, as are the bases of their selective advantage. There are at least three obvious and different selective pressures during this experiment, high temperature, growth in the chemostat, and host. Because chemostat and host have been held constant since the original lineage leading to S1, I expect that the bulk of fitness changes since S1 have been in response to host-mediated selection. Because host was the only manipulated variable, substitutions due to host-mediated selection should accumulate throughout the experiment whereas substitutions in response to high temperature and growth in the chemostat should accumulate early in the experiment.

Two lines of evidence support the above hypothesis. First, the total rate of molecular evolution was greater early in the experiment than later. The predominate role of host-mediated selection throughout the experiment vs. temperature, chemostat, and host-mediated selection together early in the experiment is supported by an overall decline in the rate of molecular evolution. The substitution rate was greater in the initial periods of selection producing the S and SC isolates (all greater than 1.2 substitutions per day) than it was in the later selection periods producing the SC. and SCSC isolates (all less than 0.75 substitutions per day).

Second, known host-specific changes (table 2) account for a much larger percentage of the total substitutions later than they do of the total early substitutions, and mutations known to be advantageous at high temperature accumulated early in the experiment and not later. Badgett *et al.*, (unpublished results) has identified three different high temperature substitutions in ϕX174 that

confer high fitness on *E. coli* , and four that confer high fitness on *S. typhimurium.* All of these changes improve plaque growth at 45°C either by improving growth rate or phage stability at high temperature. Three of these changes occured in this experiment. Badgett *et al.*, found that an A to T change at position 1613 is associated with high temperature growth on *S. typhimurium.* Accordingly, this same substitution appears during the initial selection on *S. typhimurium* producing the S1 isolate and is maintained throughout the rest of the selections (table 1). Badgett also found two other high temperature substitutions associated with selection on *E. coli* C that occurred in this study. The C to T substitution at site 1727 dramatically improves both phage stability and growth rate at high temperature, and this substitution evolved during the initial selection on *E. coli* C in isolate SC2 (table 1). The A to G substitution at site 1565 does not improve stability but does improve growth rate at high temperature on *E. coli* C. This substitution also occurred during the initial selection on *E. coli C* in isolate SC1 (table 1). These patterns and data support the predominate role of host-mediated selection throughout the experiment, but with additional selection pressures, e.g., high temperature initially.

Correlated Responses to Host-Specific Adaptation-

ϕX174 was originally isolated from an *E. coli* host (Sinsheimer, 1959) and is typically propagated on *E. coli* in the lab. ϕX174 can only infect rough strains of *S. typhimurium* (Hayashi *et al.*, 1988) and the ability of ϕX174 to grow

efficiently on *S. typhimurium* probably represents a novel phenotype for this virus. Selection for novel phenotypes commonly produce correlated responses to selection; antagonistic pleiotropy is a widely accepted mechanism for such correlated responses (Caspari, 1952; Wright, 1968; Lenski, 1988; Novella *et al.*, 1995). Antagonistic pleiotropic fitness effects can result in constraints and trade-offs either between different fitness components (e.g. growth rate and competitive ability) and/or between fitness in different environments.

I observed a dramatic host asymmetry in the correlated responses to host-specific adaptation. Whereas adaptation to *Salmonella* consistently resulted in the loss of all previous fitness gains on *E. coli,* adaptation to *E. coli* did not affect fitness on *Salmonella* (fig. 5). This is an excellent example of an antagonistic pleiotropic constraint on fitness in a different environment, the *E. coli* host (Caspari, 1952; Wright, 1968; and Lenski, 1988). Additionally, I have identified the mechanistic basis of this pleiotropic constraint. Specific amino acid replacements in the major capsid protein (fig. 12, and below) selected during adaptation to *Salmonella* cause a reduction in fitness (measured as growth rate) on *E. coli.*

There are at least two obvious molecular mechanisms by which fitness reductions produced by antagonistic pleiotropy could be recovered. First, selection could favor modifiers of the pleiotropic mutations which cause the correlated fitness reduction. These modifiers could result in recovering fitness on *E. coli* without affecting the originally selected trait, high fitness on *S. typhimurium* (Lenski, 1988). Alternatively, the pleiotropic mutations which produced the correlated fitness constraints could simply undergo reversion (e.g.,

back mutation), thereby recovering fitness directly. If the recovery of a pleiotropic fitness reductions is due to reversion, it should result in loss of the originally selected phenotype, since the pleiotropic mutations producing that phenotype have now reverted back to their original state. In this study, selection on *S. typhimurium* results in the loss of fitness on *E. coli.* Subsequent adaptation to *E. coli* results in dramatic improvements in fitness on *E. coli* without major fitness effects on *S. typhimurium.* Without knowledge of the underlying genetic basis one might reasonably infer that the recovery of fitness on *E. coli* was due to epistatic modifiers of the original pleiotropic mutations which increased fitness on *E. coli* without affecting fitness on *S. typhimurium.* However, the recovery of fitness on *E. coli* is due predominantly to the reversion of host specific changes in gene F selected during *S. typhimurium* adaptation (figs. 9 and 10). Epistatic modifiers might additionally be involved (particularly in replicate #1) but their effects on GR_C appear to be minimal.

Why does there appear to be no loss in fitness on S. typhimurium during adaptation to *E. coli* ? There may well be correlated fitness responses that I am not seeing because I only quantify one component of fitness, growth rate. Growth rate is affected by other fitness components such as rate and efficiency of attachment, rate of DNA injection into host cells, the latency period within a host, and the number of phage progeny released from an infected cell. Changes in those fitness components could go unnoticed either because they minimally affect growth rate or because changes in two or more of these other fitness components could cancel out each's individual effects on growth rate.

Molecular Basis of Host-Specific Viral Adaptation

Host specific amino acid reversions-

I identified five amino acid substitutions in the major capsid protein (gene F) involved in controlling host specific adaptation. Together they show a dramatic pattern of replacement and subsequent reversion correlated with host shifts (table 2). Each of these amino acid positions alternated between two different residues, associated with adaptation on a single host. Some amino acids (F101 in both replicates, F336 in replicate #2) changed in parallel with every host switching event, suggesting their necessity for host specific adaptation. In this experiment, amino acid changes at these positions might be necessary for host-specific adaptation, but previously published data by Bull *et al.* (1997), demonstrate that the G to A substitution at site 1305 is not necessary to achieve high fitness on *S. typhimurium.* Bull *et al.*, evolved three independent lineages of ϕX174 to *S. typhimurium* in the chemostat. All evolved relatively high fitness on *S. typhimurium*, yet one of these three lineages (S3), did not evolve the G to A substitution at site 1305 (However, it did evolve host-specific substitutions at 2 of the remaining 4 gene F sites from this study, sites 2009 and 2167).

Two or three of the host-specific changes in gene F were found in any single evolved phage isolate (table 2). This is consistent with both other research on ϕX174 adaptation to *S. typhimurium* (Bull *et al.*, 1997; Wichman *et al.*, in prep.). This pattern of consistent host correlated substitution and subsequent reversion of relatively few amino acids is very strong evidence for the adaptive

significance of these changes. Moreover, it indicates that either, there are limited adaptive pathways for improving host fitness (few "adaptive peaks"), or that these lineages are "trapped" within a single fitness peak on the adaptive landscape, or both.

An interesting question is whether the observed host-specific nucleotide substitutions and reversions from gene F are the only substitutions which achieve the observed amino acid replacements or if there are other single base substitutions which result in the same amino acid replacement. If there are other substitutions which produce the same amino acid replacements yet they are not observed, that would suggest that there might actually be a mutational bias affecting the observed spectrum of substitutions in addition to their phenotypic effects on amino acid replacements or effects on mRNA/DNA structure. The substitutions at sites 1305, 1460, and 2009 are all either first or second position codon substitutions and those nucleotide substitutions are the only way to achieve the observed amino acid replacements. Site 2093 is also a first position in a codon and the observed C to T transition during selection on *S. typhimurium* is the only substitution which produces the observed leucine to phenylalanine replacement. Yet the change back to leucine from phenylalanine during selection on *E. coli*, which occurred by the T to C reversion at site 2093 could also have occurred by a T to A or a T to G transversion at site 2095. As expected, transitions are more common than transversions in this data set (table 1) so the alternative transversions at site 2095 might not be very probable. Since this amino acid (F364) only evolved and subsequently reverted once (in isolates SCS1 and SCSC1, table 2) it is difficult to discern the role of mutational biases anyhow.

Finally, there is the T to G transversion at site 2167 which occurs in the third position of a codon and causes a histidine to glutamine replacement. A transversion of T to A at site 2167 could also cause this same amino acid replacement, but again the original T to G transversion only occurred once in the S1 isolate (table 2) so this site is not useful to attempt to determine the role of mutational bias in producing the observed spectrum of nucleotide substitutions.

Further support for the adaptive significance of these host-specific reversions in gene F is the correlation between these substitutions sweeping in and out of the populations and significant evolutionary changes in GR_C (figs. 6 and 7). This temporal pattern of change in substitution frequency is compelling evidence for molecular adaptation only during selection on *E. coli* when the gene F replacements are presumably the direct targets of selection. During selection on *S. typhimurium,* selective sweeps of host-specific changes in gene F are correlated with decreases in GR_C. These decreases in GR_C cannot be under direct selection since the phage are evolving in the chemostat on *S. typhimurium*, instead they must be due to the antagonistic pleiotropic effects of the fixation of host-specific substitutions selected during adaptation to *S. typhimurium.*

How can these host-specific substitutions in gene F rapidly sweep through the populations evolving on *S.typhimurium* and not affect GR_S? Recall that GR is only one component of fitness, and it may not be highly correlated to that aspect of fitness which is directly selected in the chemostat. Clearly GR_C during chemostat growth on *E. coli* is either under direct selection or it is tightly correlated with whatever fitness component is directly selected. But GR_S during adaptation to *S.typhimurium* in the chemostat is probably not the sole target of

selection. The magnitude of change in GR_S is not great enough to explain the rapid sweeps and hence large selection coefficients that must be associated with these host specific substitutions in gene F. Instead, some component of fitness other than GR_S is probably directly selected during adaptation to *S. typhimurium.* Two possible explanations for why GR_S is not affected are that the directly selected fitness components during adaptation to *S. typhimurium* have minimal effects on GR_S or that changes in more than one of these other fitness components might cancel out each others' effects on GR_S.

In addition to fitness components which directly affect GR such as adsorption, latency period, and burst size (discussed above) other fitness components such as intracellular-competitive ability might be under selection in this experiment. Recent work by Turner and Chao (1998) demonstrates that estimates of viral fitness vary depending on whether host cells are infected by individual or multiple viral particles (the multiplicity of infection, m.o.i.). Unpublished work by Wichman *et al.*, (in prep.) demonstrates varying growth rates for chemostat selected lineages of ϕX174 depending on the multiplicity of infection (m.o.i.) at which they were assayed.

Effects of host specific reversions on protein structure -

Three structural proteins make up the mature phage particle: F, G, and H, the capsid, spike, and pilot proteins respectively. Twelve pentameric units of gpF form the planar capsid surface surrounding each vertex and encapsulating the DNA genome (fig. 11). One copy of gpH surrounded by five copies of gpG form

the attachment spikes which lie on top of gpF, at each of the 12 icosahedral capsid vertices (Fig 11).

All five of the major capsid protein (gpF) host-specific residues are located on the capsid protein's exterior surface. They are not covered by the spike protein, yet form a loose ring around the gpG/H spike protein unit at each vertex (figs. 11 and 12). ϕX174 binds to lipopolysachariде (LPS) in the outer membrane of both *E. coli* and *S. typhimurium*. The core polysacharide of the LPS receptor from *E. coli* C and *S. typhimurium* share structural affinites such as the presence of a terminal glucose and a side chain galactose, both absolutely necesssary for ϕX174 binding (Hayashi *et al.*, 1988). The core polysacharides of these two hosts also differ primarily by the presence of an N-acetylglucosamine at the non-reducing end of the core polysacharide in *S. typhimurium* and its absence in *E.coli* C (Hayashi *et al.*, 1988). The removal of this residue from from *S. typhimurium* LPS dramatically reduces bindng effeciency. Possibly these host-specific residues in the major capsid protein interact with the LPS or some component of it which differs between the two hosts. Unfortunately the exact details and requirements of the complete phage-host receptor interactions are not well understood (Hayashi *et al.*, 1988). Moreover, the host-switching residues discovered in this study are not a part of the putative carbohydrate binding pocket of the capsid protein involved in host LPS receptor binding (McKenna *et al.*, 1994), again though, the specific details of these interactions are unknown. It is known that 2-3 spikes of the virion become imbedded in the cell wall, and the phage becomes submerged within the host cell wall up to one half its diameter

(Bayer and Starkey, 1972). There are therefore presumably numerous opportunities for these host switching residues which form loose rings around each spike, to have interactions with any number of host membrane molecules outside of the initial LPS receptor - phage carbohydrate binding pocket interactions.

Previous genetic work on ϕX174 has mapped host range mutants to genes F, G, and H (Hayashi *et al.*, 1988). Nevertheless, the host-switching capsid protein residues identified in this study have not previously been identified as important for host binding or host-specific interactions. The identification of these residues' involvement in efficient host-specific growth provides new data upon which future genetic analysis can build. The results from this study therefore demonstrate the benefit of using an experimental evolutionary approach to attempting to understand a problem, and how the results from such a study can complement and add to the body of basic genetics research.

Implications for Vaccine Production

The construction of live attenuated viruses for vaccines has historically been a haphazard procedure, moreover the molecular bases of the attenuated viral phenotypes were rarely known at the time of their construction and/or application (Sabin and Boulger, 1973; Westrop *et al.*, 1989; reviewed in Bull, 1994). Attenuation in the target host or tissue is generally achieved by selection and subsequent adaptation to novel hosts, tissues, or environmental conditions such as temperature. Viral attenuation is therefore an antagonistic pleiotropic response to

adaptation under novel selection, and an understanding of the molecular basis of such pleiotropic fitness constraints and subsequent fitness recovery should allow for the improvement of live attenuated viral vaccine design.

The results of this work have important implications for the production and application of attenuated viral vaccines. In accordance with other work (Novella *et al.*, 1995; Elena *et al*, 1998) it is clear that given the appropriate selective environment, severe debilitation in fitness on a particular host (attenuation) can rapidly, repeatably, and predictably be recovered (e.g. rapid initial improvements of GR_C during selection on C, figs. 6 and 7). Such rapid recovery of fitness is problematic for attenuated vaccines because, if they revert to pathogenecity, they can cause disease in the vaccinated individual, and because they remain infective, they can be readily spread to unvaccinated individuals. Both of these phenomena for example, occurred at appreciable frequencies with the Sabin poliovirus vaccine (Sabin and Boulger, 1973).

This work demonstrates that single nucleotide substitutions or even groups of such substitutions which cause dramatic host specific attenuation not only revert readily, but that these simple back mutations rescuing fitness might actually be expected given the correct selective environment and ecological conditions. An understanding of the molecular basis of fitness constraints producing attenuated viral phenotypes should provide the necessary information to actually target and design "smart" vaccines whose attenuated phenotypes are more difficult to revert.

Future Directions:

The continued elucidation of the specific molecular mechanism of the host-specific residues' differential interactions with the two hosts is a tractable research goal. One approach would be to use x-ray crystallography to examine the adapted viruses bound to whole cells, isolated membrane fractions, or isolated LPS of the two different hosts to examine the roles of these different hierarchical cellular structures in phage binding.

Substitutions at these host-specific sites in gene F result in dramatic changes in phage growth rates on *E. coli*. As mentioned previously GR is only one of many fitness components. Four additional fitness components, any of which can affect GR are; rate and efficiency of host cell attachment and DNA entry, the latency period or time from infection to lysis, and the burst size or the number of phage progeny released from a single infected cell. These different components are all individually quantifiable, and I am presently planning to disentangle these different fitness components and examine how they affect growth rate. A mechanistic understanding of how each of these components affects the observed evolutionary changes in growth rate will additionally provide hypotheses for the specific molecular mechanisms by which these major capsid protein residues control host-specific adaptation.

There are additional future research directions less directly related to the results if this study. For example, what would occur if phage were grown in a chemostat with both hosts present at the same time? We might predict the evolution of two different viral populations each adapted to one of the hosts, or if the results from this study are applicable, we might expect to see the evolution of

a phage similar to the *E. coli* adapted isolates in this study which grew well on both hosts. Unfortunately there are technical problems associated with a two-host chemostat system which we are presently trying to overcome.

There is some suggestion from the recent study that the time period of adaptation to a given host is important. In this study all host specific selections lasted 11 days except for one selection on *E. coli* C which lasted 22 days, and produced the SC1 sub-lineage isolate. This isolate had many more substitutions that evolved during selection than any other isolate, though the substitution rate was similar to other early isolates (table 1). I hypothesize that the additional period of selection allows for the accumulation of additional host-specific substitutions which "fine tune" or modify previous substitutions (such as the host-specific changes in gene F) improving fitness on *E. coli.* These modifying substitutions do not affect growth rate appreciably (note the lack of a difference in GR_C between SC1 and SC2) but make subsequent host specific adaption to *S. typhimurium* more difficult because of the more complex molecular basis of the previous adaptation to *E. coli.* Hence, there are more changes to "fix" during selection on the laternative host than if the period of adaptation was shorter. Obviously, this experiment was not designed to examine this question, but the continued higher substitution rate in the SCS1 and SCSC1 isolates (relative to SCS2 and SCSC2, table 1), and the significantly lower GR_C for isolate SCSC1 vs. isolate SCSC2 (fig. 5) support this hypothesis. Experiments are already underway to examine this hypothesis specifically.

Finally this study raises some interesting questions about speciation. How many changes would it take during adaptation to *S. typhimurium* for example, to

make recovery of fitness on *E. coli* improbable or impossible? Could we evolve a phage in isolation on *S. typhimurium* which will not revert to efficient growth on *E. coli* ? Would our results and conclusions from such an experiment change if the initial adaptation occurred in the presence of both hosts? These are all tractable questions, the answers of which are interesting and important to a large audience in ecology, evolution, and population biology in general.

Table 1. List of all substitutions for the 7 evolved sub-lineage isolates and ancestral ϕX174 isolate. The first column identifies the nucleotide position as defined by Sanger *et al.* (1977). The second column indicates the number of independent changes observed at the site, the third column indicates the gene(s) or genetic location of the substitution (- - , indicates an intergenic region of DNA). The fourth column indicates whether the substitution is silent or missense in the gene it's located in, and the fifth column indicates the amino acid replacement for missense substitutions (genes A and A* are in the same reading frame). Columns 6-14 give the bases at each position in the seven evolved sub-lineage isolates and ancestor. Note: both replicate lines shared the common S1 isolate during their selection, so S1 is represented in two separate columns for ease of interpretation. Lower case bases are unchanged from the wild type ancestor (column 6). Any substitutions from the ancestral sequence are shown in capitols. A bold capitol letter is used to indicate that a base substitution occurred during the selection leading to that isolate, whereas a non-bold capitol letter indicates that the base substitution occurred prior to the selection period producing that isolate. For example, site 3932 in an intergenic region between gene H and gene A, was ancestrally an "A". During the period of selection on *E. coli* producing the SC1 sub-lineage isolate, site 3932 evolved to "G" and remained a "G" throughout the next two selection periods in replicate #1.

In addition to substitutions, a 27 base deletion was observed from site 965 to 991 (removing evolved site 986 in SC1) and a 2 base insertion was observed between position 3968 and 3969 in isolate SCSC2. For each sub-lineage isolate

the total number of new mutations evolved during the previous selection are shown in the first row below the table and the accumulated total number of changes from the original wild type ancestor are shown in the second row below the table.

Site	# Δ's	Gene	s/m	Amino Acid	A	S1	SC1	SCS 1	SC SC1	S1	SC2	SCS 2	SCSC 2
3932	1	–	–	–	a	a	G	G	G	a	a	a	a
3968 +	1	–	–	–		–	–	–	–	–	–	–	2 bp insert
3974	1	–	–	–	c	c	T	T	T	c	c	c	c
4110	1	A	m	H -> Y	c	T	T	T	T	T	T	T	T
4122	1	A	m	D -> N	g	A	A	A	A	A	A	A	A
4168	1	A	m	Q -> R	a	G	G	G	G	G	G	G	G
4420	2	A	m	D -> G	a	C	A	A	A	C	C	C	C
4700	2	A,A*	m,m	N -> K	t	G	T	T	T	G	G	G	G
4804	2	A,A*	m,m	T -> I,L	c	c	A	A	A	c	c	c	c
4805	2	A,A*	s,s	–	a	a	G	A	A	a	a	a	a
5341	1	A,A*, B	m,m, m	A: K->I B:Q->H	a	a	a	a	T	a	a	a	a
5365	1	A,A*, B	m,m,s	M > T	t	C	C	C	C	C	C	C	C
22	1	A,A*, B	s,s,m	A > S	g	g	T	T	T	g	g	g	g
31	1	A,A*, B	s,s,m	N > E	a	a	G	G	G	a	a	a	a
319	2	C	m	V -> F	g	T	G	G	G	T	T	T	T
323	1	C	m	D -> G	a	a	G	G	G	a	a	a	a
384	1	C	m	K -> N	g	g	T	T	T	g	g	g	g
396	1	D	m	Q -> E	c	c	c	c	c	c	G	G	G
500	2	D	s	–	g	A	G	G	G	A	A	A	A
519	1	D	m	I -> V	a	a	a	a	G	a	a	a	a
533	1	D	s	–	c	c	c	c	c	c	T	T	T
551	1	D	s	–	c	c	c	c	c	c	c	c	T
593	1	D,E	s,m	T -> I	c	c	c	c	c	c	c	T	T
756	1	D,E	m,s	F -> L	t	C	C	C	C	C	C	C	C
795	2	D,E	m,s	T -> A	a	a	G	G	G	a	G	G	G
815	1	D,E	s,m	C -> S	g	g	g	C	C	g	g	g	g
905	1	J	m	G -> S	g	g	A	A	A	g	g	g	g
921	2	J	m	R -> H	g	g	g	g	g	g	A	G	G
986	1	–	–	–	a	G	27bp del.	27bp del.	27bp del.	G	G	G	G
1025	1	F	m	K > R	g	A	A	A	A	A	A	A	A
1126	1	F	s	–	c	c	T	T	T	c	c	c	c
1305	7	F	m	G -> D	g	A	G	A	G	A	G	A	G
1460	2	F	m	Q -> E	c	c	c	G	C	c	c	c	c
1565	1	F	m	T -> A	a	a	G	G	G	a	a	a	a
1613	1	F	m	T -> S	a	T	T	T	T	T	T	T	T
1727	1	F	m	L -> F	c	c	c	c	c	c	T	T	T
2009	5	F	m	S -> T	t	A	T	T	T	A	T	A	T
2093	2	F	m	L -> F	c	c	c	T	C	c	c	c	c
2167	3	F	m	H -> Q	t	G	T	T	T	G	T	T	T
2591	1	G	m	V -> A	t	t	t	t	t	t	C	C	C
2949	1	H	m	G -> S	g	g	A	A	A	g	g	g	g
2980	1	H	m	A -> V	c	c	c	c	c	c	T	T	T
3013	1	H	m	A -> V	c	c	T	T	T	c	c	c	c
3042	1	H	m	L -> F	c	c	T	T	T	c	c	c	c
3055	1	H	m	N ->S	a	a	a	a	G	a	a	a	a
3060	1	H	m	T->I	a	a	G	G	G	a	a	a	a
3061	1	H	m	T -> I	c	c	T	T	T	c	c	c	c
3111	1	H	m	V -> I	g	g	g	A	A	g	g	g	g
3120	1	H	m	P -> S	c	c	c	c	c	c	T	T	T

3154	1	H	m	A -> V	c	c	c	c	T	c	c	c	c
3189	1	H	m	T -> A	a	a	a	a	G	a	a	a	a
3337	2	H	m	A -> V	c	c	T	T	T	c	T	T	T
3378	2	H	m	N -> H	a	a	C	C	C	a	C	C	C
3584	1	H	m	N -> K	t	t	t	t	t	t	G	G	G
3826	1	H	m	H -> R	a	a	G	G	G	a	a	a	a
3873	1	H	m	D -> N	g	g	g	g	g	g	A	A	A
Total # of new changes / lineage :						**15**	**28**	**6**	**8**	**15**	**14**	**3**	**4**
Total # of changes/lineage from A:						**15**	**27**	**29**	**33**	**15**	**23**	**26**	**25**

Table 2. Nucleotide positions that revert to ancestral or previous states according to host. The first column identifies the nucleotide position as defined by Sanger *et al.* (1977). The second column indicates the number of independent reversions observed at any one site (only 6 possible), the third column indicates the gene or genetic location of the substitution (an - represents an intergenic location). The fourth column indicates whether the substitution is silent or missense and shows the affected amino acid in the protein for missense changes. The fifth column indicates the amino acid replacement for missense substitutions. Columns 6-14 give the bases at each position in the seven evolved sub-lineage isolates and ancestor. Note: both replicate lines shared the common S1 isolate during their selection, so S1 is represented in two separate columns for ease of interpretation. Lower case bases are unchanged from the wild type ancestor, A (column 6). Any substitutions from the ancestral sequence are shown in capitols. A bold capitol letter is used to indicate that a base substitution occurred during the selection leading to that isolate, whereas a non-bold capitol letter indicates that the base substitution occurred prior to the selection period producing that isolate.

For example, site 1305 occurs in gene F. There were 6 independent reversions at this location summed across all sub-lineage isolates. The substitution of the ancestral "G" for the evolved "A" in isolate S1 results in an amino acid replacement at position 101 of aspartic acid for the ancestral glycine. During subsequent selection on *E.coli* both isolates SC1 and SC2 independently evolved reversions at site 1305 back to the ancestral "G". These are the first two reversions to occur at site 1305. Subsequent selection on *S. typhimurium* producing isolates SCS1 and SCS2 resulted in the independent reversion in both

replicates back to the evolved "A" at site 1305. The final selection on *E. coli* producing the sub-lineage isolates SCSC1 and SCSC2 resulted in the fifth and sixth independent reversions back to the ancestral "G" at this site. Thus site 1305 evolves and reverts perfectly at every host switching opportunity.

The total number of reversions for each sub-lineage isolate are shown in the bottom row below the table. Note that 10 sites experience reversals in replicate #1, whereas only four sites revert in replicate #2. Half of all reverting sites in replicate #1, and three of four in replicate #2 occur in the major capsid (gene F) protein.

Site	# Rev	Gene	A.A. #	Amino Acid	A	S1	SC1	SCS 1	SCSC 1	S1	SC2	SCS 2	SCSC 2
4420	1	A	A147	D -> G	a	**C**	**A**	A	A	**C**	C	C	C
4700	1	A,A*	A240 A* 68	N -> K	t	**G**	**T**	T	T	**G**	G	G	G
4805	1	A,A*	s,s	--	a	a	**G**	**A**	A	a	a	a	a
319	1	C	C63	V -> F	g	**T**	**G**	G	G	**T**	T	T	T
500	1	D	s	--	g	**A**	**G**	G	G	**A**	A	A	A
921	1	J	J7	R -> H	g	g	g	g	g	g	**A**	**G**	G
1305	4	F	F101	G -> N	g	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**
1460	1	F	F153	Q -> E	c	c	c	**G**	**C**	c	c	c	c
2009	3	F	F336	S -> T	t	**A**	**T**	T	T	**A**	**T**	**A**	**T**
2093	1	F	F364	L -> F	c	c	c	**T**	**C**	c	c	c	c
2167	2	F	F101	H -> Q	t	**G**	**T**	T	T	**G**	**T**	T	T
Total # reverting sites per lineage:							**7**	**2**	**3**		**3**	**3**	**2**

Table 2

Figure 1. Atomic structure of the ϕX174 virion adapted from McKenna *et al.*, 1994. Color represents distance from the virion center. Reddish colors are closer to the center of the virion, blue colors are farther away from the center of the virion. The figure clearly shows the isometrical and icosahedral shape of the virion. The virion is composed of 60 copies of proteins F and G and 12 copies of protein H.

ϕX174 Virion

Adapted from Mckenna et al., 1994

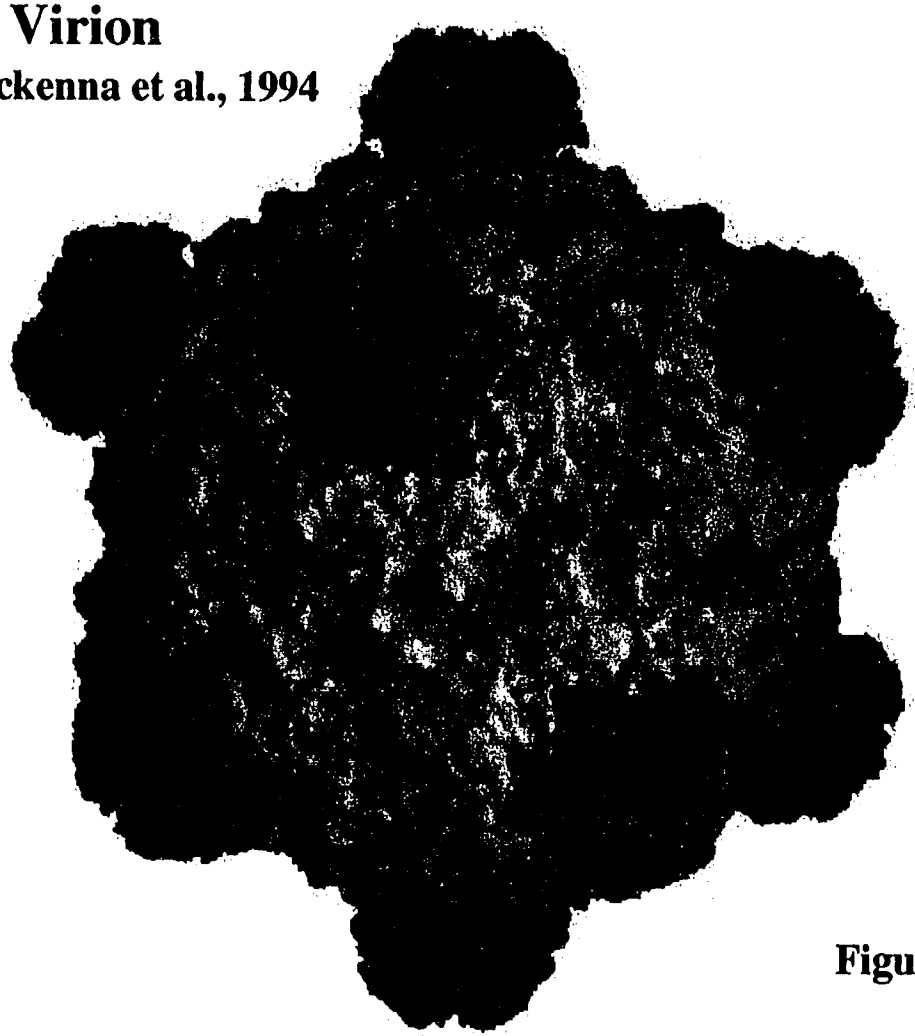


Figure 1

Figure 2. Schematic diagram of the ϕX174 genome. The genome consists of a single stranded, circular DNA molecule, 5386 bases in length, containing nine essential and two non-essential genes, some of which are transcribed from overlapping reading frames embedded within other open reading frames (Hayashi et. al., 1988).

ϕX174 Genome

5386 bases
circular ss DNA
9 essential genes

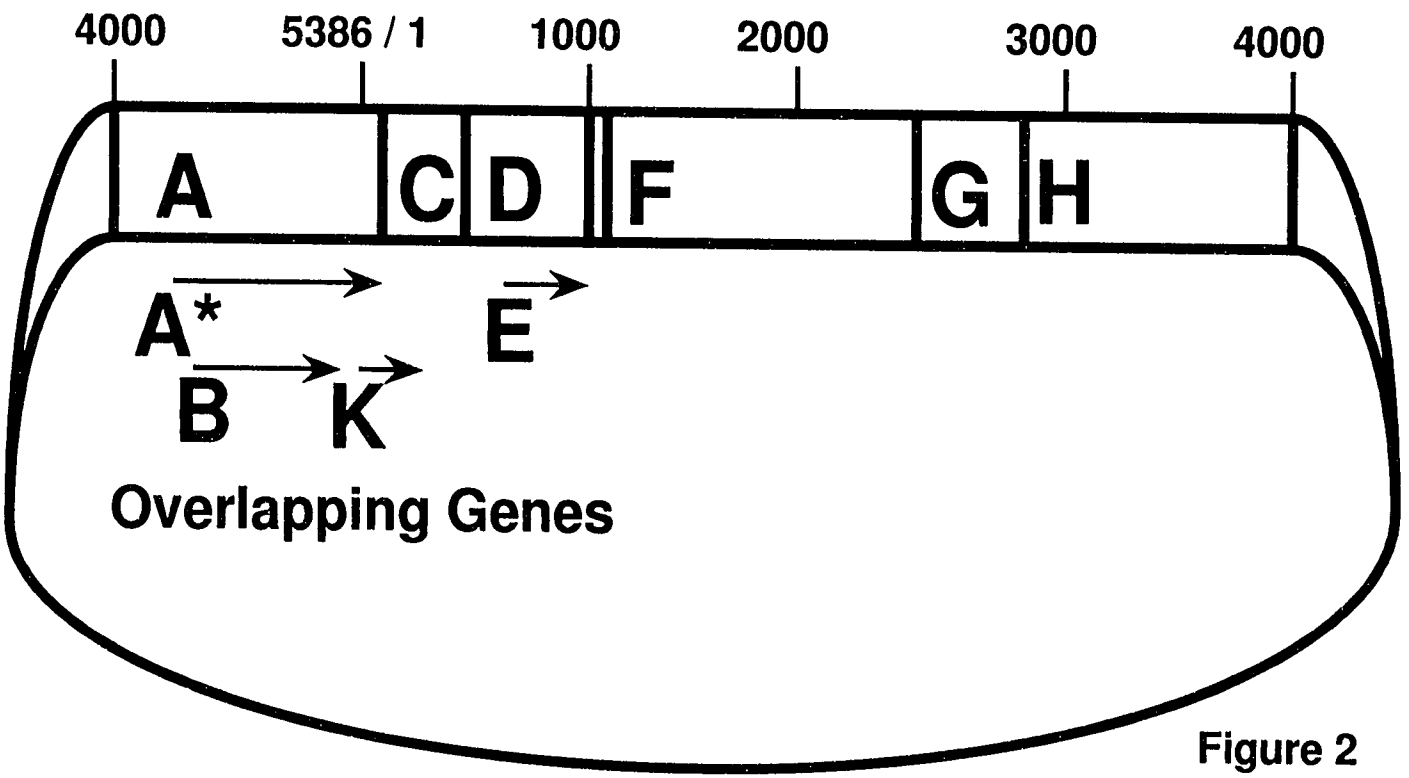


Figure 2

Figure 3. Schematic diagram of the chemostat apparatus. Phage populations were evolved in such a continuous culture, 2-chambered chemostat. Media was pumped continuously into the first chamber which contained cells only. Excess suspension from this cell chamber was drained continuously and unidirectionally into the 2nd chamber, which contained phage. Excess suspension from this phage chamber was drained off continuously. Aeration of both chambers was achieved with bubbling of filtered air through chamber one, drawn with negative pressure through chamber two and exhausted with excess liquid from this chamber. Both chambers consisted of 100 x 20 mm glass test tubes with stoppered tops.

The chemostat was sterilized and re-inoculated with cells from a frozen stock every two to three days. The most recent phage sample, treated with chloroform (which kills cells but not phage), was used to inoculate the phage tube. Thus, any bacteria adapted to the chemostat or the phage were discarded but phage adaptation accumulated.

A flow rate of 6-10 ml/hr resulted in total turnover of the cells and phage in the phage tube of about 80-100 times per day. Phage samples were collected from the chemostat, mixed with chloroform, and saved for future analyses and/or used to inoculate the chemostat after sterilizing and changing tubes, media, and cells.

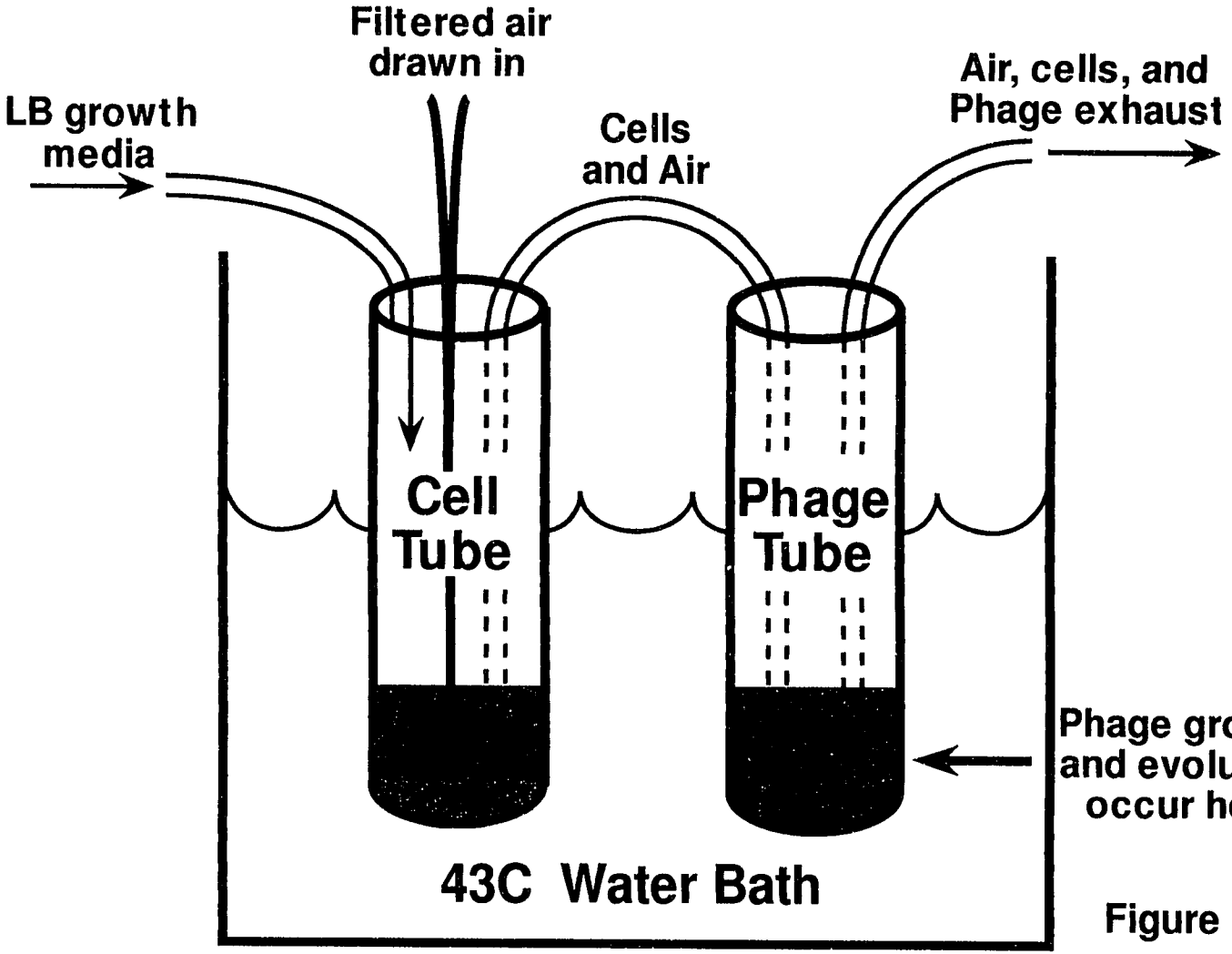
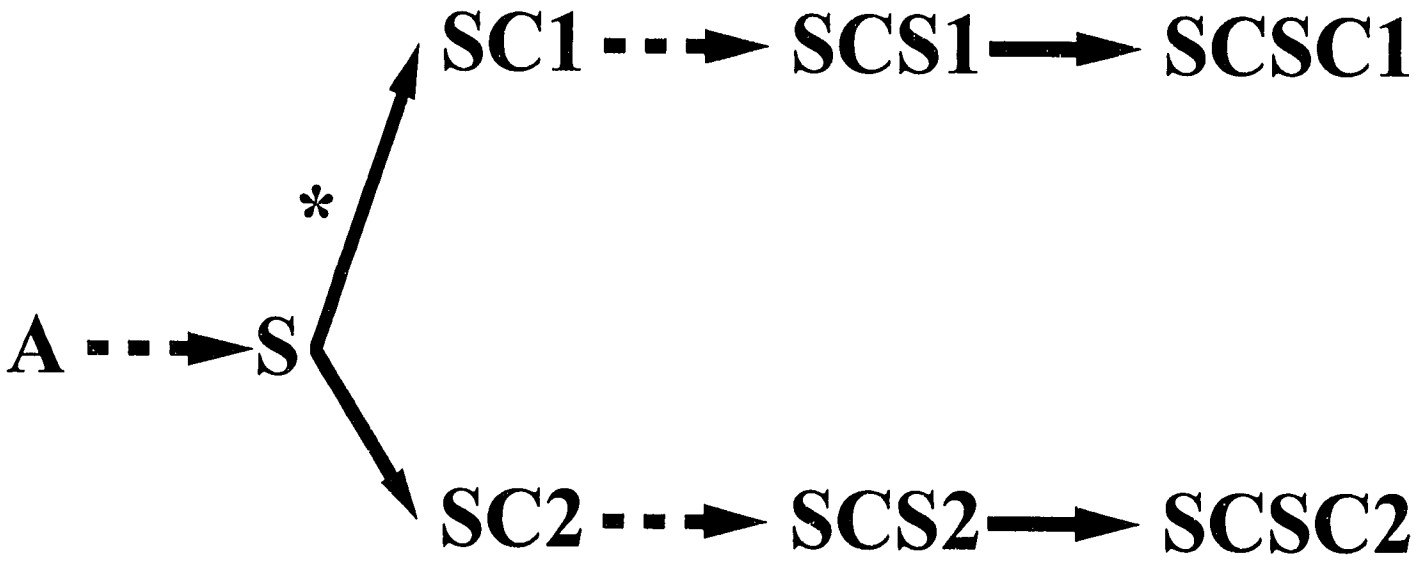

Filtered air
drawn in
LB growth
media
Cells
and Air
Air, cells, and
Phage exhaust
Cell
Tube
Phage
Tube
Phage growth
and evolution
occur here
43C Water Bath

Figure 3

Figure 4. Experimental design and historical relationships of sub-lineage isolates. The ancestral ϕX174 isolate (A) seeded the initial 11 day period of selection on the *S. typhimurium* host (S). A single plaque isolate (S1) was selected from this large population of 10^7-10^8 phage/ml. and saved for fitness and molecular analyses, and used to inoculate two new replicate lineages on the *E. coli* C (C) host. After adaptation to the *E. coli* hosts, individual plaque isolates (SC1 and SC2) from these populations were again selected for analysis and used to seed the next round of chemostat selection back on host S. This final period of selection on S resulted in the selection of individual plaque isolates (SCS1 and SCS2) for analyses and to seed the final period of selection in the chemostat on host C. After this final period of selection on C, individual plaque isolates SCSC1 and SCSC2 were chosen for analyses. All selections on a given host lasted 11 days except for the S1->SC1 *E. coli* selection which lasted 22 days. Individual plaque isolates and the periods of selection which produced them are referred to as sub-lineages of the two replicate lineages. This experimental design results in two replicate lineages each containing three independently derived sub-lineage plaque isolates (SC1 &2, SCS1 & 2, and SCSC1 & 2) and two shared, ancestral sub-lineage isolates (A and S1).

The selection of a single plaque isolate after each period of selection to begin selection on the alternative host results in a severe bottleneck of N=1 occurring at each host switching event. This ensures that all substitutions accumulated in the individual isolate are "fixed" in the population at the beginning of each selection period.

Daily whole culture samples were removed from the chemostat during the final periods of selection on S and on C for the population level fitness and molecular dynamics analyses. All plaque isolates were selected by plating the final whole culture sample at low density.



Selection on *Salmonella enterica*

Selection on *Escherichia coli C*

Figure 4

Figure 5. Fitness measured as growth rate of all sub-lineage plaque isolates. Fitness is measured as $\log_2$ of the phage increase per hour. This can be thought of as a doubling rate. The growth rate on *E. coli* is shown above the horizontal, the growth rate on *S. typhimurium* is shown below the horizontal. All 7 sub-lineage isolates including ancestor are shown across the top chronologically (in the order of selection) from left to right. Both replicates are depicted for the SC, SCS and SCSC sub-lineages. Replicate #1 is the left column, #2 the right column. Shaded bars represent growth rates of isolates measured on the host that the isolate was most recently selected on. Error bars represent 95% confidence intervals based on standard errors of the mean of 4 or 5 independent estimates of growth rate.

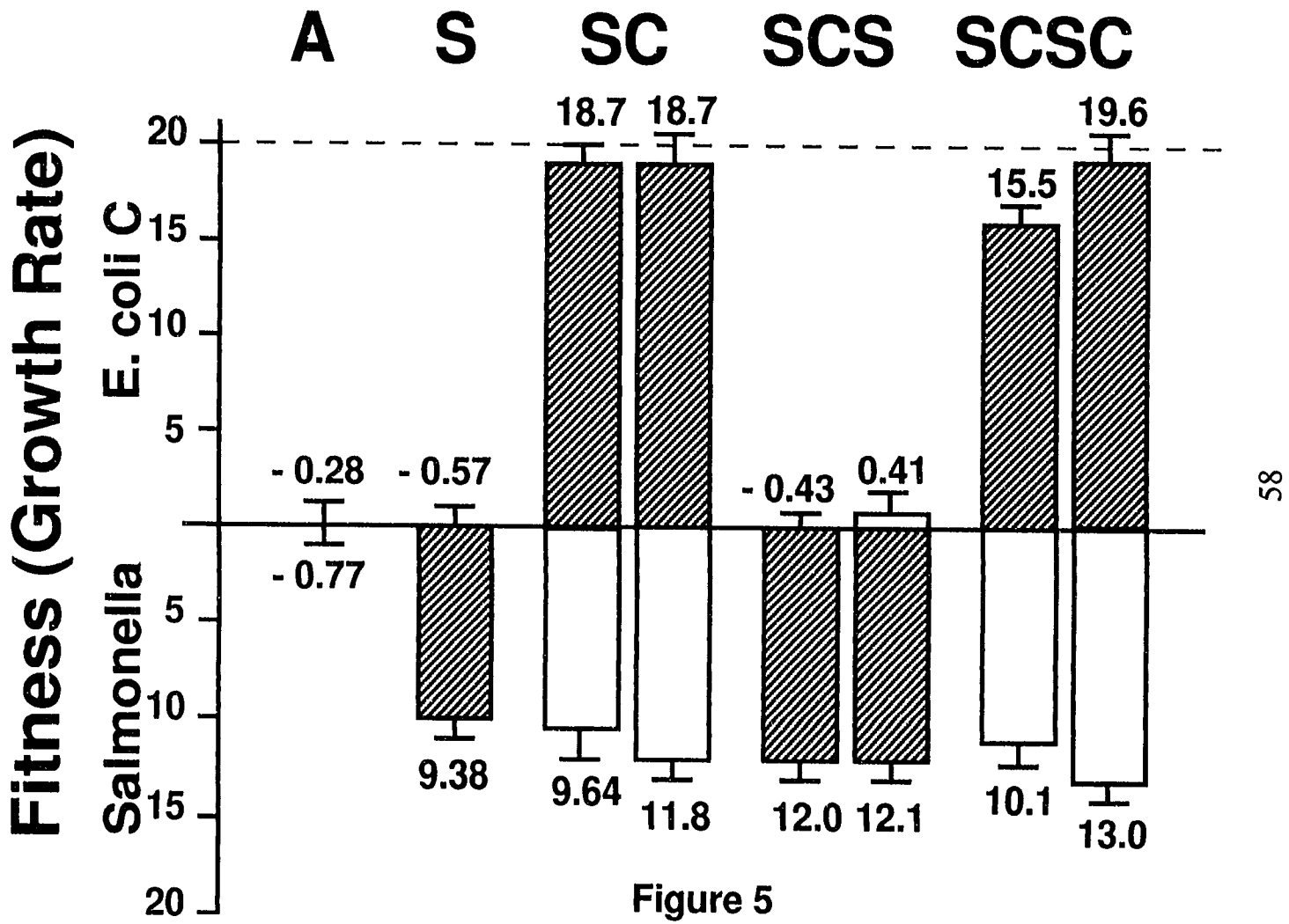

A
S
SC
SCS
SCSC
Fitness (Growth Rate)
E. coli C
Salmonella
20
15
10
5
5
10
15
20
- 0.28
- 0.77
- 0.57
9.38
18.7
9.64
18.7
11.8
- 0.43
12.0
0.41
12.1
15.5
10.1
19.6
13.0

Figure 5

Figure 6. Temporal pattern of fitness evolution and correlated selective sweeps of host-specific reversions in gene F for replicate #2.

The upper panel shows the daily pattern of fitness (measured as growth rate (GR), $\log_2$ phage increase per hour) evolution on each host first during selection on *S. typhimurium* , and then during selection on *E. coli* C. Error bars represent 95% confidence intervals based on standard errors of the mean of 3-5 independent estimates of GR per day. The bottom panel shows the population frequency at particular nucleotide sites. Frequencies were determined using an oligo screening assay (see methods). Host-specific reversions in gene F are depicted in thick bold lines. An "E" following a site represents the evolved nucleotide substitution associated with selection on S, as these sites change in frequency the ancestral state (back mutations associated selection on C) are always present at the frequency of 1-(the frequency of the evolved substitution). See table 1 or 2 for the actual nucleotide and amino acid changes. All amino acid changing substitutions evolved during these selections (SC2 to SCS2 and SCS2 to SCSC2, table 1) are shown.

During selection on S, selective sweeps at host-specific gene F sites 1305 and 2009 between days 4 and 8 are correlated with significant decreases in GR_C (day 4: $GR_C = 16.5\pm2.3$ vs. day 8: $GR_C = 0.91\pm2.1$, t-test: $p<0.001$). During selection on C, the rapid initial improvement in GR_C by day 1 (SCS2 isolate: $GR_C = 1.41\pm1.3$ vs. day 1: $GR_C = 14.4\pm0.91$, t-test: $p<<0.001$) is correlated with a sweep of the ancestral back mutation at site 1305. The second improvement in GR_C between days 10 and 11 (day 10: $GR_C = 16.3\pm0.75$ vs. day 11: $GR_C =$

21.5±0.33, t-test: $p<0.001$) is correlated with the sweep of the back mutation at site 2009. The exact nature of the dynamics of sites 593 and 921 are not known. Dots represent known frequencies at given times and the dotted line and question mark represents a lack of knowledge of the pattern of rate and timing of the change in frequency. The two amino acid altering substitutions are shown for the SCS2 to SCSC2 selection period on *E. coli*. Two additional mutations differentiate these sub-lineage isolates. A 2 base insertion between 3968 and 3969 and a silent substitution at site 551. Neither of these changes were present in the chemostat population prior to day 10 of selection on *E. coli*.

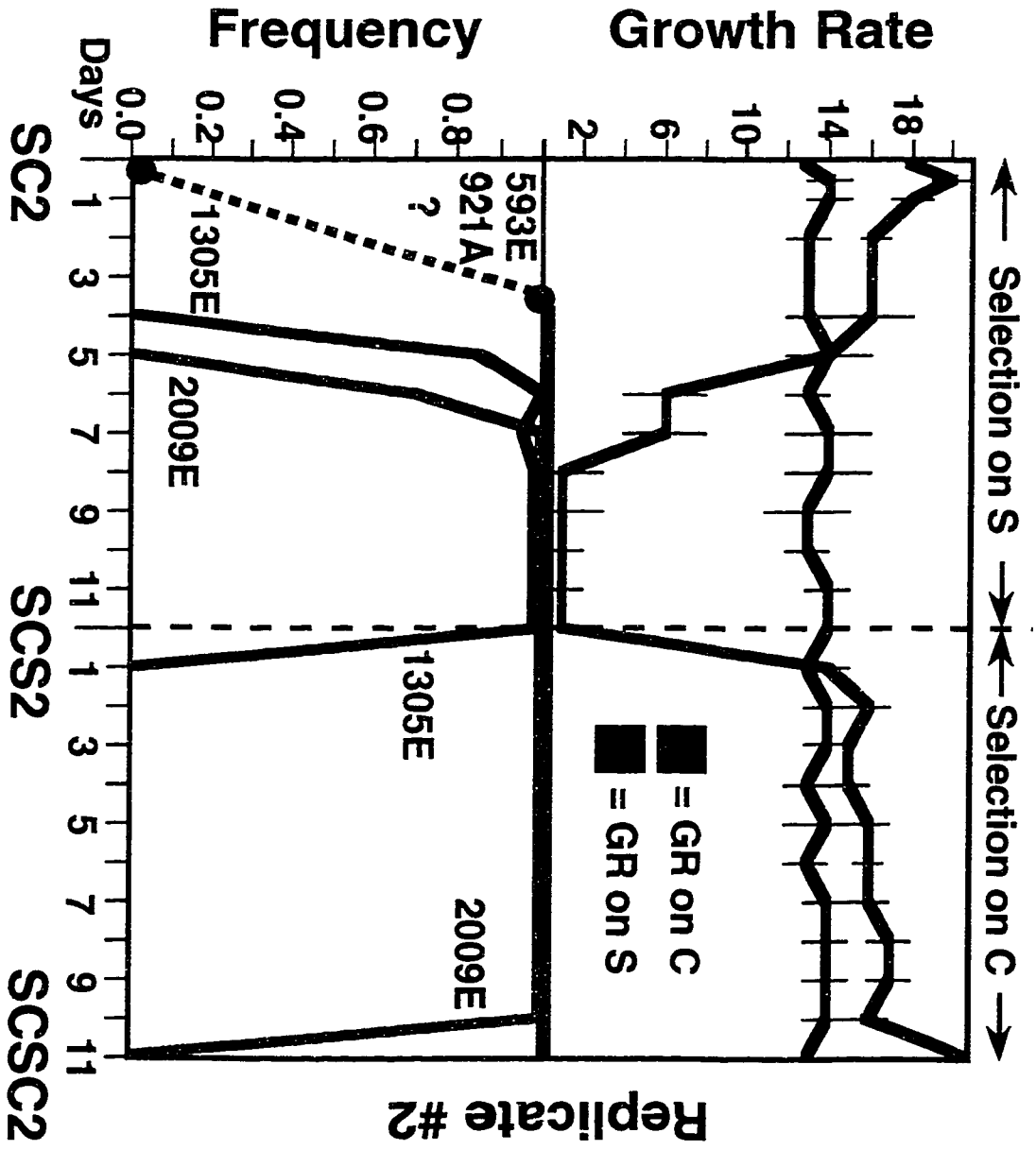

Frequency
Growth Rate
Days
0.0
0.2
0.4
0.6
0.8
2
6
10
14
18
SC2
1
3
5
7
9
11
SCS2
SCSC2
593E
921A
?
1305E
2009E
1305E
2009E
= GR on C
= GR on S
Selection on S
Selection on C
Replicate #2

Figure 6

Figure 7. Temporal pattern of fitness evolution and correlated selective sweeps of host-specific reversions in gene F for replicate #1.

The upper panel shows the daily pattern of fitness (measured as growth rate (GR), $\log_2$ phage increase per hour) evolution on each host first during selection on *S. typhimurium* , and then during selection on *E. coli* C. Error bars represent 95% confidence intervals based on standard errors of the mean of 3-5 independent estimates of GR per day. The bottom panel shows the population frequency at particular nucleotide sites. Frequencies were determined using an oligo screening assay (see methods). Host-specific reversions in gene F are depicted in thick bold lines. An "E" following a site represents the evolved nucleotide substitution associated with selection on S, as these sites change in frequency the ancestral state (back mutations associated selection on C) are always present at the frequency of 1- (the frequency of the evolved substitution). See table 1 or 2 for the actual nucleotide and amino acid changes.

All amino acid changing substitutions and subsequent reversions evolved in gene F during these selections (SC1 to SCS1 and SCS1 to SCSC1, table 1) are shown. Additional changes from these lineages are shown in table 1. Three novel transient polymorphisms (at sites 2015, 2168 and 2170) discovered via oligo screening but not correlated with fitness changes are not depicted in this figure.

During selection on S the selective sweep at host-specific gene F site 1305 between days 0 and 3 was correlated with significant decreases in GR_C (SC1 isolate: $GR_C = 18.1 \pm 0.97$ vs. day 3: $GR_C = 11.7 \pm 0.98$, t-test: $p << 0.001$). A sweep at the host-specific gene F site 1460 between days 3 and 5 is correlated with the

final drop in GR_C during selection on S (day 3: $GR_C = 11.7 \pm 0.98$ vs. day 5: $GR_C = 0.37 \pm 0.53$, t-test: $p << 0.001$). During selection on C the back mutation at site 1305 is again correlated with the rapid initial improvement in GR_C (SCS1 isolate: $GR_C = 0.089 \pm 1.67$ vs. day 2: $GR_C = 11.2 \pm 1.27$, t-test: $p << 0.001$). A sweep of the back mutation at gene F site 2093 may also be correlated with this initial GR_C improvement. It had swept through the population by day 4, and was present in a single plaque isolate on day 2, but other than that, its pattern of rate and frequency change is unknown.

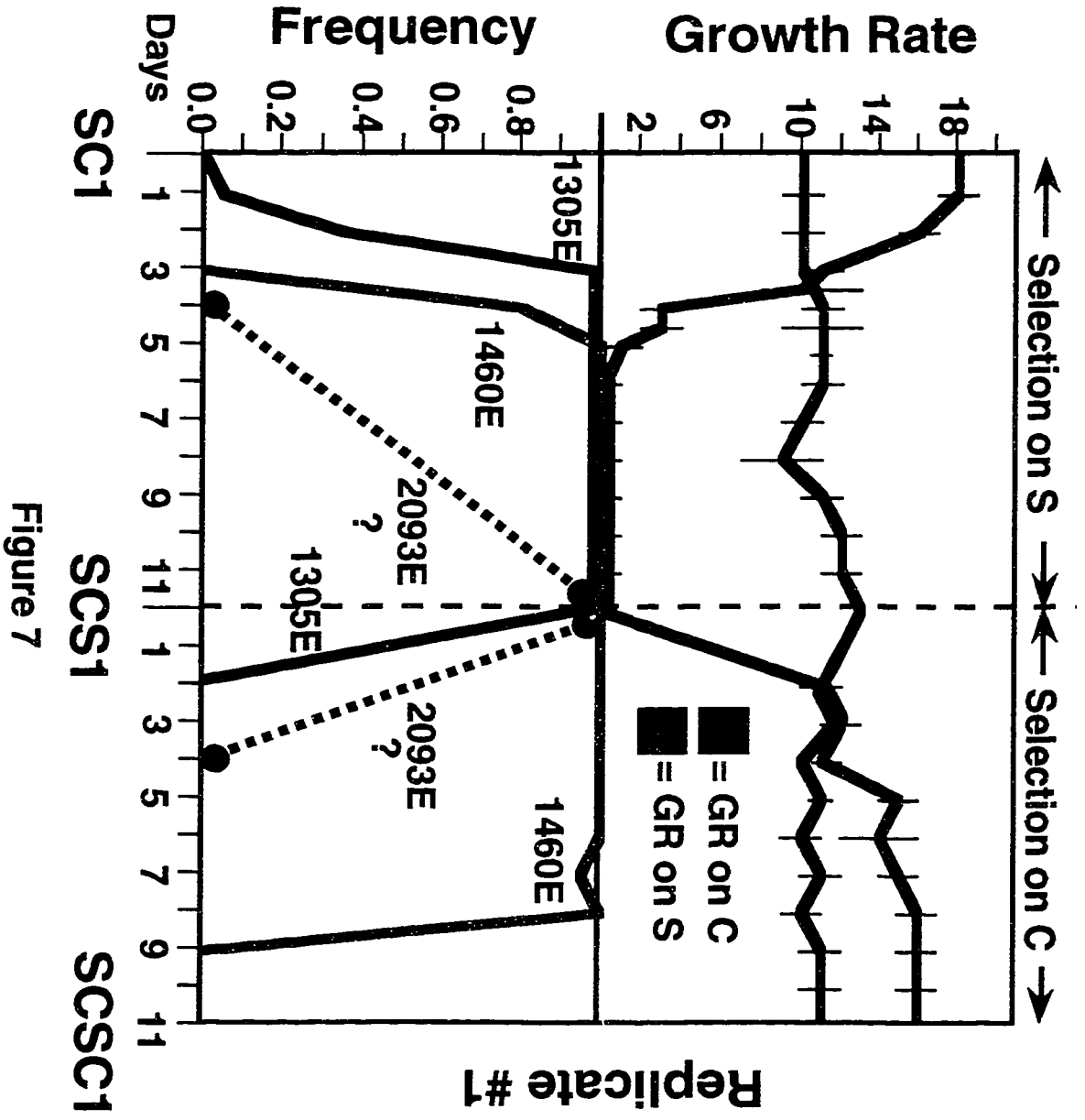

Frequency
Growth Rate
Days
0.0
0.2
0.4
0.6
0.8
2
6
10
14
18
1
3
5
7
9
11
1305E
1460E
2093E
?
1305E
2093E
?
1460E
= GR on C
= GR on S
Selection on S
Selection on C
SC1
SCS1
SCSC1
Replicate #1

Figure 7

Figure 8. Fitness effects of changes at site 1305 in different replicate genetic backgrounds. This figure shows the effect on GR_C of changing only site 1305 to its "evolved" or *S. typhimurium* adapted state in the *E. coli* adapted genetic background of isolates SC1 and SC2. Also shown is the effect on GR_C of changing only site 1305 back to its "ancestral" or *E. coli* adapted state in the *S. typhimurium* adapted genetic background of isolates SCS1 and SCS2. Each pair of columns represent replicate #1 in the dark and replicate #2 in the lighter shading. From left to right the first two columns are the GR_C values for the *E. coli* adapted isolates, SC1 and SC2. The second two columns are the GR_C values for these two isolates with only site 1305 changed from the ancestral "G" to an evolved "A". The third two columns are the GR_C values for the *S. typhimurium* adapted isolates SCS1 and SCS2. The fourth pair of columns show the GR_C values for the SCS1 and SCS2 isolates with only site 1305 changed from the evolved "A" back to the ancestral "G". The last two columns are the GR_C values for the final *E. coli* adapted isolates SCSC1 and SCSC2. All 1305 mutants were created using site-directed mutagenesis and sequenced entirely to be sure that there were no other changes.

The substitution of the evolved "A" for the ancestral "G" in the *E. coli* adapted SC isolates reduced GR_C dramatically and to similar degrees in both replicates. Clearly GR_C is not reduced to the extent that occurred in the *S. typhimurium* adapted isolates SCS1 and SCS2. GR_C is recovered to varying degrees by the substitution of the ancestral "G" for the evolved "A" in the *S. typhimurium* adapted SCS isolates depending on the genetic background of the replicate the substitution occurs in. Site 1305 has a much larger effect on

improving GR_C in replicate #2 than in replicate #1. The difference between the two replicates in the effect on GR_C of changing site 1305 can only be due to site 1305 acting epistatically with substitutions that have diverged since the splitting of the two replicate lines.

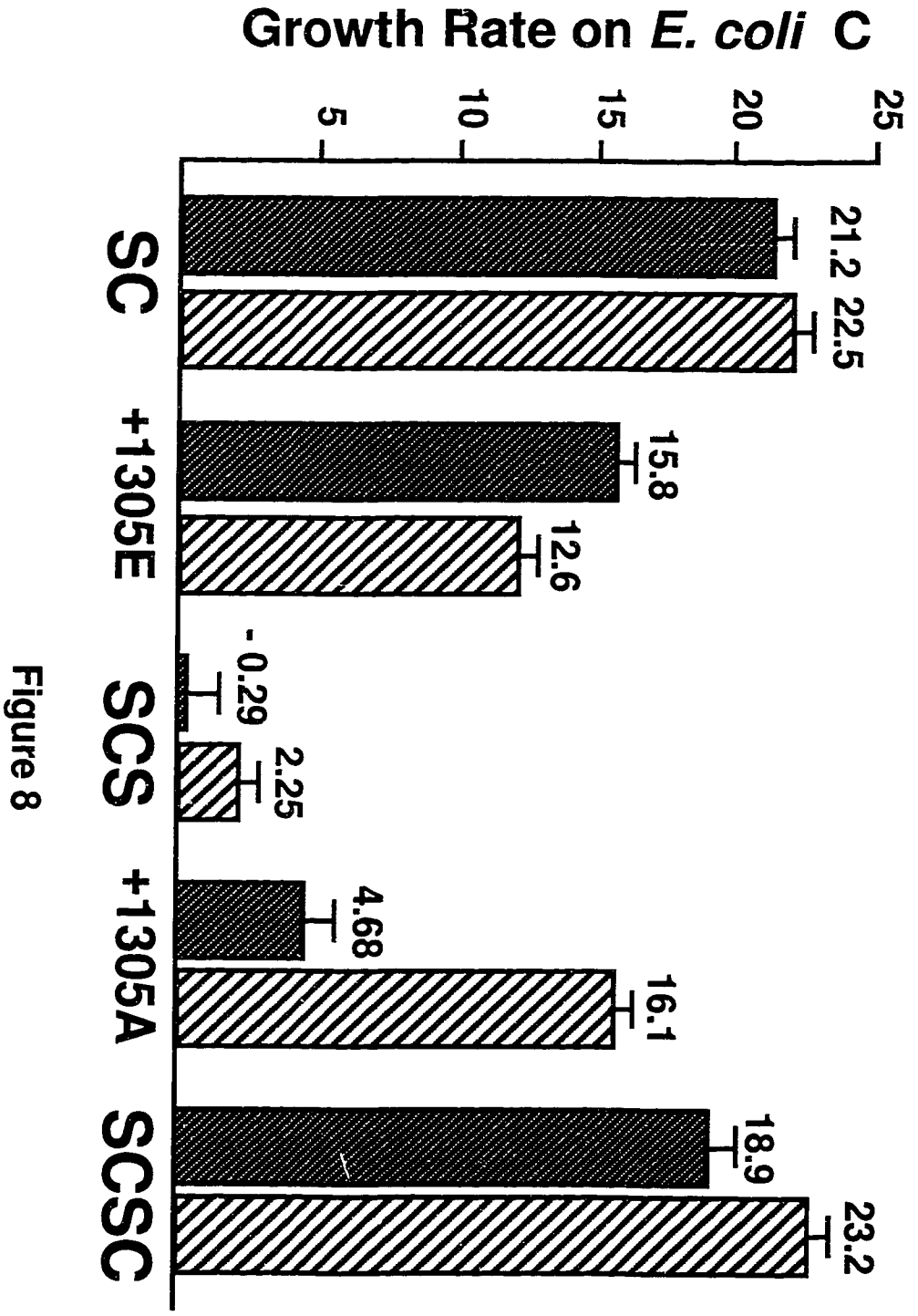

Growth Rate on E. coli C
5
10
15
20
25
SC
21.2
22.5
+1305E
15.8
12.6
SCS
- 0.29
2.25
+1305A
4.68
16.1
SCSC
18.9
23.2

Figure 8

Figure 9. Fitness effects of individual host switching mutations in the replicate #2 genetic background. This figure shows the effect of individual and groups of substitutions on GR_C in replicate #2. The vertical axis shows GR_C. Specific genotypes are listed below each fitness measurement. All listed genotypes are in the sub-lineage SC2 isolate genetic background, see table 1 for that specific genotype. Columns 1, 5, and 7 are the sub-lineage isolates SC2, SCS2, and SCSC2 respectively. All other columns represent isolates generated either by site-directed mutagenesis or removal of specific genotyped isolates from the evolving chemostat population. All isolates were verified by sequencing the entire genome. An "E" represents the evolved nucleotide associated with adaptation to *S. typhimurium,* an "A" represents the ancestral nucleotide associated with adaptation to *E. coli* C, see table 1 for the exact nucleotide changes at these sites. Note that the four sites listed for the SC2 and SCS2 isolates (593, 921,1305, and 2009) are the only sites that differentiate the SCS2 isolate from its SC2 ancestor (table 1). Moreover, sites 1305 and 2009 are the only two missense changes that evolved during the selection on C producing the SCSC isolate (site 551 is silent, table 1).

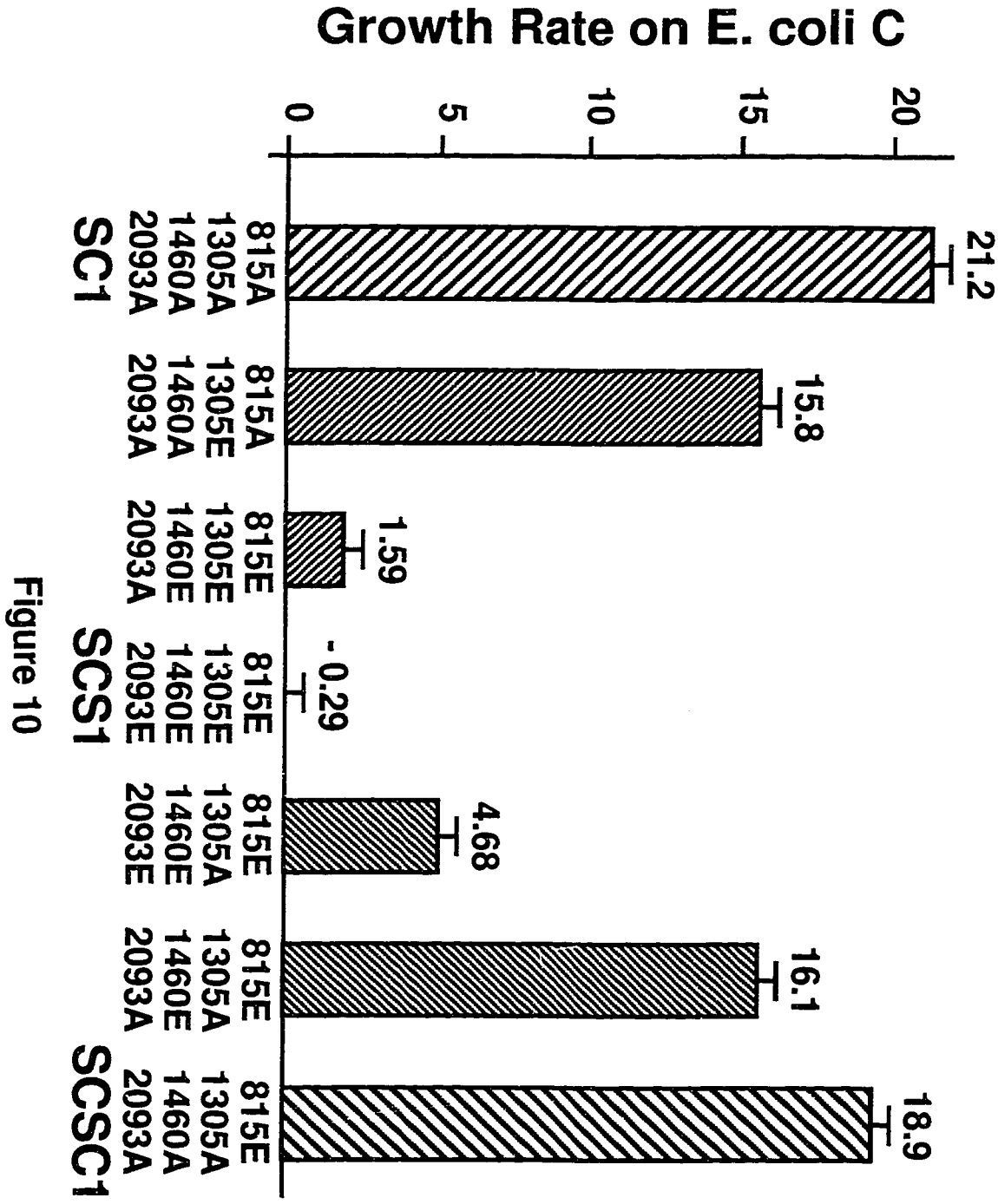

Growth Rate on E. coli C
0
5
10
15
20
815A
1305A
1460A
2093A
SC1
815A
1305E
1460A
2093A
815E
1305E
1460E
2093A
SCS1
815E
1305E
1460E
2093E
815E
1305A
1460E
2093E
815E
1305A
1460E
2093A
SCSC1
815E
1305A
1460A
2093A
21.2
15.8
1.59
- 0.29
4.68
16.1
18.9

Figure 10

Figure 10. Fitness effects of individual host switching mutations in replicate #1 genetic background. This figure shows the effect of individual and groups of substitutions on GR_C in replicate #1. The vertical axis shows GR_C. Specific genotypes are listed below each fitness measurement. Columns 1, 4, and 7 are the sub-lineage isolates SC1, SCS1 and SCSC1 respectively. All other columns represent isolates that were generated either by site-directed mutagenesis or removal of specific genotyped isolates from the evolving chemostat population. All isolates were verified by sequencing the entire genome. Sub-lineage isolate genotypes are listed only for the sites of interest, and have additional changes from ancestor which can be found in table 1. An "E" represents the evolved nucleotide associated with adaptation on *S. typhimurium*. An "A" represents the ancestral nucleotide associated with adaptation to *E. coli* C, see table 1 for the actual nucleotide changes at these sites. Mutant isolates are identical at all non listed sites to the sub-lineage isolate to its left (SC1 or SCS1). For example, SCS1 actually has two additional changes from the isolate to its left at sites 3111 and 4805, and all isolates to its right also have these changes.

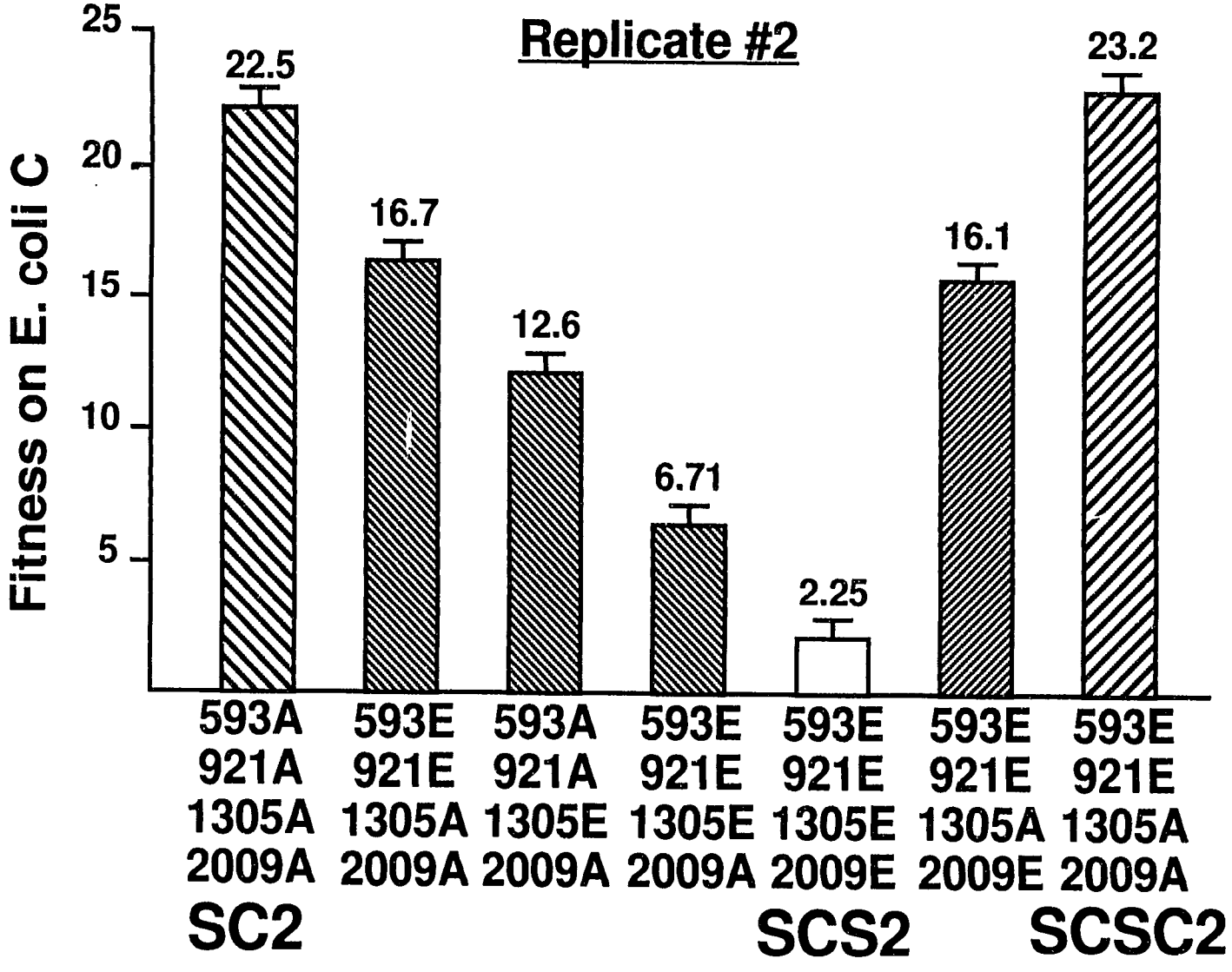

Replicate #2
Fitness on E. coli C
25
20
15
10
5
22.5
16.7
12.6
6.71
2.25
16.1
23.2
593A 593E 593A 593E 593E 593E 593E
921A 921E 921A 921E 921E 921E 921E
1305A 1305A 1305E 1305E 1305E 1305A 1305A
2009A 2009A 2009A 2009A 2009E 2009E 2009A
SC2
SCS2
SCSC2

FIGURE 9

Figure 11. Structure of the ϕX174 virion adapted from McKenna *et al.*, 1994. Color represents distance from the center of the virion. Reddish colors are closer to the center of the virion, blue colors are farther away from the center of the virion. I have outlined how single copies of the F protein (major capsid) and G protein (major spike) are individually arranged in triangular units. These triangular sub-units are organized in pentameric units surrounding each of the 12 icosahedral vertices for a total of 60 copies each of the F and G proteins in the complete virion.

Location of Pentameric gpF and gpG units in the complete ϕX174 Virion

12 copies of gpF pentameric units form the planar capsid surface

5 copies of gpG sit atop gpF pentamers at each vertex

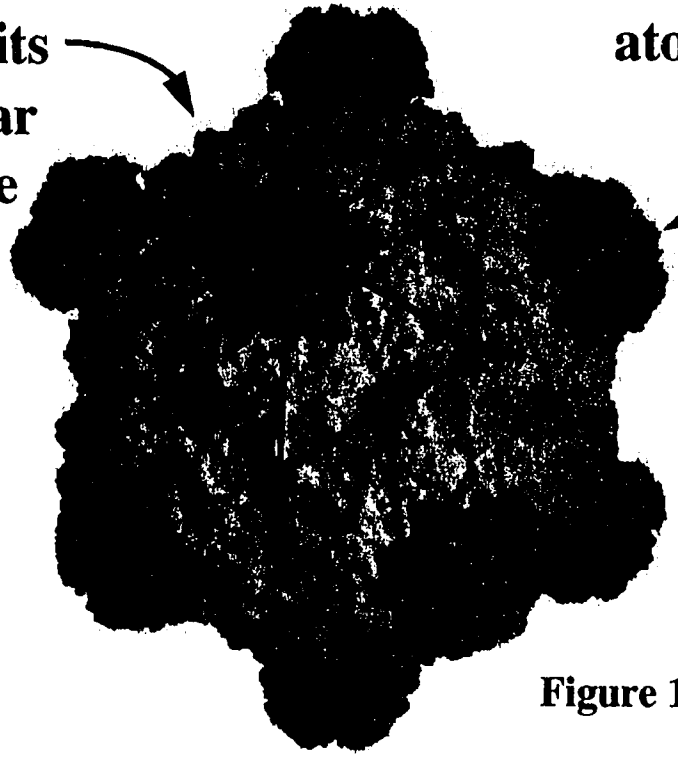


Figure 11

Figure 12. Molecular location of the five major capsid protein amino acid replacements which evolve and then revert with adaptation to the alternating hosts. A space filling model of a single copy of the major capsid protein is shown in red, the major spike protein is shown in blue. All five host-specific residues are highlighted in yellow. The nucleotide sites in gene F are shown to the left with the ancestor to evolved amino acid replacements. The light triangle identifies the single unit which forms pentamers around each vertex in the complete virion (fig 11).

All five host switching residues are located on the surface of the major capsid protein and none of them are covered by the major spike protein which sits on top of the capsid protein towards the 5-fold axis of symmetry. Once the five-fold symmetry of these individual F and G protein units is taken into account in the complete virion (fig. 11) it is clear that together the host-specific residues form a loose ring around the "spikes" found at each vertex, and formed by five copies of the spike protein (gpG).

Location of the Five gF Host Switching Residues

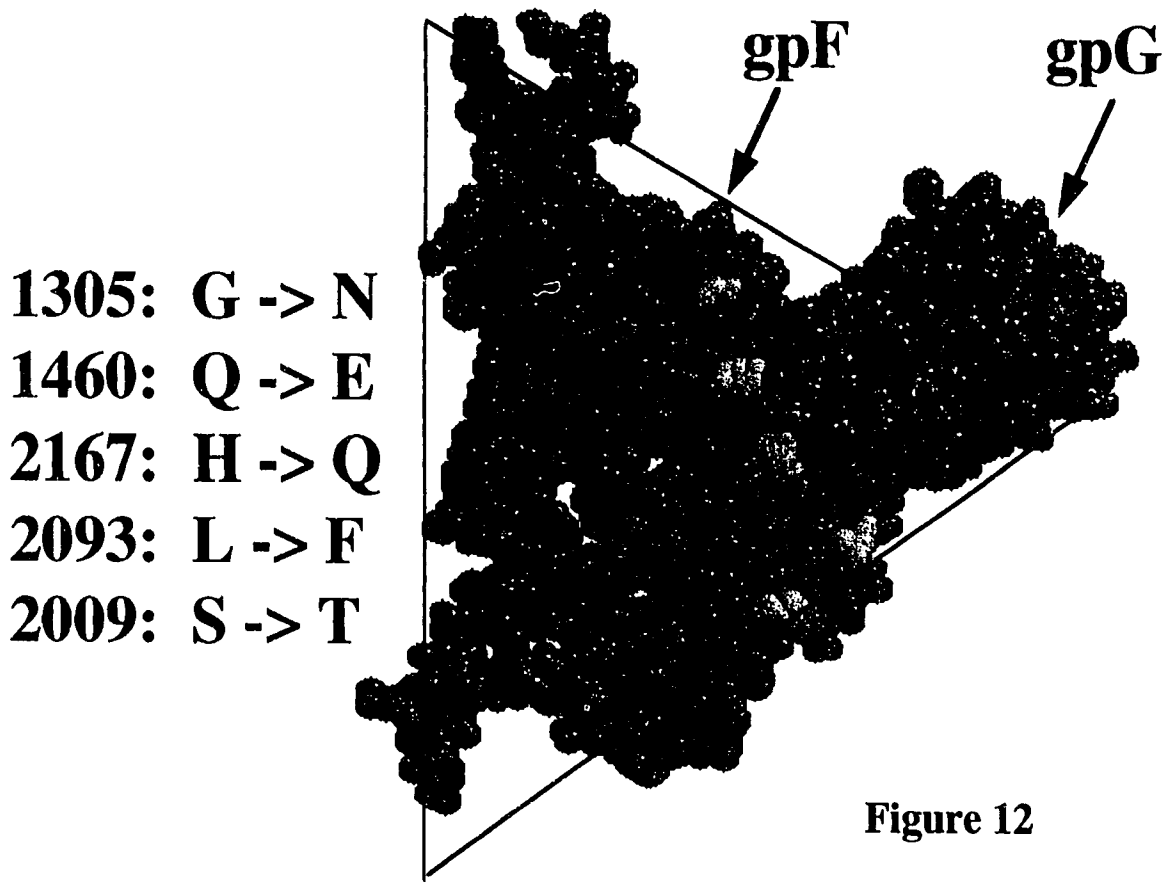


Figure 12

REFERENCES

AVERY, O. T., C. M. MACLEOD, AND M. MACCARTY. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp. Med., **79**:137-158.

BENZER, S. 1955. Fine structure of a genetic region in bacteriophage. Proc. Nat. Acad. Sci. USA, 41:344-354

BULL, J. J. 1994. Perspective : virulence. Evol. **48**: 4123-1437.

Bull, J. J., A. Jacobson, M. R. Badgett, and I. J. Molineux. 1998. Viral escape from antisense. Mol. Micro. **28**:835-846

BULL, J. J., M. R. BADGETT, H. A. WICHMAN, J. P. HUELSENBECK, D. M. HILLIS, A. GULATI, C. HO, AND I.J. MOLINEUX, 1997. Exceptional convergent evolution in a virus. Genetics **147**: 1497-1507

CASPARI, E. 1952. Pleiotropic gene action. Evolution **6**:1-18.

DUBOSE, R. F., D. E. DYKHUIZEN, AND D. L. HARTL. 1988. Genetic exchange among natural isolates of bacteria: recombination within the *pho*A gene of *Escherichia coli.* Proc. Nat. Acad. Sci. USA. **85**:7036-7040.

ELENA, S. F., AND R. E. LENSKI. 1997. Long-term experiemntal evolution in *Escherichia coli*. VII. Mechansims maintaining genetic variability within populations. Evol. **51**:1058-1067

ELENA, S. F., M. DAVILA, I. S. NOVELLA, J. J. HOLLAND, E. DOMINGO, AND A. MOYA, 1998. Evolutionary dynamics of fitness recovery from the debilitating effects of Muller's ratchet. Evol. **52**: 309-314.

FENNER, F. AND J. CAIRNS. 1959. Variation in virulence in relation to adaptation to new hosts. pp. 225-249 in *The viruses, Vol. 3. Animal viruses* , edited by F. M. BURNET AND W. M. STANLEY. Academic Press, New York.

HAYASHI, M., A. AOYAMA, D. L.RICHARDSON JR., ANDM. N. HAYASHI. 1988. Biology of the bactriophage ϕX174. in *The Bacteriophages,* R. Calendar editor, Plenum Press, New York.

HERSHEY, A. D., AND M. CHASE. Independent functions of viral protein and nucleic acid in growth of bacteriophage. 1952. J. Gen. Physiol., **36**:39-56.

JANDA, Z. AND V. VONKA. 1968. Variation in neurovirulence of NWS influenza virus repeated passages in different culture systems. I. Derivation and

neurotropic activity of NWS virus lines. Archiv für die Gesamte virusforschung **24**:192-196.

LENSKI, R. E. 1988. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in Competitive fitness among mutants resistant to virus T4. Evolution **42**:425-432.

Li, W-H. 1997. *Molecular Evolution*. Sinauer, Sunderland, MA.

LWOFF, A. 1961. Mutations affecting neurovirulence. in *Poliomyelitis, 5th international Poliomyelitits conference*. J. B. Lipincott, Philedelphia, PA.

MCKENNA, R., L. L. ILAG, AND M. G. ROSSMAN. 1994. Analysis of the single stranded DNA bacteriophage ϕX174, refined at a resolution of 3.0 Å. J. Mol. Biol. **237**:517-543.

Morse, S. S. 1993. *Emerging Viruses*. Oxford University Press, NY.

NICHOL, S. T., C. F. SPIROPOULOU, S. MORZUUNOV, P. E. ROLLIN, T. G. KSIAZEK, H. FELDMANN, A. SANCHEZ, J. CHILDS, S. ZAKI, AND C. J. PETERS. 1993. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. Science **262**:914-917.

NOVELLA, I. S., D. K. CLARKE, J. QUER, E. A. DOMINGO, AND J. J. HOLLAND. 1995. Extreme fitness differences in mammalian and insect hosts after continuous replication of vesicular stomatitis virus in sandfly cells. J. Virol. **69**: 6805-6809.

OCHMAN, H. AND A. C. WILSON. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**:74-86.

PETERS, C. J., E. D. JOHNSON, P. B. JAHRLING, T. G. KSIAZEK, P. E. ROLLIN, J. WHITE, W. HALL, R. TROTTER, AND N. JAAX. 1993. Filoviruses. In *Emerging Viruses*, S. S. MORSE editor, Oxford University Press.

RAYSSIGUIER, C., D. S. THALER, AND M. RADMAN. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutatnts. Nature **342**:396-401.

SABIN, A. B. AND L. R. BOULGER. 1973. History of Sabin attenuated poliovirus oral live vaccine strains. J. Biol. Stand. **1**:115-118.

SAMBROOK, J., E. F. FRITSCH, AND T. MANIATIS. 1989. *Molecular Cloning A laboratory Manual. Second Edition.* Cold Springs Harbor Laboratory Press, New York.

SANGER, F. G. M., A. R. COULSON, T. FRIEDMAN, G. M. AIR, B. C. BARRELL, N. L. BROWN, J. C. FIDDES, C.A. HUTCHINSON III, P. M. SLOCOMBE, AND M. SMITH. 1978. Nucleotide sequence of bacteriophage ϕX174 DNA. Nature **265**:687-695.

SAYLE, R. 1994. *RasMol v2.5, a molecular visualisation program*. Compuserve Incorporated.

SHARP, P. M. 1991. Determinants of sequence divergence between *Escherichia coli* and *Salmonella typhimurium:* codon usage, Map position, and concerted evolution. J. mol. Evol. **33**:23-33.

SINSHEIMER, R. L. 1959. Purification and properties of bacteriophage ϕX174. J. Mol. Biol. **1**: 37-55.

TESSMAN, E. S. AND I. TESSMAN. 1978. The genes of the isometric phages and their functions. in *The Single Stranded DNA Phages.* Cold Springs Harbor Laboratory, New York.

TURNER, P. E. AND L. CHAO. 1998. Sex and the evolution of intra-host competition in the RNA virus ϕ6. Genetics, In Press

TRUYEN, U., A. GRUENBERG, S. CHANG, B. OBERMAIER, P. VEIJALAINEN, AND C. R. PARISH. 1995. Evolution of the feline-subgroup parvoviruses and the control of canine host range in vivo. J. Virol. **69**: 4702-4710.

TRUYEN, U., J. F. EVERMANN, E. VIELER, AND C. R. PARRISH, 1996 Evolution of canine parvovirus involved in loss and gain of feline host range. Virol. **215**: 186-189.

WEBSTER, R. G., W. J. BEAN, O. T. GORMAN, T. M. CHAMBERS, AND Y. KAWAOKA. 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. **56**: 152-179.

WESTROP, G. D., K. A. WAREHAM, D. M. A. EVANS, G. DUNN, P. D. MINOR, D. I. MAGRATH, F. TAFFS, S. MARSDEN, M. A. SKINNER, G. C. SCHILD, AND J. W. ALMOND. 1989. Genetic basis of attenuation of the sabin type 3 oral poliovirus vaccine. J. Virol. **63**: 1338-1344.

WICHMAN, H. A., M. R. BADGETT, I. J. MOLINEUX, AND J. J. BULL. 199X. Temporal pattern of viral adaptation. In Prep.

WRIGHT, S. 1968. *Evolution and the genetics of populations, Vol. 1. Genetics and biometric foundations.* Univ. Chicago Press, Chicago, IL.

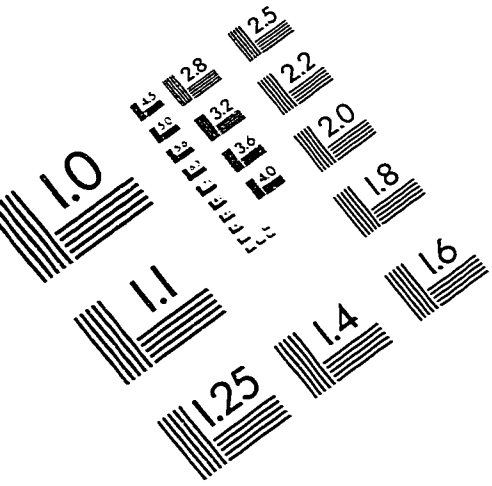
Vita

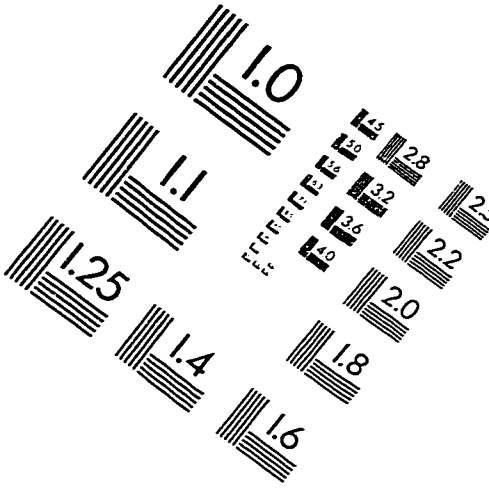
Wayne Douglass Crill was born in Seattle, Washington on May 27, 1968, the son of Wayne Elmo Crill and Jean Douglass Crill. After completing his high school education in 1986 he entered the University of Washington that fall. He left the university in 1988 and opened and managed a glass engraving store "A Glass Act" for one year in Tacoma Washington. He returned to the University of Washington in the fall of 1989. In the spring of 1990 he entered the BS program, and recieved his Bachelors of Science degree, in zoology in the spring of 1992. He entered the Graduate School of the University of Texas in the fall of 1992. He has published various articles in a wide variety of scientific journals from The Emu to Heredity and Evolution. He maintains his sanity by climbing up cliffs and jumping off of them.

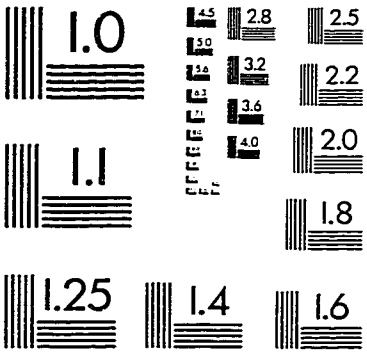
Permanent address: 5320 E. Mercer Wy., Mercer Island, WA 98040

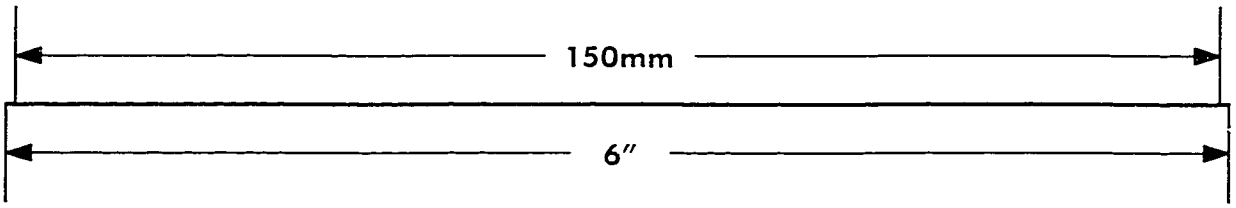
This dissertation was typed by the author.

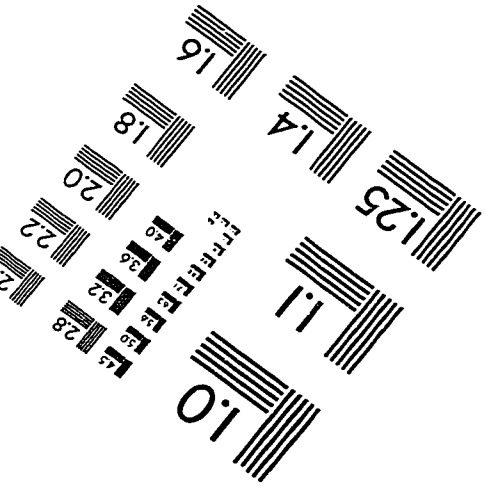
IMAGE EVALUATION TEST TARGET (QA-3)

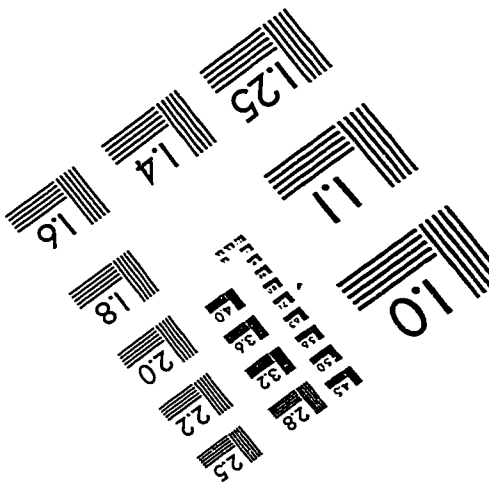












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989