

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.

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.

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

UMI[®]

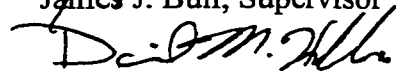
Copyright
by
Matthew Jonas Brauer
2000

Geometry and Genetics of Microbial Adaptation

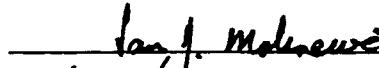
**Approved by
Dissertation Committee:**



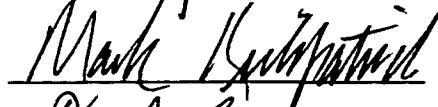
James J. Bull, Supervisor



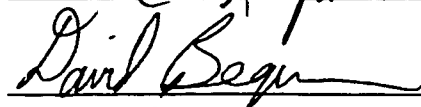
David M. Hillis, Supervisor



Ian J. Moloney



Mark Kunitz



David Beggs



H. Hillis

Geometry and Genetics of Microbial Adaptation

by

Matthew Jonas Brauer, B.A., M.S.Stat.

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, 2000

UMI Number: 3004221

UMI[®]

UMI Microform 3004221

Copyright 2001 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Dedication

I have been the most fortunate of students. From Mrs. Jane Carmichael, in the first grade at Lakeview Elementary, to my advisors in the UT Section of Integrative Biology, at every turn there has been someone encouraging me. Among these of course are my very first teachers, my parents Beth, Bryan and Mary Ann, and my grandparents Benno and Alma, Oliver and Agnes.

This dissertation is dedicated to them, all of my teachers.

Acknowledgements

*“It is a rare and tantalizing fruit
our hands reach for, but nothing absolute.”*

Countee Cullen's *Sonnet* is about the complex ambiguity of love, but the couplet applies equally well to most things outside of our grasp. We commit to a path without knowing where it leads, thinking we know our ultimate goal. But the path circles and meanders, and we must be sustained by the trust that the journey will be worth the effort.

It would be simple to acknowledge the following people as important contributors to the present work. However, their contributions extend far beyond the bounds of this dissertation. Each of the people I acknowledge have in some way helped me to maintain that trust in the process and in my ability to follow it through to the end.

The members of my committee deserve special thanks. My advisors, Jim Bull and David Hillis have been friends as well as mentors. Mark Kirkpatrick and David Begun have kept consistently been willing to put aside whatever they were doing

to help me puzzle through a particular problem or discuss an idea. Ian Molineux has been the intellectual origin and inspiration for much of the experimental work. Finally, I owe a great deal to Keith Crandall, who, in the too-brief time that we overlapped as office- and lab-mates, pushed me along the glorious road to Quantitative Genetics. I only wish I had had more time for the Yabbies.

Others were part of my 'virtual committee', having as much of an impact, with none of the glory. Holly Wichman heads this list, and I owe her a great deal for her ideas and inspiration.

I was fortunate to have two extremely talented undergraduate students helping me at various phases of these projects. Shawn Gerum spent a phenomenal summer keeping several of my experiments running while I was at meetings or working on the computer. Arundathi Shenoy was a welcome presence and great help during her two years in the lab.

Several members of the Andy Ellington Lab in the Department of Biochemistry contributed substantially to my thinking about adaptability and *in vitro* evolutionary systems. Andy himself constantly challenged me to think about what makes an interesting question in science. In addition to his significant intellectual contributions, Jamie Bacher helped tremendously in his heroic synthesis of the very long oligonucleotides used for constructing the "shallow random" library

described in chapter 3. That experiment would have failed (or at least would have had to have been repeated several times) with oligos of lesser purity or yield.

Without the support, inspiration, intellect and good-cheer of the other members of the lab, I would have not been able to finish and remain sane. These contributors to my growth and well being include students John Huelsenbeck, Sharon Messenger, Barb Mable, Jim McGuire, Wayne Crill, Laurie Dries, Kelly Agnew, Rafe Brown, Mark Holder and Kris Kichler-Holder. I'd also like to thank current and former post-docs Anna Graybeal, Dan Brumbaugh, Tom Wilcox, Joe Bernardo, and Mike Charleston. Finally Marty Badgett deserves a warm thank-you all to himself. These people have had a profound influence on who I have become, as a scientist and as a person.

A Graduate Research Fellowship from the National Science Foundation and a University Continuing Fellowship from the University of Texas School of Graduate Studies lightened the financial burdens of my graduate career substantially.

This work was funded in part by a Sigma Xi Grant-in-Aid-of-Research and by National Science Foundation Dissertation Improvement Grant DEB-9801639.

I would also like to thank the late great Department of Zoology at the University of Texas, and its generous and consistent donors, for ongoing support for travel to meetings, workshops and conferences. Dorothea Bennett deserves particular praise for her dramatic role in enhancing the graduate student experience.

My family has been very supportive of my extended stay in Austin. I would especially like to thank my aunt and uncle, Kevin and Renee, for extraordinary love, support and encouragement.

Finally, I would like to acknowledge the contributions of my wife, Jennifer. She has been a constant support and soul mate, and my tenure and trials through graduate school were made much easier by the fact that they were shared with her.

Having now come within sight of this particular path's end, I realize how little, in fact, it resembles an end. The product of this journey, the "rare and tantalizing fruit", is not much like what I had originally expected. Yet, notwithstanding the words of the poet, the love and support of my wife and family, my colleagues, mentors and friends have been, in fact, nothing *short of* absolute.

Geometry and Genetics of Microbial Adaptation

Publication No. _____

Matthew Jonas Brauer, Ph.D.

The University of Texas at Austin, 2000

Supervisors: James J. Bull and David M. Hillis

The process of adaptation was examined in the context of a simple microbial system. Simulation experiments showed that minor changes in the assumptions of Fisher's geometrical model of adaptation, consistent with the conditions of a microbial culture, have a large effect on the distribution of mutational effects. Experimental evolution studies showed that recovery of a simple virus from fixed deleterious mutations occurs primarily by reversion of those mutations. Further experiments showed that the virus has a number of equivalent mechanisms for mounting a robust response to an agent of strong selection.

Table of Contents

List of Tables	xiv
List of Figures	xv
Chapter 1: Geometry of Microbial Adaptation	1
INTRODUCTION	1
The Development of the Geometrical Model	1
A General Model of Adaptation.....	3
METHODS	5
Simulations.....	5
Clonal Competition	7
Asymmetry in Mutational Effect	9
RESULTS	11
DISCUSSION	12
Competition Among Beneficial Mutations.....	12
Asymmetry in Mutational Effect	15
Chapter 2: Recovery from Deleterious Mutations	32
INTRODUCTION	32
METHODS	33
Strains and General Methods.....	33
Evolution of Mutant Phage of Low Fitness.....	34
Recovery Conditions	35
Fitness Assay	36
Sequence Analysis	36
Site-Directed Mutagenesis	37
Using a <i>mutD</i> Host to Enhance Mutation Rate.....	39

Recombination of Mutant Lines	39
RESULTS	39
Mutation Accumulation	39
Recovery from Mutagenesis	41
Identification of Mutant Genotypes.....	42
Recombination Between Mutant Lines	43
Targeted Mutagenesis	43
DISCUSSION	43
Pattern of Accumulation of Fixed Deleterious Alleles	43
Recovery from Deleterious Mutations.....	44
Microbial Adaptation and Fisher's Model.....	46
Factors Affecting Compensatory Adaptation.....	47
Is Microbial Adaptation Primarily Accomplished by Compensatory Mechanisms?	48
Reversion as the Primary Mode of Fitness Recovery in ϕ X174	49
 Chapter 3: Adaptation of Bacteriophage to Escape from Autoinhibition.....	 58
INTRODUCTION	58
The Evolution of Microbial Drug Resistance	58
Designing New Drugs	59
Evolving Old Drugs.....	60
A Drug-Pathogen Arms Race: <i>E. coli</i> versus ϕ X174	61
METHODS	62
Strains and General Methods.....	62
Construction of Cloning Vector pUCUT	63
Cloning of ϕ X174 Sequences into pUCUT	64
Screening of Library for Inhibitory Activity	65
Sequencing of Plasmid and Bacteriophage DNA	66

Fitness Assay	67
Selection of Escape Mutants.....	69
Construction of Randomly Mutagenized Library.....	70
Challenge by Related Viral Inhibitory Sequences	72
RESULTS	73
Selection and Analysis of Resistance-Conferring Sequences	73
Selection of Viral Escape Mutants.....	74
Fitness Effect of Escape from Inhibition.....	74
Challenge by Related Viral Inhibitory Sequences	75
Selection of Resistant Sequences from the Randomly Mutagenized Library	75
DISCUSSION	77
Evolutionary Significance of Autoinhibitory Sequences	77
Autoinhibition and Evolutionary Versatility in ϕ X174.....	78
The Evolution of Viral Escape from Inhibition.....	79
Cost of Escape.....	81
Selection of Sequences from Random Library	82
Inhibitory Sequences from Related Phages	84
Regimes of Less Stringent Selection – Evolution of Intermediate Inhibitors.....	85
Appendix A: Source code of program ‘asym’	90
MAKEFILE	90
“ASYM.H”.....	91
“ASYM.CC”	91
Appendix B: Source code of program ‘fixx’.....	96
MAKEFILE	96
“FIXX.H”	96
“FIXX.CC”	101

Literature Cited.....	106
Vita	114

List of Tables

Table 1: Mutagenic primers, listed according to position in the ϕ X genome. The PCR primers used for synthesis (“PCR1” and “PCR2”) are from the pool used for sequencing. Labels for these primers refer to the approximate location in the genome. PCR1 primers are parallel to the positive strand, and PCR2 primers are parallel to the negative strand. PCR primers are between 12 and 17 bases long.	52
Table 2: Mutation accumulation in line A1.	53
Table 3: Mutation accumulation in line A3.	54
Table 4: Genotype and fitness for cultures and isolates from mutagenesis and recovery lineage A3. Sites of silent substitutions are shown in light-colored type. A “.” indicates no change in state from the previous recovery passage, and a “P” indicates a polymorphism at that site. Single plaques isolated from a culture are indicated by (i). Shaded cells highlight the substitutions presumed to be responsible for fitness changes.....	55
Table 5: Genotype and fitness for cultures and isolates from mutagenesis and recovery lineage A3. Legend follows that of table 4.....	56
Table 6: Correspondence between elements of pathogen/drug evolution and a bacteriophage model system.....	62
Table 7: Primers used for construction of randomly mutagenized library. Lowercase nucleotides are degenerate positions, as described in the text.	71
Table 8: Pair wise sequence differences between entire genomes (below diagonal) and between the autoinhibitory region (above diagonal) for the related bacteriophages ϕ X174, S13 and G4.	72
Table 9: Boundaries and orientation of resistance-conferring inserts. pCB102A is identical in sequence to the plasmid isolated by Van der Avoort.....	73
Table 10: Mutations in ϕ X resistant clones from randomly mutagenized library.	76
Table 11: Nucleotide and amino acid changes in wild type and escape phages. ...	81

List of Figures

Figure 1: Illustration of the method used for scaling mutational axes.....	21
Figure 2: The distribution of phenotypic factors fixed during the course of adaptation. The underlying distribution of mutation effect sizes was exponential with a mean of 0.1. These simulations were done with a population at a constant size of 10^8	22
Figure 3: The distribution of selection coefficients of mutations fixed during the course of adaptation. Experimental conditions are as described in figure 2. .	23
Figure 4: The distribution of phenotypic factors fixed during the course of adaptation. The experimental conditions are as described in figure 2, except the underlying distribution of mutation sizes was gamma, with a mean of 0.1 and shape parameter of 0.5.	24
Figure 5: The distribution of selection coefficients of mutations fixed during the course of adaptation. Experimental conditions are as described in figure 4. .	25
Figure 6: Distribution of effect sizes for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^6	26
Figure 7: Distribution of selection coefficients for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^6	27
Figure 8: Distribution of effect sizes for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^{10}	28
Figure 9: Distribution of selection coefficients for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^{10}	29
Figure 10: Time to first fixation as a function of variance in mutational axis scaling factors.....	30
Figure 11: Time to last mutation fixed as a function of variance in mutational axis scaling factors.....	31
Figure 12: Fitness profile for recovery of A3-5, showing frequency of early genotypes.....	57
Figure 13: Plasmid pUCUT and its construction.	86
Figure 14: Cloning of autoinhibitory sequences into pUCUT.	87
Figure 15: Strategy for synthesizing “shallow random” inhibitor library.....	88
Figure 16: Streak test for bacterial resistance. Strains B, D and E are susceptible, strain A is resistant, and strain C is partly so.....	88
Figure 17: Location and orientation of pF1 inhibitory sequence in phage genome.	88

Chapter 1: Geometry of Microbial Adaptation

INTRODUCTION

The Development of the Geometrical Model

In 1930, Fisher proposed that adaptation should occur primarily as a result of many beneficial mutations of small effect (Fisher 1930). He argued by means of an analogy: just as the movement of a microscope into perfect focus depends mainly on the fine-focus knob, so the movement of a population towards an optimum phenotype requires very small phenotypic changes.

In a formalized version of this argument, Fisher presented a “geometrical model of adaptation”, as it came to be known. The model was derived from a primarily phenotypic view of adaptation. In the model, an organism’s phenotype is partitioned into independent traits, each of which contributes to fitness. The phenotype of a monomorphic population is represented as a point in the n -dimensional trait-space defined by the n orthogonal traits. Likewise, the optimum phenotype (the combination of traits with the highest fitness) is represented by a single point in the space. The closer a phenotype is to this optimum, the higher fitness it has. The process of adaptation is then the movement of the population’s phenotype towards the optimum phenotype, with a concordant increase in mean fitness.

A population can change in phenotype *via* the action of mutation and selection (population size is assumed to be large, and drift is ignored). Each mutation generates a ‘proposal phenotype’ that may become the new population

phenotype. A mutation is therefore simply a potential change in phenotype: beneficial mutations are those that potentially move the phenotype closer to the optimum, while deleterious mutations move it further away. If its effect is large enough, a beneficial mutation will become fixed, and the entire population will move to the point defining the new phenotype.

The geometrical explanation for the prevalence of small mutational effects during the course of adaptation derives from the fact that very small mutations have a 50% probability of being beneficial. As a mutation increases in size, its probability of being beneficial decreases rapidly. This effect is heightened for trait-spaces with a large number of dimensions, *i.e.*, for complex organisms.

Fisher used this argument to calculate the probability that a mutation of a given size would be beneficial. As a result of this calculation, it became clear that the distribution of phenotypic effect sizes for beneficial mutations (or mutational “factors”) should be very strongly leptokurtic, and that the mode should be very close to zero. The vast majority of beneficial mutations should have a very small effect on phenotype.

Later Kimura (Kimura 1983) noted that, regardless of the distribution of beneficial mutation effects, the effects of beneficial alleles *observed in a population* would have a very different distribution. This is because the probability of fixation for mutations of very small effect is extremely low. Mutations of miniscule size, while very common, have a fixation probability close to that of neutral alleles. Larger-effect beneficial mutations, though more rare, fix

with much greater frequency. Kimura derived a distribution of *fixed* beneficial mutational effects, in which the mode is at mutations of intermediate size.

Fisher and Kimura both considered the distribution of beneficial effects arising in a population at a given distance from its optimum phenotype. In contrast, recent extensions of the model by Orr (1998; 1999; 2000) followed the progress of the population as successive beneficial mutations arise and are fixed. As a population approaches its optimum, large-effect mutations that might have formerly been beneficial no longer are so. With each step in the progress of adaptation, the average effect size of fixed mutations decreases. This iterative process shifts Kimura's distribution – which emphasized intermediate effect sizes – back towards zero.

A General Model of Adaptation

After accounting for all of the steps in the adaptive walk, Orr's conclusion was that beneficial mutations of small effect are those most commonly fixed during adaptation. In fact, his result was much more precise: the effect sizes of fixed beneficial mutations are distributed exponentially. His result was also extremely robust to changes in the overall distribution of mutation effect sizes. The precision and robustness of the result have supported Orr's claim that the model has value as a general model of adaptation.

The geometrical model of Fisher, Kimura and Orr makes several important assumptions about the interactions between organism and environment. The most important of these relate to the shape of the fitness surface: there is a

single fitness peak towards which the population moves, and its location does not change over time. The fitness surface is smooth on a scale comparable to that of mutation effect sizes, and the traits contributing to fitness are all under stabilizing selection.

Specific properties about the nature of mutational effects are also assumed. Mutations affect the values of the various orthogonal traits with equal loading, that is, on average each mutation affects each of the traits equally. For example, in a two-dimensional phenotype space defined by the traits of *color* and *size* (which itself might be a linear combination of height, weight, mass, *etc.*), the color component of a mutation's effect will be, on average, equal to the size component of that effect. Mutational phenotypic effects are additive: the effect of a mutation is independent of the position of the population experiencing that mutation.

A final assumption relates to the rate of appearance of new mutations. Fisher, Kimura and Orr imagined a serial process, whereby a single mutation is either fixed or lost before the next mutation arises. Competition among beneficial alleles is assumed not to occur, implying that the time between appearances of mutations that are destined to be fixed must be much less than the fixation time. In order for the appearance rate of 'fixing mutations' to be less than the inverse of the fixation time, as this implies, the appearance rate of all new mutations must be very low, the proportion of new mutations that eventually fix must be very low, or both.

In this chapter I examine the performance of the Fisher-Kimura-Orr model when some of these assumptions are relaxed. The first result will show that the exponential distribution of fixed effects is strictly true only when the time between new beneficial mutations is longer than the time taken for a beneficial mutation to fix. The effect of including a modest increase in the rate of beneficial mutations is to shift the distribution significantly towards a mode of intermediate size. The distribution of the selection coefficients of mutations fixed during adaptation likewise shifts away from the exponential.

The second result demonstrates the effect of introducing a consistent bias in the degree to which mutations affect different traits. In this case, the distribution of fixed effects remains exponential, but the proportion of mutations that are fixed declines dramatically. Including this complication drastically increases the time it takes for a population to reach the optimum.

Individually, these results imply that the general conclusions of the geometrical model might not apply to a large population composed of organisms with a high genomic mutation rate. However, taken together the results suggest a mechanism by which adaptation might proceed in the manner of the Fisher-Kimura-Orr model, even in such populations.

METHODS

Simulations

The analytical approach developed by Orr (1998, 1999) allows the derivation of a distribution of relative mutational effects. These are the same as

the ‘factors’ described by Fisher, and are the mutational effects scaled to the distance remaining between the population and its optimal phenotype. That approach cannot be used to determine the distribution of absolute effect sizes or mutational fitness effects. Furthermore, Orr’s 1998 derivation assumes that the population is monomorphic. For these reasons a Monte Carlo approach was adopted, in which adaptation through a Fisher-like trait space was simulated.

In these simulations each point in trait-space represents a combination of trait values, or a composite phenotype. For simulations that allow variation within the population, each of several points represent the individual phenotypes that comprise the population. In simulations that do not incorporate variation, a single point represents the phenotype of the monomorphic population. Each point has an associated fitness, defined by the distance of the point from the origin, which is assumed to represent the optimum combination of trait values.

The first set of simulations model clonal competition and so allow multiple phenotypes to exist simultaneously and to compete with each other. Each phenotype is represented by a separate point, and kept track of individually. In the second set of experiments, the phenotype of a monomorphic population is modeled. At each iteration a mutation defines a proposal phenotype which will fix, becoming the new population phenotype, depending on the strength of the mutation’s effect.

Clonal Competition

This experiment simulated adaptation in a population with a large number of new mutants in each generation, that is, with a large value of population size times mutation rate. Because the time to fixation was assumed to be longer than a single generation, multiple phenotypes were present in the population throughout the course of the adaptive walk. This was a substantial departure for the models of Fisher and Orr, who each assumed that mutations were rare enough that they could be considered only one at a time and that the population was uniform in phenotype. In these experiments the frequency, mutational composition and fitness of each newly arising phenotype in the population thus had to be accounted for.

In these simulations, each phenotype was represented as a vector in an n -dimensional phenotype space, where n is the number of traits under simultaneous stabilizing selection. Mutations were potential displacements of the phenotype, and were likewise n -dimensional vectors. Mutations that moved the phenotype closer to the optimum were beneficial, and increased the fitness of the phenotype.

Each generation followed these steps:

1. The frequency, p , of each extant phenotype was adjusted deterministically by the equation $\Delta p = p \left(\frac{w}{\sum_i w_i p_i} - 1 \right)$, where w is the fitness of the phenotype and $\sum_i w_i p_i$ is the product of frequency and fitness summed over all extant phenotypes. A phenotype's fitness was

proportional to a gaussian function of its Euclidean distance z from the optimum: $w = \exp\left\{\frac{-z^2}{2}\right\}$.

2. For each extant phenotype a value was drawn from a binomial(Np, μ) distribution (where p is the frequency of the phenotype, N is the population size and μ is the mutation rate). This became the number of potential new descendant phenotypes. Each of these was given its parent's position vector in trait-space and its parent's complement of mutations. Each new phenotype was also given an additional mutation with magnitude drawn randomly from the distribution $m(r)$, and direction distributed uniformly. This mutation vector was added to the phenotype's position vector, and the length of this new position vector was determined. If this new position was further from the origin than the parent's position, the mutation was deleterious, and the new phenotype was discarded. Beneficial mutations were those that moved the position closer to the optimum, and either of two fates could befall the new phenotype with that mutation. With probability e^{-2s} the phenotype was lost. Otherwise its frequency was instantaneously adjusted upwards to a 'threshold of deterministic behavior'. This threshold was defined as $1/sN$ (equal to $1/s$ individuals), where the selection coefficient $s = \frac{w}{\bar{w}} - 1$. This method compresses the time taken to reach the threshold frequency, but does not limit the amount of time taken for further increases in frequency.

3. Phenotypes of low frequency were subjected to stochastic loss. If the frequency of a phenotype dropped below the threshold $\frac{1}{sN}$, the phenotype was lost with probability e^{-2s} . If not lost, the phenotype's frequency was set to the threshold.

Each trial of the simulation started with a single phenotype of frequency 1.0 at a distance of 1.0 from the optimum, and progressed until the average fitness of the population had increased to 90% of the maximum possible fitness gain. At that point the frequencies of the mutations present in extant phenotypes was calculated: those that were present in 70% or more of the population were considered fixed. (In practice, no high-frequency mutation that was present at the end of the simulation had a frequency of less than 95%.)

Up to 100 trials were run to accumulate data on the distribution of fixed mutation effects for each of three mutation rates and at three population sizes: 10^6 , 10^8 and 10^{10} . In addition, a simulation similar to that described by Orr (1999) was included for comparison. This set of experiments was done with each of two distributions of underlying mutation effect sizes: an exponential with mean 0.1 and a gamma with mean 0.1 and shape parameter 0.5.

Asymmetry in Mutational Effect

I examined the result of asymmetrical trait effects by simulating an adaptive walk of a population towards an optimum in a trait space of 10 dimensions. In contrast to the simulations described above, and in common with the simulations of Orr, the entire population was considered to be monomorphic.

Mutations were potential displacements of the population that occurred only if the mutation was fixed by selection.

The simulation followed the following procedure:

1. At the start of each adaptive walk, n weighting factors were drawn from a gamma distribution with mean = 1 and variance depending on the experiment.
2. Mutation vectors were drawn from a multivariate gaussian distribution, and scaled by the gamma-distributed weights such that the scaled mutation vector was of the same length, but biased in direction according to the weights;
3. Adaptive walks were simulated, with the starting phenotype equally poorly adapted in each trait. Throughout the adaptive walk the size of the fixed effects was recorded. Additionally, the time taken to adaptation, in number of mutations evaluated was recorded.

The effect of changing the scaling of mutation-effect axes is shown in figure 1, for a simplified phenotype-space of two dimensions. The trait represented by the x-axis is on average more affected by mutations, while the mutational effect on that represented by the y-axis is decreased. In this example, the x-axis is dilated by a factor of 1.5, while the y-axis is contracted by a factor of 0.5.

In the simulations of (Orr 1998), all axes were equally weighted, and mutations were therefore uniformly distributed in direction. Since all of the weighting factors were equal, the variance of the axis weighting factors equaled 0. For the current experiments, four conditions were examined: axis weighting factors were taken from four gamma distributions, each with a mean of 1.0, and with variances of 0.0, 0.1, 1.0 and 10.0.

In each simulation, axis weights were drawn from the appropriate distribution and the actual sample variance was calculated. The results of the simulation were correlated with this sample variance of the axis weighting factors.

RESULTS

For populations of size 10^8 , the exponential distribution of fixed phenotypic factors disappears completely with a value of μ as small as 10^{-10} (fig. 2 and fig. 4, $N\mu = 0.01$). Increasing the mutation rate to the more reasonable values of 10^{-8} ($N\mu = 1$) or 10^{-6} ($N\mu = 100$) result in a dramatic departure from the exponential distribution. The distribution of selection coefficients of fixed mutations change similarly with the increase of $N\mu$ (fig. 3 and fig. 5). This result is observed whether or not the distribution is exponential (figs. 2 and 3) or gamma (figs. 4 and 5). This departure from the exponential distribution for fixed effect sizes is not an artifact of the particular population size used in the simulations. Additional simulations, with populations of size 10^6 and 10^{10} show a similar effect (figs. 6-9).

The distribution of fixed phenotypic effects does not change substantially with the increase in mutational asymmetry, and neither does the distribution of fitness effects from fixed mutations (not shown). However, there is a large difference in the variance of the total number of mutations that need to be sampled before the first mutation is fixed (figure 10), and in the total number of mutations sampled before adaptation (figure 11). Assuming that $N\mu$ remains constant during the course of the adaptive walk, a very small increase in the mutational asymmetry dramatically increases the amount of time needed for adaptation.

DISCUSSION

Competition Among Beneficial Mutations

The results show that even a moderately large value of $N\mu$ substantially changes the distribution of fixed phenotypic effects. This effect is present, although less pronounced, when the distribution of selection coefficients is examined. The result is not dependent on the shape of the underlying distribution of mutational effects.

The reason for this result is simple: when the rate of beneficial mutations is higher than the rate of fixation, potential fixed mutations are no longer the *first* beneficial mutations encountered during an iteration, but rather the winners from among many beneficial mutations arising during the course of fixation. The qualitative difference between the model considered by Orr and the model

described here suggests that there is a limited range of population parameters over which Orr's general conclusions should be expected to apply.

This variation may be especially important in considering microbes with large population sizes and high mutation rates (Drake 1994). An effect of high mutation rates and large population sizes is that potentially all genotypic variants that lay one mutational step away from the canonical phenotype may be present in the population (Swetina and Schuster 1982; Schuster and Swetina 1988; Schuster and Sigmund 1989). This constant introduction of variation results in a "cloud" of mutants surrounding the wild-type sequence and competing with themselves and the wild-type genome.

The effect of a pool of variants on the adaptation of microbial populations has long been recognized. The concept of viral "quasispecies" in particular was developed in part as an attempt to understand the unique qualities of microbial evolution (Eigen and Schuster 1977; Steinhauer and Holland 1987; Domingo *et al.* 1996; Domingo *et al.* 1998; Domingo *et al.* 1998; Elena *et al.* 1998). Although the quasispecies concept is perhaps now used as an extreme case of mutation-selection balance, it does provide a counter-example to the model of adaptation assumed by Fisher, in which variation is not considered.

The strength of selection also plays a role in the importance of variation in the course of adaptation. Under weak selection, the probability that any one beneficial mutation fixes is small (roughly the inverse of population size, in fact) (Kimura 1983). In that case, any single beneficial mutation has roughly the same probability of fixation as any other. This situation corresponds to the assumptions

of Orr's model. It is sufficient to sample mutations one at a time, sequentially, until one is found that successfully fixes. When selection is strong, the difference in fitness effects among beneficial mutations becomes significant. The best mutation from among a large number sampled would be expected to have a larger effect than the first beneficial mutation found to fix. This correlates with the conclusion drawn by Lande (Lande 1983) that large-effect mutations are more likely to be seen as a result of artificial selection, rather than in natural populations, which generally experience weaker selection.

The importance of competition among beneficial mutations has been discussed previously. Fisher (Fisher 1930), Muller (Muller 1932) and, more recently, Gerrish and Lenski (Gerrish and Lenski 1998) have examined the effect of so-called "clonal interference" on the rate of adaptation. In asexual populations, a beneficial mutation will be fixed only if it escapes the effects of drift and is also not out-competed by a mutation of larger beneficial effect. Fixed mutations are thus expected to have larger effects than they would in the absence of clonal interference.

An upper bound can be calculated for the rate of appearance of mutations that eventually fix, assuming that the population is behaving according to Orr's model. The rate of appearance of new mutations is the product of the population size N and the genomic mutation rate μ . The average fixation time (in units of N generations) for a mutation with selection coefficient s in a population of size N is given by $\bar{t} = \frac{1}{Ns \cdot e^{Ns} - 1} \int_0^1 \frac{(e^{Nsx} - 1)(e^{Ns(1-x)} - 1)}{x(1-x)} dx$ (Kimura, 1969). If there is to be only one mutant in the population at a time, mutations destined to fix must appear

at a rate less than the inverse of this value. The proportion of mutations that ultimately fix must then be less than $\frac{1}{iN^2\mu}$. Assuming that $N=10^8$, a mutation with selection coefficient $s=0.01$ has an average fixation time of almost 3000 generations. For a mutation rate of 10^{-6} , the proportion of all mutations that ultimately fix must be less than about 3×10^{-6} , otherwise beneficial mutations will exist simultaneously in the population, and competition among them will begin to be important.

The departure from the exponential distribution for fixed effects was seen as well in simulations with a much simpler set of assumptions (results not shown). In those earlier experiments, newly arising mutations were simply allowed to compete among themselves for fixation, which occurred immediately. The absence of a qualitative difference suggests that competition among simultaneously arising mutants is more significant for these results than is competition among mutants appearing sequentially.

Asymmetry in Mutational Effect

The analytical tractability of the geometrical model depends on spherical symmetry in both the fitness function and the distribution of mutational effects. The first condition is equivalent to assuming that all traits contribute equally, on average, to fitness. In this case the variance matrix of the fitness function is a scalar multiple of the identity matrix. The second condition implies that mutations have equal effects, on average, on all phenotypic traits. Formally, the direction of

a mutation is equally likely to lie within one area of a hypersphere as in any other area.

These conditions can be relaxed by assuming that either the mutational effects or the trait axes (but not both) can be rescaled. However, in so doing, the problem becomes much more difficult to solve. It seems unlikely that both simplifying conditions can hold true; certainly there is no reason to think that they must. In the context of the example given in the introduction, both assumptions holding true would be equivalent to a unit change in the *color* trait having equal fitness effect as a unit change in the *size* trait. But, more implausibly, mutations would on average act to change both traits by the same amount. The pleiotropic effects of mutations would be equally distributed across all traits, proportional to the trait's contribution to fitness.

It is therefore reasonable to relax this assumption of uniform or symmetrical pleiotropy. With such a change the distribution of fixed effects does not seem to be affected much – it remains approximately exponential. However, even a modest asymmetry in pleiotropic effects dramatically lengthens the adaptive process. The reason should be fairly intuitive: for some traits the mutations' effects are increased, resulting in fewer beneficial mutations. For the others, the mutational effects are decreased, leading to beneficial mutations with smaller fixation probabilities.

For a symmetrical distribution of mutational effects, between 1 and 1000 mutations must be sampled before one that is sufficiently beneficial to be fixed is found. As many as 10000 mutations must be sampled before enough beneficial

mutations are fixed to move to within 10% of its original distance to the optimum. As the mutations are biased towards affecting particular traits more than others, the number of mutations that must be sampled can increase by up to a factor of 10000.

With some distributions of mutational effect sizes, the time taken to the first fixation increases with increasing asymmetry, but then decreases with further asymmetry. This phenomenon was recognized by Chao (*pers. comm.*), who considered the most extreme case of asymmetry in mutational pleiotropy, in which case the mutation's effects are effectively zero for some of the traits. In this case, the effective number of dimensions has effectively decreased.

If the number of mutations needing to be sampled is much greater than the number of possible mutations (given an organism with a small genome size), a small amount of asymmetry will serve to greatly slow adaptation. This is the unavoidable consequence of the geometrical model's treatment of the mutation-space as continuous. In reality, of course, the number of possible phenotypes, and the number of points in Fisher's phenotype space that an organism can occupy is finite.

Pleiotropic asymmetry greatly increases the number of mutations that need to be sampled before one that is sufficiently beneficial to become fixed arises. Another way of saying this is that the proportion of 'fixing' beneficial mutations is greatly decreased. Some amount of asymmetry in pleiotropy may thus serve to support a major assumption of the Orr model, that is, that multiple beneficial mutations do not exist simultaneously in the population.

Orr's has argued that his extension of Fisher's geometrical model refutes the 'infinitesimal' model (Orr 1998; 1999), in which phenotypic variation is assumed to be the result of an infinite number of genes acting with infinitesimal effect. Orr's prediction of exponential distribution of fixed phenotypic effects does not assume infinitesimal effects: nevertheless it still concludes that mutations with small effects should greatly outnumber those with large effects.

Empirical results from quantitative trait loci (QTL) and other studies generally find that many significant trait differences are caused by a few genes of large effect. These studies do not support the infinitesimal model (see for example (Bradshaw *et al.* 1995; Keightley 1995; Keightley 1998)), and Orr has indicated this as implicit support for the exponential distribution of effects. However, the exponential prediction does not match these data very well either, no matter how robust it is to changes in the model parameters.

The results of this chapter suggest that a geometrical model might be roughly supported by QTL and other data, if the competition among beneficial alleles is taken into account. Even for populations as small as 10^6 with mutation rates of 10^{-8} , the distribution of fixed effects corresponds more closely to the QTL data when clonal competition is included. However, there are some caveats to this conclusion; the assumption of completely monomorphic population is not the only potentially unrealistic assumption of the geometrical model. Other unrealistic assumptions include the smooth fitness surface with a single stationary peak and

the complete lack of epistasis among loci. These features of adaptation in real populations are not accounted for by any variation of the model.

More significantly, perhaps, for organisms with limited genome sizes, the model assumes that the mutational landscape is continuous. In this, Orr's extensions of Fisher's model do not differ from the original assumptions of the model: mutations can be arbitrarily small in effect and there are an infinite number of states in the phenotype space. The approximation of a continuous state-space is especially likely to break down in considering the genotype space of, for example, ϕ X174, which has a genome of roughly 5000 nucleotides.

The long-term success of the geometrical model rests on the fact that it captures essential features of adaptation in spite of its conceptual simplicity. Orr's extensions of the model seem to further this success. Orr's results imply that the model can capture enough of the phenomenon of adaptation to warrant extremely precise predictions.

How accurate are these predictions? It is intrinsically difficult to determine how the features of the model relate to measurable features of adaptation in natural or experimental populations. Although the fixed phenotypic effects might correspond to effect sizes in QTL studies, the correspondence is not absolute. The magnitudes of phenotypic factors in the Fisher-Kimura-Orr model are over many orthogonal traits, while QTL studies look at individual traits. Furthermore, the results of most QTL studies do not support the theoretical results given by the Fisher-Kimura-Orr model, except in the most general and vague terms.

Testing the predictions of the model in natural populations would be very difficult. The most likely experimental system for such a test would be a microbial population, in which the environment and population parameters could be strictly controlled, and the genotype frequencies could be completely monitored. However, as the simulations described here have shown, the properties that make a microbial system attractive (large population size, for example) may violate the assumptions of the model.

Incorporating clonal competition and pleiotropic asymmetry into the model leads to substantial changes in its predictions. Further features of adaptation (such as multiple fitness peaks, epistasis in mutational effects, and the discrete nature of the genotype space) could in principle be included as well, perhaps leading to additional changes in the predictions. In that case, however, a great deal of the conceptual simplicity, generality and elegance of the model would be lost. For that reason, further attempts at general models of adaptation should best be developed from different foundations.

Figure 1: Illustration of the method used for scaling mutational axes. In the first figure, the direction of mutations is distributed uniformly. The second figure illustrates a transformation of the axes determining the direction of the mutation. In this case, the x-axis is dilated by a factor of 1.5 and the y-axis is contracted by 0.5. In the example, the third mutation shown would be deleterious – it would move the population to a point further from the optimum. For this reason, it would not be fixed. A result of this transformation is that more mutations need to be sampled to continue the progress of adaptation.

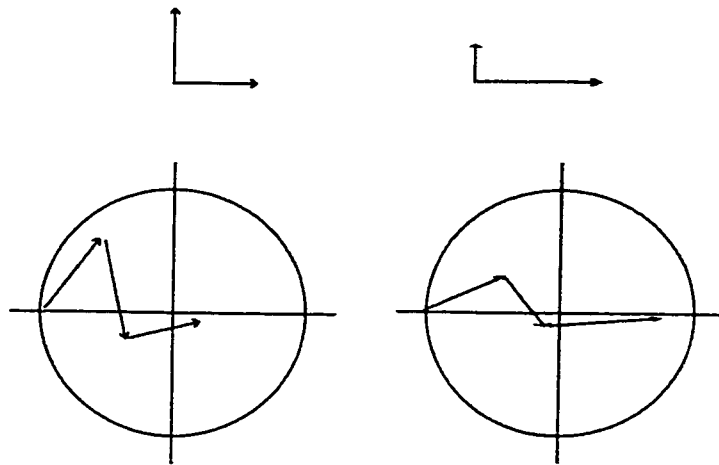


Figure 2: The distribution of phenotypic factors fixed during the course of adaptation. The number of new mutants appearing in the population was determined by the parameter $N\mu$. In these experiments, the underlying size of mutational effects was distributed exponentially with mean 0.1. The curve for $N\mu \ll 0.01$ corresponds to the conditions described in Orr (1998, 1999). In this case the rate of beneficial mutations was sufficiently low that only one was present in the population at any given time. These simulations were done with a phenotype space of 25 dimensions and a population at a constant size of 10^6 .

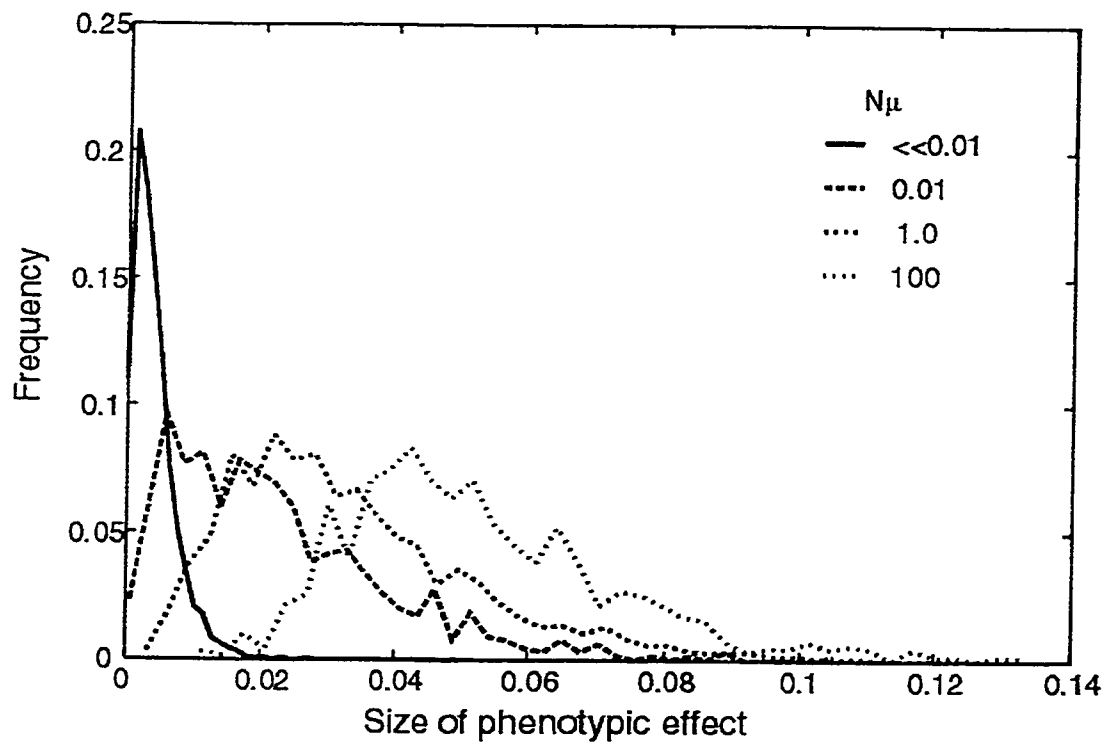


Figure 3: The distribution of selection coefficients of mutations fixed during the course of adaptation. Experimental conditions are as described in figure 2.

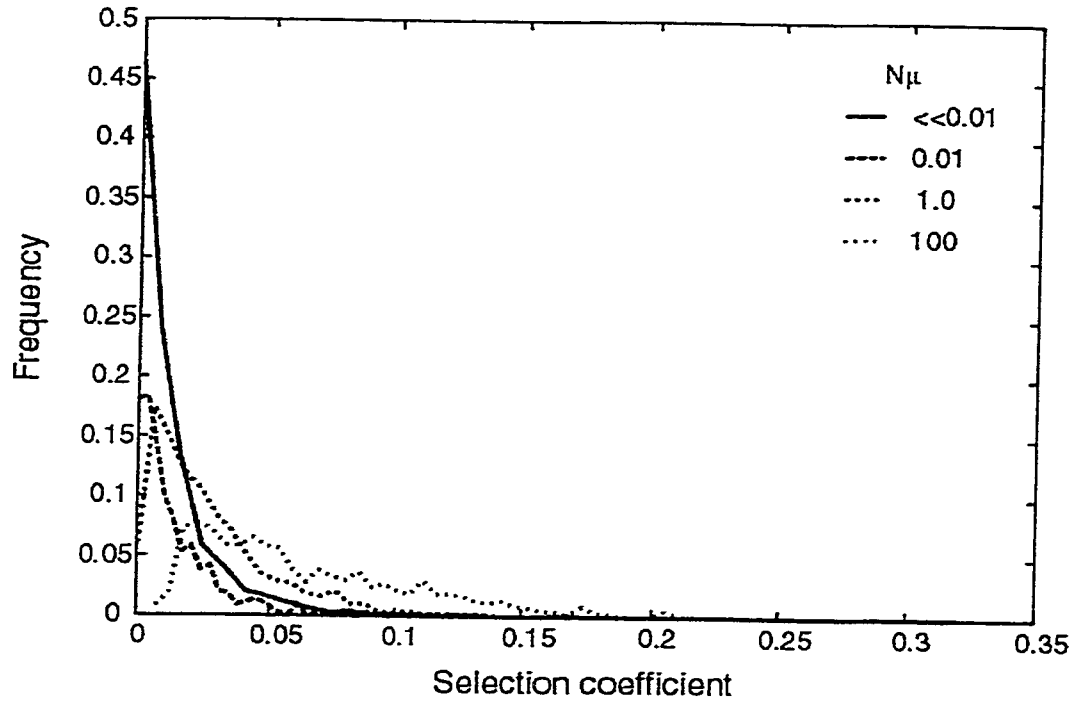


Figure 4: The distribution of phenotypic factors fixed during the course of adaptation. The experimental conditions are as described in figure 2, except the underlying distribution of mutation sizes was gamma, with a mean of 0.1 and shape parameter of 0.5.

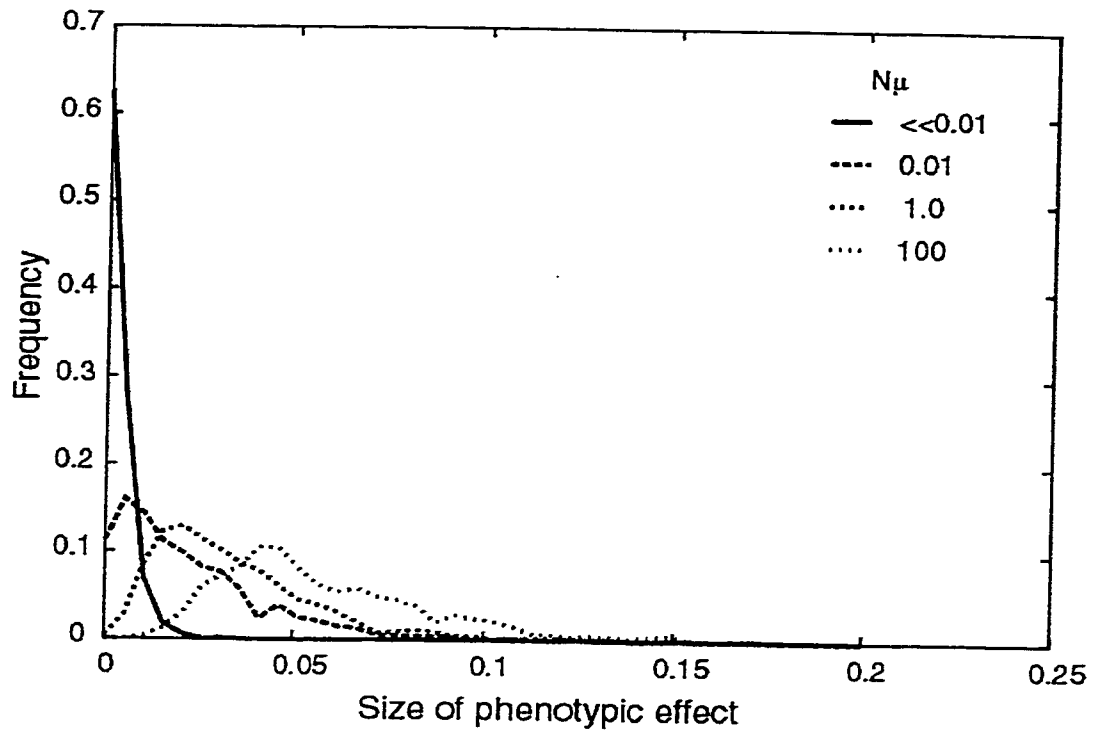


Figure 5: The distribution of selection coefficients of mutations fixed during the course of adaptation. Experimental conditions are as described in figure 4. The plots for $N\mu \ll 0.01$ and $N\mu = 0.01$ largely overlap.

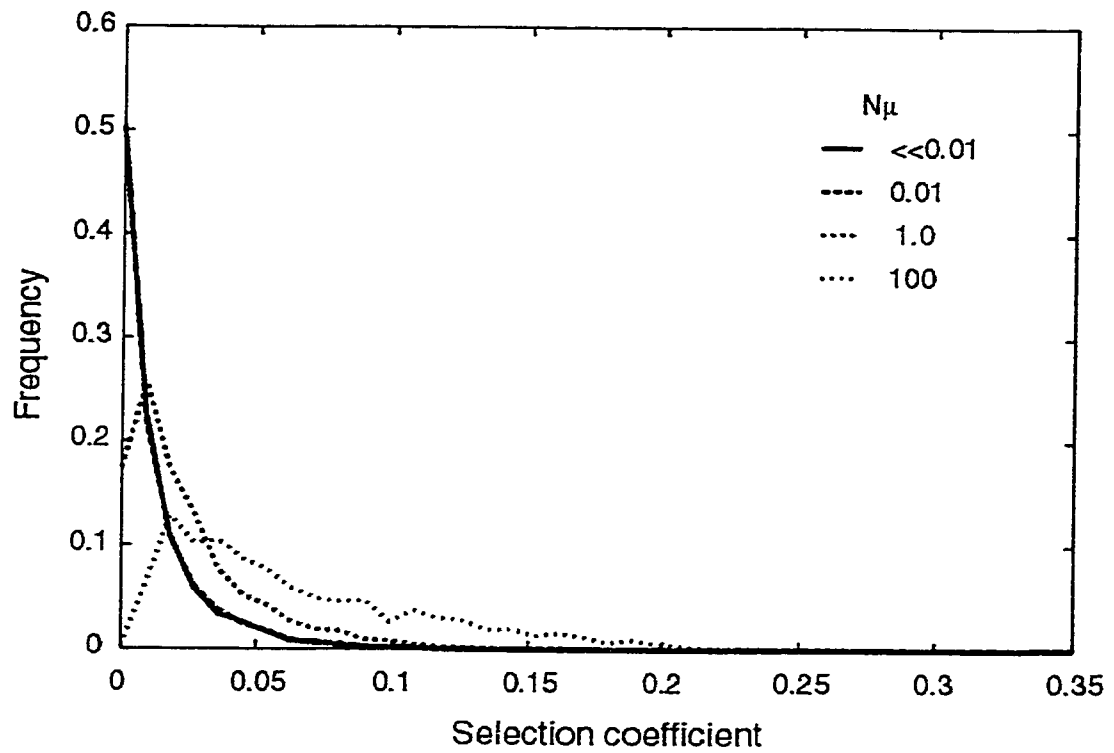


Figure 6: Distribution of effect sizes for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^6 .

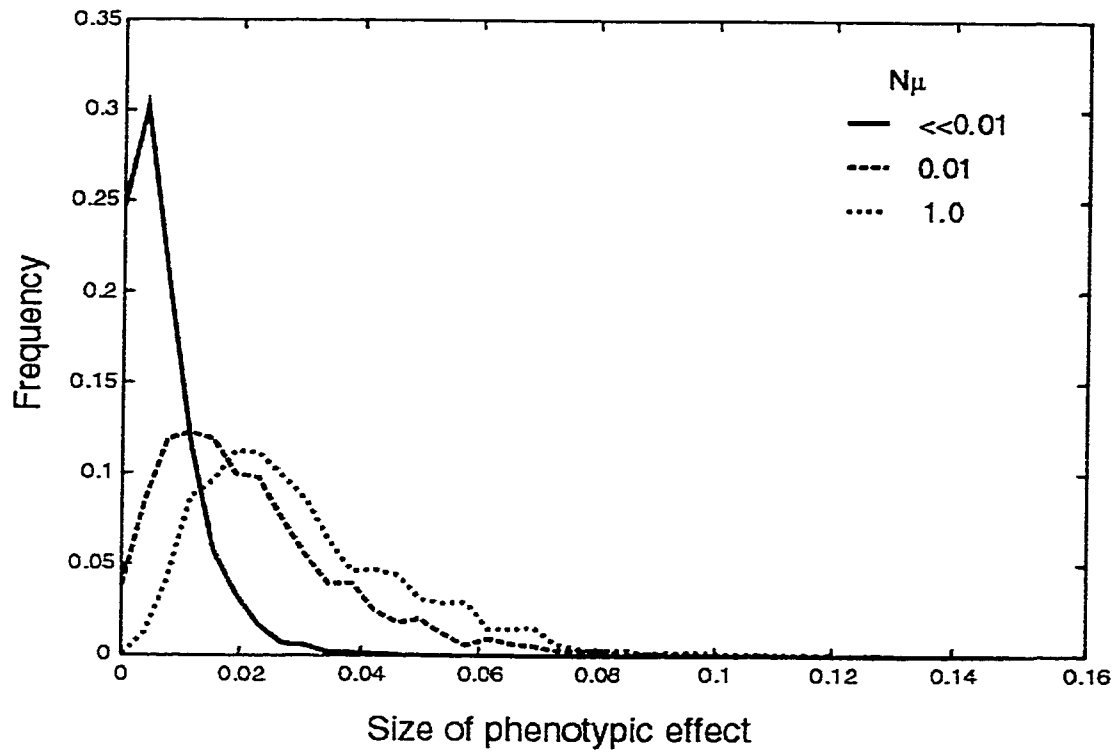


Figure 7: Distribution of selection coefficients for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^6 .

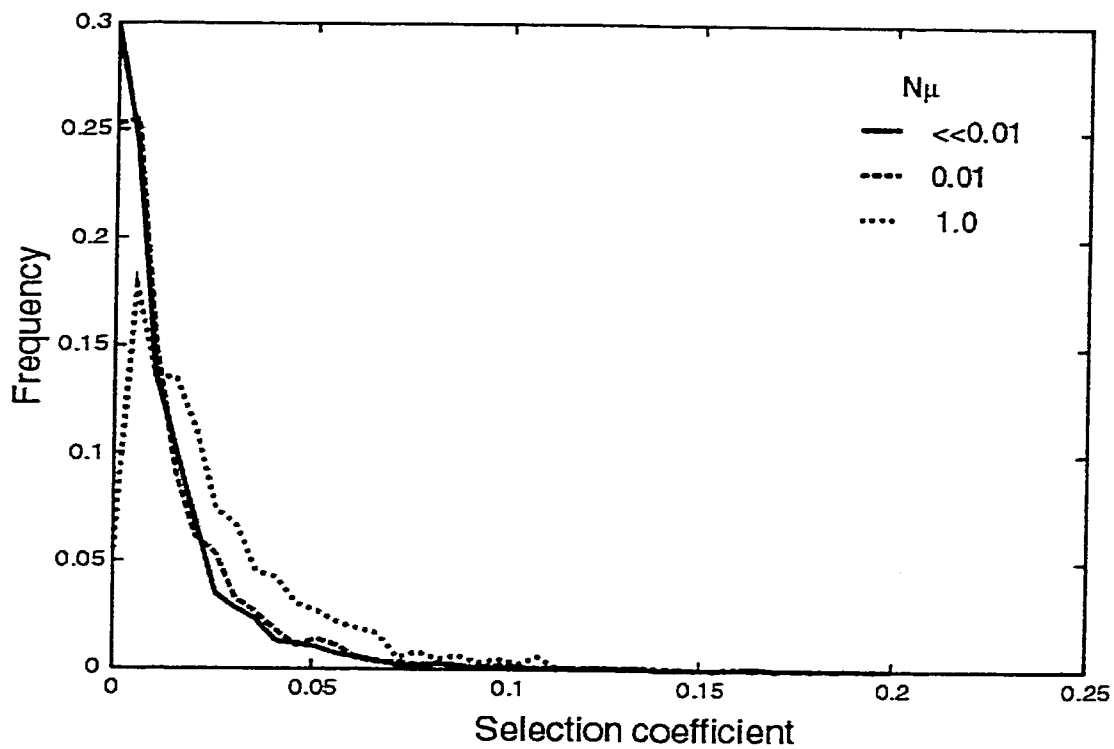


Figure 8: Distribution of effect sizes for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^{10} .

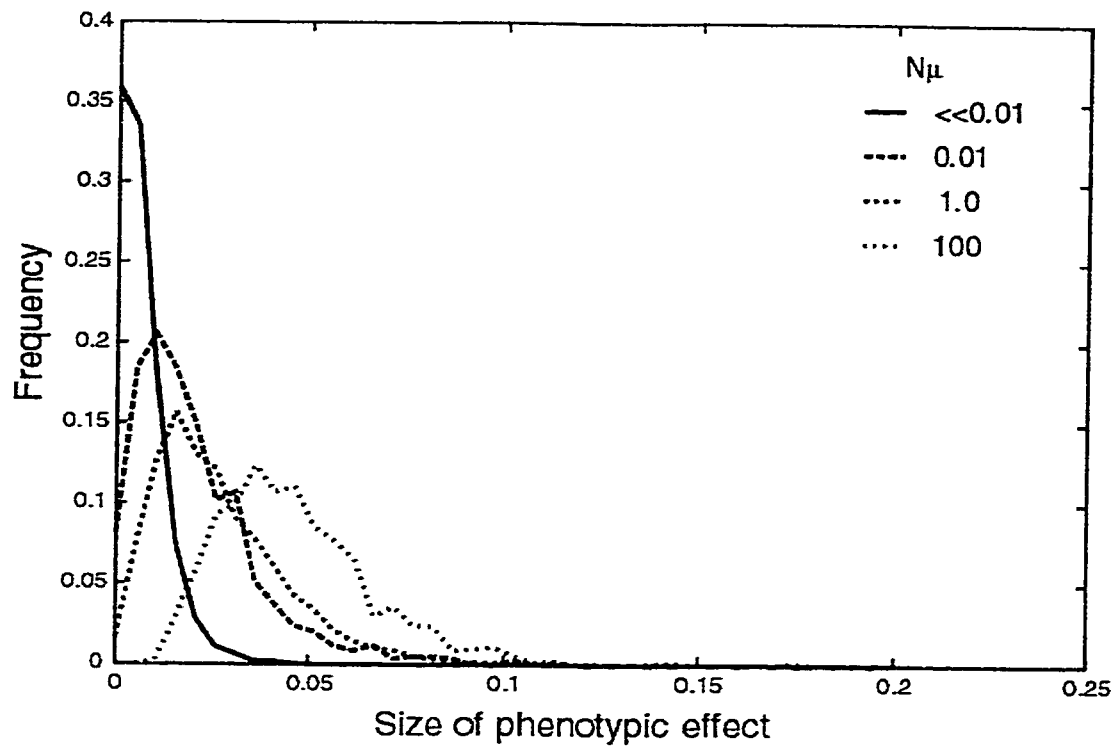


Figure 9: Distribution of selection coefficients for mutations fixed during adaptation. The underlying distribution of mutation effects is an exponential with mean 0.1, and the population size is 1×10^{10} .

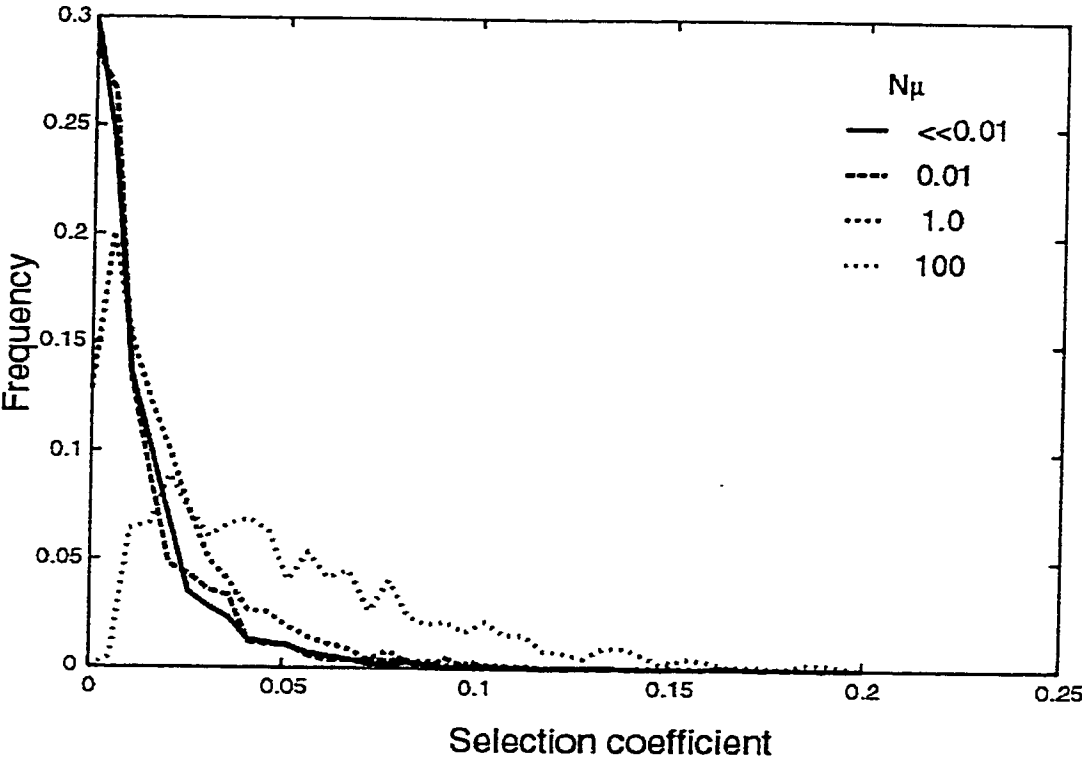


Figure 10: Time to first fixation as a function of variance in mutational axis scaling factors.

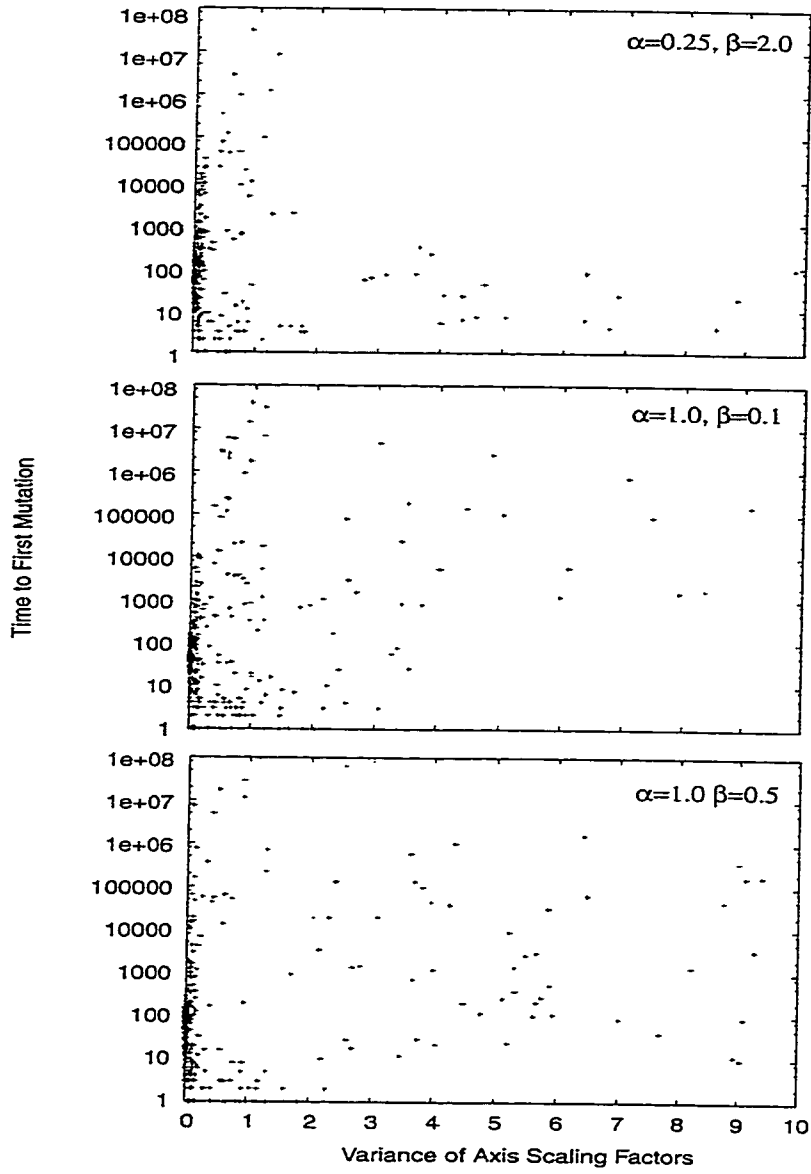
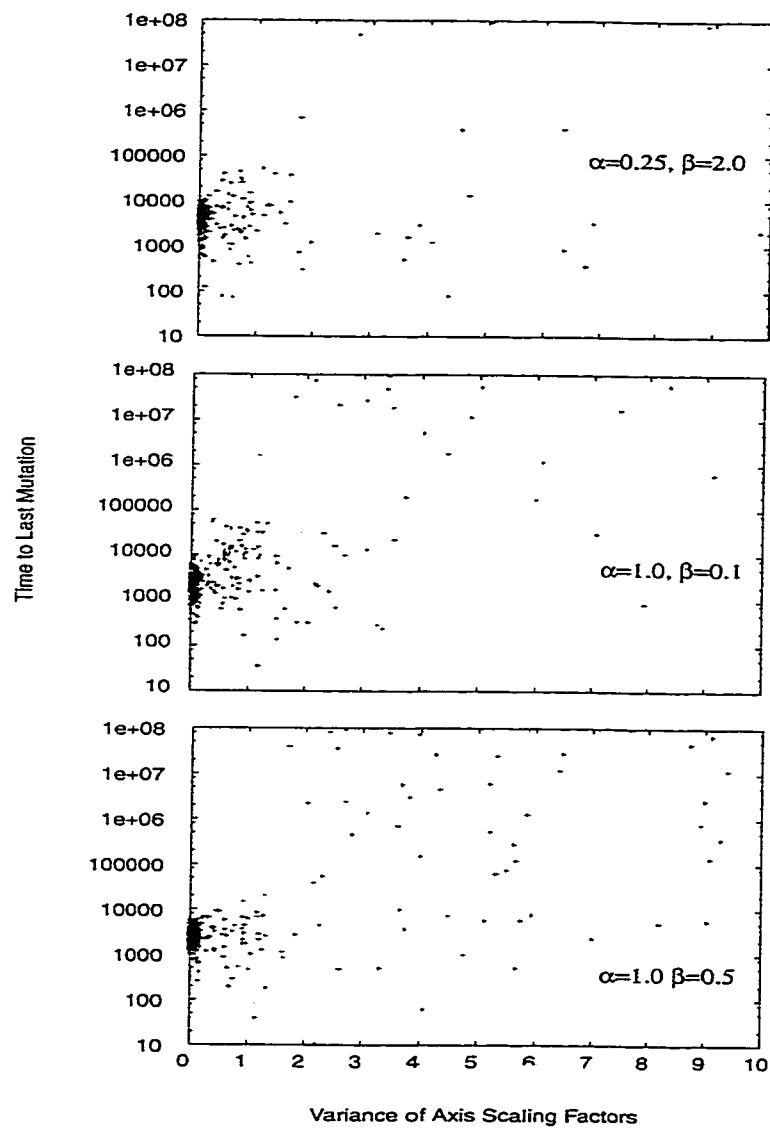


Figure 11: Time to last mutation fixed as a function of variance in mutational axis scaling factors.



Chapter 2: Recovery from Deleterious Mutations

INTRODUCTION

Because of the stochastic nature of evolution, mutations with deleterious fitness effects can become fixed in a population. Failure to respond to fixed deleterious mutations dooms a population to the spiral into extinction of the “mutational meltdown” (Lynch *et al.* 1995; Otto and Whitlock 1997). Populations are limited in the mechanisms they can use to respond to a fixed mutation. The negative effects of the mutation can be suppressed by a mutation elsewhere, or the deleterious mutation itself can be removed by direct reversion.

The most likely of these would seem to be the suppression of the deleterious effects by a mutation at a second site: there are potentially many such sites, while there is at most one specific mutation that results in a reversion. For this reason, second-site suppressors of deleterious effect (or “compensatory mutations”) are thought to be strongly relevant to the process of adaptation (Burch and Chao 1999; Bjorkman *et al.* 2000; Levin *et al.* 2000; Moore *et al.* 2000). Compensatory mutations are beneficial, but must be distinguished from strictly beneficial mutations, which have a positive effect regardless of genetic background (Burch and Chao 1999). Compensatory effects are those that accrue only to specific low-fitness genotypes. They are suppressors of a deleterious allele, and represent interactive fitness effects between mutations that would be deleterious or neutral in an optimal-fitness genome.

Notwithstanding their small relative probability, reversions may offer the best means of recovering the full fitness lost by the accumulation of deleterious mutations. In a fitness landscape with a single optimal peak, reversions present a more straightforward mechanism for mutation to reduce the effects of deleterious substitutions, as they do not require complex epistatic interactions among genes. There is the potential for the entire fitness loss to be regained by reversion, while this is not necessarily the case for compensatory changes.

Here I present a study in which a small viral genome was fixed for several deleterious mutations and allowed to recover via serial passaging of large populations. The goal of the study was to evaluate the mechanism (either reversion or compensatory mutation) by which populations recover fitness from the fixation of deleterious mutations. Throughout the study, the fitness and genotype of the recovering populations were assessed.

METHODS

Strains and General Methods

LB broth was used for growth of cells, both infected and uninfected. The preparation of LB broth was mixed as 10 g/l NaCl, 10 g/l Bacto-tryptone and 5 g/l yeast extract in water, autoclaved, then supplemented with 2 mM CaCl₂. For plates, 15 g/l Bacto-agar was added to LB broth. For phage overlay plating (top agar), 7 g/l Bacto-agar was added to LB broth. Phage stocks were stored at 4°C after an equal volume of Borate-EDTA (BE) was added to the phage suspension in LB. BE consists of 80 mM sodium borate and 80 mM EDTA, pH adjusted to 9.5 with NaOH. Phage were propagated in *Escherichia coli* C.

In order to limit the response of the population to either reversion or compensatory mutation, all strictly beneficial mutations were excluded. This was done by starting with a strain of phage that was at a fitness optimum (that is, maximally fit under the conditions of the experiment). To be certain that I was starting at the optimum, I began the experiment with a line of bacteriophage (A_f for “Flask-adapted Ancestor”) that had been grown under the experimental conditions to a point that further increases in fitness had stopped (Bull *et al.* 2000). The final fitness of the A_f strain was approximately 24 doublings per hour, under the defined growth conditions at 37°C (Bull *et al.* 2000).

Evolution of Mutant Phage of Low Fitness

During the constructing mutation-accumulation lines, it is important to minimize the effects of selection. The best way to do this is to keep population size small, so that deleterious mutations may be fixed. I adopted a strategy of chemical mutagenesis coupled with single-plaque bottlenecking and selection for small plaque size to introduce fixed deleterious mutations into the ancestral stock (Chao *et al.* 1992; Stephan *et al.* 1993). The basic cycle of propagation was to mutagenize the phage with hydroxylamine, plate, choose a small plaque, and grow the phage to increase number. This cycle was repeated until fitness was sufficiently low and the phage had accumulated enough mutations to require multiple steps in recovery. During this cycle, selection only has the opportunity to act during the phase of growth that occurs as a single virion grows to a plaque.

The mutagenesis protocol followed that described in (Tessman 1968). Phage were incubated 90 min. in 250 mM hydroxylamine, in 1 mM EDTA pH 7

at 37°C, followed by dilution with LB and immediate plating on LB with agar. This procedure resulted in a survival ratio of 1%.

Starting with A_f , two lines were subjected to repeated cycles of mutagenesis and single-plaque propagation. Plaques were observed at each stage to be progressively smaller. At the end of five such cycles, plaques were tiny, and no further size decrease was attainable. This most likely indicates that a lower limit of fitness had been reached, at which further fitness loss would have resulted in an inability to recognize and recover plaques.

Recovery Conditions

Phage recovery was selected by serial passage of phage at large population size. The conditions of each passage of recovery consisted of growth in rich medium at 37°C, in cells that had been taken to the logarithmic phase of growth. Bacteria were grown from a frozen stock for 1 hr. to a density of $2-4 \times 10^8$ cells/ml in 10 ml LB contained in 125 ml flasks, aerated at 200 rpm in an orbital water bath. Between 10^5 and 10^6 phage were added and grown for 60 min., before treatment with chloroform. (In the earliest stages of recovery, low titers limited the amount of phage available to add. At these stages no fewer than 10^3 phage were used). Growth was limited to one hour or less, in order to avoid the transition to the stationary phase of bacterial growth, and to keep the multiplicity of infection (MOI) relatively low. Free phage density was not allowed to exceed 10^9 phage/ml, to ensure that the ratio of infected to uninfected cells remained less than unity. (The burst size of wild-type $\phi X174$ is approximately 100 phage per infected cell).

Fitness Assay

Fitnesses were calculated from phage growth rate over 40 min. and under the conditions used for recovery. These were the same conditions used to adapt the starting phage (A_f) prior to its subjection to mutagenesis (Bull *et al.* 2000). The measure of fitness used is doublings/hour: $w = \frac{3}{2} \log_2 \left[\frac{\text{phage density at 40'}}{\text{phage density at 0'}} \right]$.

This fitness assay is less precise than the competition-based assays used by, for example, Burch and Chao (2000), but it has the advantage of being an absolute measure of fitness, not scaled to the fitness of any standard competitor. For the large changes in fitness observed here, our measure is more than adequate. Furthermore, it has been shown (Paquin and Adams 1983) that competition-derived fitnesses are not necessarily transitive among subsequent isolates from a microbial lineage because of interaction between the strains.

Sequence Analysis

For mutagenesis and recovery passages, sequencing was generally done from whole cultures; the resulting sequence thus represents the majority consensus. Site-directed mutants were sequenced individually, as were occasional isolates from whole cultures.

Phage cultures and isolates were sequenced from PCR products using cycle sequencing with fluorescent-labeled (BigDye) dideoxynucleotides to terminate the extension. Terminated PCR products were visualized on an ABI 377 automated sequencer.

Sequencing primers were positioned across the genome at approximately 500 base intervals. Sequencing was generally done from the minus-strand

template, supplemented with sequences from the other strand when needed to resolve ambiguities.

Sequences were aligned to the known ancestral sequence, and differences between them were noted. When PCR products of phage lysates were sequenced, peaks from the sequence of a population of viruses could occasionally be resolved into major and minor components, suggestive of polymorphism within a population. For some genotypes, the presence of multiple peaks was used to confirm suspected polymorphisms, but, in general, the presence of multiple peaks was not by itself taken as an indication of polymorphism. Genotyping of phage isolates by oligo hybridization followed the protocol of Crill *et al.* (2000).

Site-Directed Mutagenesis

Specific mutations were introduced into specific phage genotypes using a DNA Polymerase I extension reaction from a mutagenic primer. These primers, extending 250-600 nt, were constructed by PCR amplification from strains carrying the desired genotype within the amplified region. Each fragment was purified using a Promega PCR prep purification column and resin. The concentration of mutagenic primer was quantified by absorbance at 260nm. Table 1 lists the primers used, their genotypes, the template genomes and the PCR primers used to construct them.

DNA from the phage genome to be mutated was purified by extraction with phenol and phenol/chloroform, followed by isopropanol precipitation. After preparation, the effective amount of DNA recovered was estimated by transformation into electrocompetent *E. coli* cells. The number of plaque-forming

units per unit volume was used to calculate the minimum amount of mutagenic primer to use. A minimum of 100-fold molar excess primer to template was used for the mutagenesis.

When the concentrations of the mutagenic primer and template were determined, I annealed the two in PolI buffer (New England Biolabs). Annealing consisted of heating a mixture of the template and primer to 80°C, then cooling in small increments to room temperature, then to ice. DNA Polymerase I and deoxynucleotides were added to the ice-cold solution, which was then elevated to 37°C. After allowing the extension reaction to proceed for 1 hr, the product was re-extracted with phenol, ethanol precipitated and washed.

The products of the extension reaction were transformed into *E. coli C* using electroporation, and the transformation mixture was plated onto LB/agar plates in top agar containing additional *E. coli C*. The results of this plating gave an indication of a titer for the transformation mixture, which was then plated onto large LB/agar plates at a density of approximately 1000 plaques per plate.

Transformants were screened by oligonucleotide hybridization. Instead of transferring plaques into microtiter plate wells (as in Crill *et al.* 2000), nitrocellulose lifts were made using the method of Maniatis *et al.* (1982). Plate lifts were otherwise treated as described by Crill *et al.* (2000).

Using a *mutD* Host to Enhance Mutation Rate

To examine the possibility that mutation rate limits the rate of adaptation, I passaged isolates from the recovery line in a mutator strain *E. coli C mutD*. The mutation in this strain is in the 3' exonuclease editing function of DNA

polymerase, and results in an increase in mutation rate of approximately 1000-fold. Isolates chosen for additional recovery through *mutD* were from recovery passages 1, 29 and 41 from the mutated lineage A3.

After growing phage for 1-hour in *mutD*, each isolate was passaged up to seven one-hour cycles in non-mutagenic *E. coli* C to enable the ascent of the highest fitness clones. These descendant strains were sequenced, and their fitnesses evaluated.

Recombination of Mutant Lines

In order to get an idea about how many reversions might be sufficient to allow fitness recovery, I recombined the two mutant lineages and selected for the highest-fitness recombinant. Recombination was achieved by coinfection of the two mutant lineages in *E. coli* C at high MOI, and subsequent passage of the resulting culture by the protocol described for fitness recovery.

RESULTS

Mutation Accumulation

Of the three lines mutagenized by HA and single-plaque bottlenecking, two dropped significantly in fitness (as estimated by plaque size) over five passages. Further passages did not decrease plaque size: this was probably due to the fact that there is a lower limit to the fitness needed to form a visible plaque. For the two lines that decreased significantly (A1 and A3), I took the fifth mutagenic passage as the starting point for the beginning of the recovery experiments. These end-points, and the intermediates (A1-1, A1-2, A1-3, A1-4, A1-5 and A3-1, A3-2, A3-3, A3-4, A3-5) were sequenced, and the fitnesses were

determined by the standard liquid-culture fitness assay described above. The decrease in plaque size for the third line (A2) was much less than for A1 and A3, and the line was not investigated further.

In the two retained accumulation lines (A1 and A3), 13 and 16 mutations were fixed, respectively. With one exception (nucleotide 4494, resulting in a silent substitution in gene A), the mutations were unique to each lineage. Each step of the mutation accumulation carried more mutations than its ancestors (tables 2 and 3). All of the mutations that accumulated during the five mutagenic cycles in each line were single nucleotide substitutions of C→T. This is consistent with the use of hydroxylamine, which deaminates cytosine residues to uracil (Pol and Arkel 1965; Tessman 1968). Tables 2 and 3 describe the location and timing of the mutations, the genes in which they occurred, and the amino acid substitution, if any, that resulted. To determine if these mutations were randomly distributed across the genome, I performed a chi-square test, partitioning the genome into seven segments by gene, so that untranslated sequences were assigned to the gene immediately downstream, and overlapping genes were ignored. The results of the chi-square test (not shown) indicated that the distribution of mutations is not significantly different from random for each of the mutation-accumulation lines.

Fitnesses of the mutation accumulation lines declined with each step, except that A3-5 was slightly higher than its immediate ancestor A3-4 (table 3). Most of the fitness drop occurred during the first 2-3 passages in both lines.

Recovery from Mutagenesis

Mutant lineages A1-5 and A3-5 were allowed to recover in fitness via repeated passage in flasks. For each lineage a duplicate line of recovery passages was done, under similar conditions, although for a shorter number of recovery passages. The recovery lineages were A1-5R and A1-5S, and A3-5R and A3-5S. A1-5R was recovered for 29 passages, A3-5R were recovered for 40 passages, while A1-5S and A3-5S were recovered to the 17th passage.

The entire genome of each recovery lineage, taken from whole culture, was sequenced at various points throughout the recovery. Whole-culture sequences were compared at the end-points, and at passage 17 with the sequencing of two single isolates. In all cases, the sequences of the isolates matched that of the whole culture.

At the end of the recovery passages, A1-5R40 and its replicate A1-5S17 each had the same single nucleotide substitution (a reversion), and had both increased in fitness. A3-5R40 had two substitutions (both reversions) and a large increase in fitness. Its replicate, A3-5S17, had no changes from the unrecovered mutant, and did not have a substantial increase in fitness from its starting point.

Because it had the most substitutions and largest fitness gain during recovery, lineage A3-5R was examined in greater detail than the other lineages. Sequences and fitnesses for A3-5R are shown in detail in table 4. The results for the other three lineages are summarized below.

In contrast to lineage A3-5R, the replicate (A3-5S) had no new mutations or reversions. After 6 passages, the fitness was 9.6 ± 0.5 and after 11 passages was

10.3±0.3. These fitness values are not substantially different from that of the unrecovered mutant with the same genotype.

After 6 recovery passages of mutant lineage A1-5, a single reversion was observed, which persisted alone across all remaining passages (until the 29th). The reversion, at site 1061, restored the identity of residue 21 of gpF from the mutant phenylalanine to the wild-type leucine. The fitness of the single-reversion genotype was 13.3±0.25, and the fitness of the lineage never got above 14.4±0.06. The replicate recovery lineage had the same reversion, and the fitness of the last passage was 12.9±0.1.

Identification of Mutant Genotypes

For lineage A3-5R the frequencies of mutant genotypes were evaluated at recovery passages 3, 6, 9, 12 and 15. These frequencies are shown along with the population fitnesses in figure 12. The plot of mutant genotypes shows a transient polymorphism of the new mutation at site 1628, coincident with a transient polymorphism of the reversion at site 2252. These polymorphisms are swept from the population by the double revertant at sites 632 and 2252. No further changes in genotype frequency were observed for the remainder of the recovery.

Recombination Between Mutant Lines

The fitness of the recombined A3-5 and A1-5 lineages increased after 9 passages to near wild-type (22.7±0.72). An isolate of this last passage (called ρR9) was sequenced, and found to contain six of the 25 mutations present in mutant isolates A1-5 and A3-5. These mutations (those at 1393, 3331, 3485, 3550, 4493 and 4629) imply a recombination event with crossover points

occurring between 1297 and 1393, and 4474 and 4540. Of the six mutations present in the recombinant, three are silent, and two of the remaining three missense mutations arose late in the course of mutagenesis, when fitness decreases were small. Not surprisingly, this recombination event yielded much higher fitness than that of any recovery line.

Targeted Mutagenesis

The six genotypes of A3-5 constructed by site-directed mutagenesis were sequenced in their entirety to confirm identity, and assayed for fitness. They are shown, with their fitnesses, in table 5. All mutants exhibited an increase in fitness. Thus all were beneficial, though to different degrees.

DISCUSSION

Pattern of Accumulation of Fixed Deleterious Alleles

During the mutagenic passages, both lineages accumulated fixed mutations, and decreased substantially in fitness. It is evident that not all of the mutations contributed to the fitness decline, but rather that some are compensatory for others. (This is almost certainly true of the mutation C4015T in A3-5, which has a fitness that is slightly higher than that of A3-4). The fact that most of the fitness decline occurred in the early passages may reflect either (1) a bias in selection of mutant plaques or (2) a pattern of diminishing returns of deleterious fitness effects (*e.g.* Bull, *et al.* 2000).

Recovery from Deleterious Mutations

In three of four lineages, nucleotide substitutions occurred during the recovery passages, with a coincident increase in fitness. All substitutions observed were reversions of mutations fixed during the mutagenic passaging. Two replicate recovery lineages (A1-5R and A1-5S) had identical reversions and nearly identical increases in fitness.

In a third lineage (A3-5R), a compensatory change was present in early passages, but was quickly lost when a reversion swept through the population. Genotypically, isolates from the final passage differed from the initial mutant only at two sites that had reverted to the ancestral state. The site-directed mutants demonstrate that the compensatory change at site 1628 had a larger fitness effect than either of the reversions at sites 632 and 2252. However, a virus with both reversions had significantly higher fitness, and rapidly out-competed the mutant with the single compensatory change.

None of the lineages approached the fitness of the ancestral strain. This may be due to either of two possibilities. First, the population might be at a genuine fitness plateau, having reached an alternate peak from that of the ancestral strain. That is unlikely: further reversions introduced by site-directed mutagenesis (*e.g.* the single reversion of site 3154) increase the fitness substantially.

More likely, further progress is mutation-limited. Presumably, an additional reversion such as that at 3154 would enter the population and fix, given either enough time, or an increased mutation rate. Recovery in a mutagenic host

shows that with an increase in the rate of mutation, further fitness gains are in fact possible, by the reversion of one of the mutations.

The reversion seen in the mutagenic host was accompanied by a silent substitution in gene F, at site 1702. These substitutions arose from passaging of cultures A3-5R29 and A3-5R41, each predominantly composed of phage with the same genotype. That the substitution at site 1702 arose twice independently may reflect a strong interaction with site 4474. The reversion at 4474 was expected to have a large fitness effect: the mutagenic passage at which the mutation fixed had a large fitness drop. However, reversion of the site (along with that of a neighboring site, 4329) had a negligible effect on fitness. The silent change at 1702 may increase the fitness gain of a reversion at 4474. If this is the case, the mutations found in the mutagenic recovery passages might represent an alternate fitness peak.

Another, more prosaic, explanation for the simultaneous occurrence of these sites in the mutagenic host might simply be a bias in the mutational distribution of *mutD*.

Microbial Adaptation and Fisher's Model

Fisher's geometrical model (Fisher 1930) has been presented as the possible basis for a general model of adaptation (Orr 1998; Orr 1999), and as a framework for understanding the distribution of beneficial mutational effects. There are aspects of the model and its derivatives that make it appealing for asexual microbial populations in particular. It has important implications for mutation load and for the long-term health of populations, particularly those that

are asexual. The application of the model to microbial populations is also helped by the easy testing afforded by microbial systems.

The model assumes that a population adapts towards an optimum by a number of steps in a continuous p -dimensional phenotype space. Because the phenotype (and mutational) space is continuous, it also assumes that there are an infinite number of routes by which adaptation can occur. Specifically, if a population is displaced from an optimum by the fixation of even one deleterious mutation, it should be able to return by any of an infinite number of sequences of mutations. The adaptation of the population is extremely unlikely to take place by reversion of the mutations that took it to the sub-optimal point. Mutations that return a population to its optimum by alternate pathways are by definition compensatory: they increase the fitness of a genotype, but conditionally. A consequence of Fisher's model is therefore that compensatory changes are the exclusive mode of fitness recovery from the fixation of deleterious mutations.

Factors Affecting Compensatory Adaptation

At some point in the continuum of genome sizes and genetic complexity, one might expect that the number and effect of potential compensatory changes becomes less than the number and effect of reversions. As pointed out by Levin *et al.* (2000), and discussed at some length by Burch and Chao (2000), the relative importance of compensatory change depends upon several population genetics parameters. Strength of selection should be critical to the relative abundance of compensatory changes and reversions: under weak selection, the fitness difference between points on the surface is decreased, so there are more near-optimal points,

and more accessible paths to the optimum. To the extent that it affects the strength of selection, effective population size should thus be critical to the relative importance of compensatory change (Novella *et al.* 1995; Novella *et al.* 1995; Elena *et al.* 1998).

If the population is considered as a collection of genotypes (rather than as a point in phenotype-space), the mutation rate should also affect the prevalence of compensatory adaptation. In a large population with a high mutation rate, there is a significant probability that all possible neighboring genotypes, including that of the revertant, can be sampled.

The genome architecture also dictates in part whether recovery can occur *via* reversion or compensatory change. If mutational effects are exclusively additive, compensatory change is impossible. Similarly, the type of fixed deleterious mutations that are responsible for a fitness drop determine whether reversion can occur. For example, deletions and insertions will probably be much less likely to spontaneously revert than will a deleterious point mutation.

Is Microbial Adaptation Primarily Accomplished by Compensatory Mechanisms?

In most recent studies (including all of those cited above), compensatory changes have been found to be the predominant factors responsible for adaptation. The picture emerging from this apparent prevalence of CA is one of a fitness surface with a multitude of potential paths between points. This suggests that the problem of sub-optimal fitness has many solutions – that microbial organisms generally have a high degree of genomic versatility.

In contrast to the studies cited above, experimental evolutionary studies using $\phi X174$ suggest a more constrained adaptive landscape. For example, Crill *et al.* (2000) found a set of residues that responded in a near switch-like fashion to transfers between hosts. (In chapter 3, I describe the response of ϕX to strong selection for escape from autoinhibitory sequences).

Since the relative contributions to adaptation of reversion and compensatory change in adaptation turns on specific environmental, genomic and population parameters, there may be no general set of conditions, there may be no sense in which compensatory change is consistently “more relevant” to adaptation than is reversion. However, there are certain extreme conditions that are interesting for biomedical reasons and that are accessible to laboratory manipulation. Strong selection, in particular, is a condition faced in epidemiological contexts. Likewise, large values of $N\mu$ – the product of mutation rate and population size – is relevant to considerations of microbial evolution, particularly in pathogens.

This study differs from those discussed above in that all of the following properties hold true:

1. Selection is directional, uniform over time and *strong*;
2. The displacement from the optimum is *large*: more than a single mutation away;
3. The displacement from the optimum is in principle *reversible*: all mutations contributing to the displacement are single nucleotide transversions of C→T;

4. Population size is *large*: although the mutation rate is smaller than for RNA viruses, more potentially beneficial mutations may be sampled;
5. The population started at a defined optimum: adaptation consisted entirely of recovery to that optimum.

I found that, in contrast to most recent studies, adaptation occurred primarily by reversion of the original deleterious point mutations. Compensatory changes, when they occurred, were transitory and replaced by higher-fitness revertant genotypes.

Reversion as the Primary Mode of Fitness Recovery in ϕ X174

This study demonstrates that under strong selection, reversions are the main mechanism of adaptation in ϕ X174. That this conclusion differs both from those suggested by Fisher's model and from empirical studies, can be attributed to several factors.

First, the fixed deleterious mutations were single nucleotide substitutions. Reversions were in principle much easier than if the mutations had been insertions or deletions, as were seen in Escarmís *et al.* (1996; 1999).

Second, also *contra* Escarmís *et al.*, the ancestor was fully adapted to the passage conditions. The foot-and-mouth disease virus that Escarmís *et al.* (1996) used in their mutation accumulation experiments was not fully adapted to the passaging conditions. The recovery line described in Escarmís *et al.* (1999)

achieved a fitness much higher than the ancestral strain, indicating that strictly beneficial mutations were included in the recovery.

Third, the large population size reduces the effect of drift and increases the relative importance of selection. Recoveries that go through population bottlenecks (as in Burch and Chao, 2000) have reduced effective population sizes, and the fixed mutations are more likely to be the result of sampling error. If the deleterious mutations are of large effect, reversions may be of larger effect than any possible compensatory change. In regimes of strong selection, these would be more likely to become fixed. The observed plateau in fitness may be due to the fact that most of the large-effect mutations had at that point reverted. For further progress, it might be that small-effect compensatory changes can compete with reversions of the remaining small-effect mutations. If that were true, one might expect to see a prevalence of reversions early in the recovery, and a prevalence of compensatory changes later.

Finally, the genetics of the phage may be such that there are not many mutations available to ameliorate the effects of the fixed deleterious mutations. In the recovery, one compensatory change was observed. However, the main argument for an expected prevalence of compensatory change is that compensatory sites are much more abundant than the single reversion. If the geometrical model were to describe microbial adaptation, an infinite number of pathways between low-fitness and high-fitness populations would be possible, and the fitness steps available to the recovering population would be distributed continuously. In contrast to the abundance of potentially compensating sites in *E.*

coli, for example, the ϕX genome may be so constrained that this prediction does not hold.

Table 1: Mutagenic primers, listed according to position in the ϕ X genome. The PCR primers used for synthesis (“PCR1” and “PCR2”) are from the pool used for sequencing. Labels for these primers refer to the approximate location in the genome. PCR1 primers are parallel to the positive strand, and PCR2 primers are parallel to the negative strand. PCR primers are between 12 and 17 bases long.

Primer	Template	Genotype	PCR1	PCR2	Approx. Size	Min. T_m
μ 2	FANC	632C	521	939	418	63.1°C
μ 5	A3-5R6	1628C	1440	2000	560	49.1°C
μ 8	FANC	2252C	1923	2500	577	46.9°C
μ 12	FANC	3042C	2900	3150	250	56.7°C
μ 14	FANC	3042C 3154C 3365C	3381	3849	481	56.7°C
μ 27	A3-5R29	4329C 4474C	4240	4753	513	

Table 2: Mutation accumulation in line A1.

Site	Mutagenic Passage					Gene	Change
	1	2	3	4	5		
491	-	-	-	+	+	D	S→L
722	-	-	+	+	+	D/E	Silent
771	-	+	+	+	+	D/E	R→C/Silent
852	-	-	-	+	+	J	S→F
1061	-	+	+	+	+	F	L→F
1297	+	+	+	+	+	F	Silent
1538	-	-	-	+	+	F	L→F
3331	-	-	-	-	+	H	A→V
3485	-	+	+	+	+	H	Silent
3550	-	-	-	+	+	H	T→I
4493	+	+	+	+	+	A	Silent
4540	-	-	-	-	+	A/A*	A→V
5319	-	+	+	+	+	A/A*/B	P→S/T→I
w	23.5	10.1	8.9	8.7	8.5		

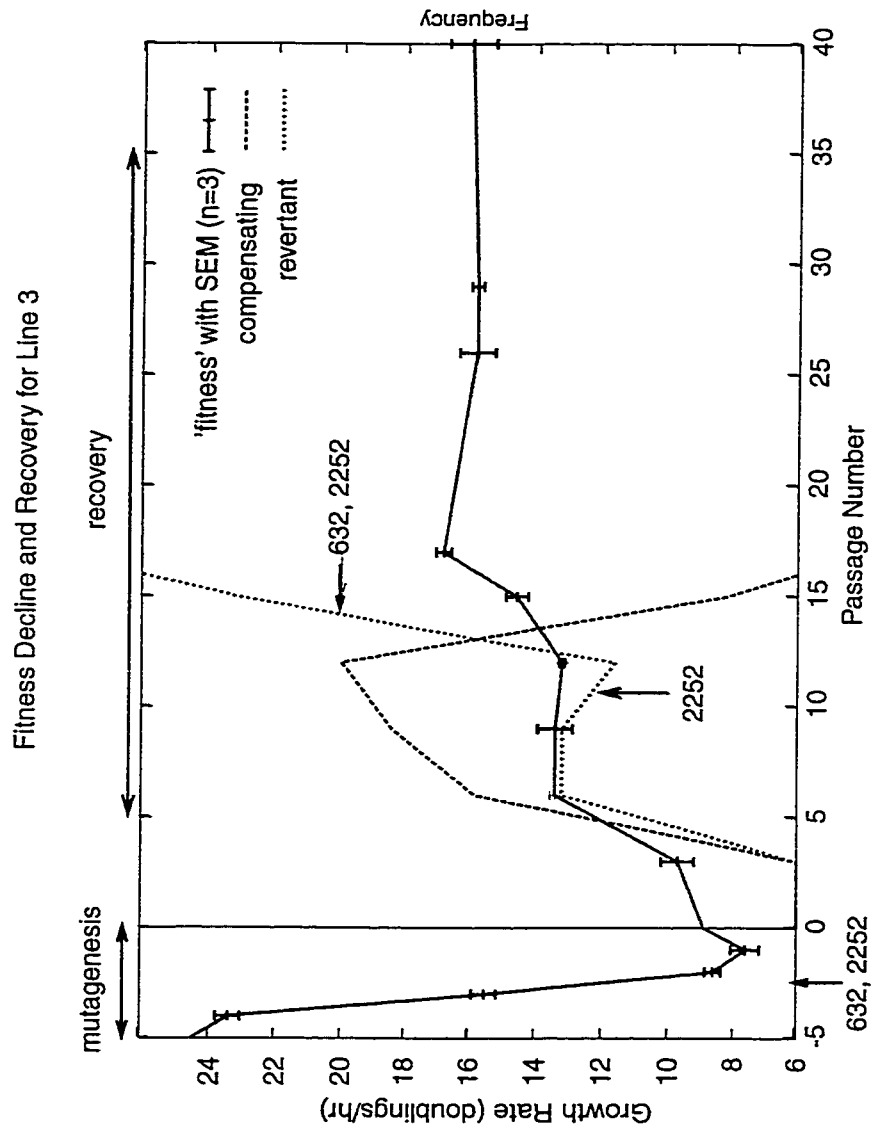
Table 3: Mutation accumulation in line A3.

Site	Mutagenic Passage					Gene	Change
	1	2	3	4	5		
632	-	-	-	+	+	D/E	S→L
1393	+	+	+	+	+	F	Silent
2252	-	-	+	+	+	F	P→S
2769	-	+	+	+	+	G	Silent
2788	-	+	+	+	+	G	L→F
3042	-	-	-	+	+	H	L→F
3154	-	+	+	+	+	H	A→V
3365	+	+	+	+	+	H	Silent
3553	-	+	+	+	+	H	A→V
3896	-	-	+	+	+	H	Silent
3922	-	-	-	+	+	H-A	(intergenic)
4015	-	-	-	-	+	A	A→V
4329	-	+	+	+	+	A	R→C
4474	-	-	+	+	+	A	S→F
4493	-	+	+	+	+	A/A*	Silent
4629	+	+	+	+	+	A/A*	P→S
W	23.4	15.5	8.6	7.6	8.9		

Table 4: Genotype and fitness for cultures and isolates from mutagenesis and recovery lineage A3. Sites of silent substitutions are shown in light-colored type, "." indicates no change in state from the previous recovery passage, "P" indicates a polymorphism at that site. Single plaques isolated from a culture are indicated by (i). Shaded cells highlight the substitutions presumed to be responsible for fitness changes.

Genotype	Site																\bar{w}	sd				
	632	1393	1628	1702	2252	2769	2788	3042	3154	3365	3553	3896	3922	4015	4329	4474			4493	4629		
Gene	D/E	F	F	F	F	G	G	H	H	H	H	H	--	A	A	A	A/A*	A/A*	A/A*			
Af(-)	C	C	G	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
mut(+)	T	T	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Af(-)	S	I	D	R	P	D	L	A	D	A	G	--	--	A	R	S	S	S	P			
mut(+)	L	J	H	R	S	D	F	F	V	D	V	G	--	V	C	F	S	S	S			
Af																				24.5		
Mutagenesis																						
3-1	+									+										+	23.4	
3-2	+									+										+	15.5	
3-3	+				+					+										+	8.6	
3-4	+				+					+										+	7.6	
3-5	+				+					+										+	8.9	
Recovery																						
R3	+																				9.7	0.5
R6	+		P																		13.4	0.5
R9	+		P																		13.4	0.5
R12	-		P																		13.2	0.1
R15	-		-																		14.6	0.3
R26	-		-																		15.8	0.6
R29	-		-																		15.8	0.2
R40	-		-																		16.0	0.7
R40(i)	-		-																		16.0	

Figure 12: Fitness profile for recovery of A3-5, showing frequency of early genotypes.



Chapter 3: Adaptation of Bacteriophage to Escape from Autoinhibition

INTRODUCTION

The Evolution of Microbial Drug Resistance

With the development of antibiotics, and sanitation and vaccination programs, twentieth century humans faced for the first time the possibility of a triumph over microbial infection (Lederberg 2000). However, the stunning medical successes are proving to be ephemeral. Because of recent microbial adaptation, human medicine and agriculture are on the verge of revisiting the dangerous realm of pre-twentieth century medicine. The success of antibiotics in treating bacterial infection has led to their overuse, and this, combined with agricultural and domestic applications, have changed the environment in which bacteria live. The ubiquity of antibiotics provides a constant challenge to bacterial populations: we are just now learning about the effectiveness of the bacterial response to this challenge. Largely through misapplication of technology, the effectiveness of antibiotics is diminishing dramatically (Anderson and May 1992; Anderson 1998; Anderson 1999; Austin *et al.* 1999).

Although our ability to control viral populations has been extraordinarily successful in certain domains (*i.e.*, in vaccination against polio, smallpox and an array of childhood diseases), other viruses, most notably HIV, show little sign of vulnerability to conventional anti-viral approaches (Lederberg 1998). Particularly

troubling is the rapidity with which viruses adapt to treatment conditions, developing resistance to anti-viral drugs.

The crisis faced by biomedicine and public health, then, is evolutionary. No antibiotic or anti-viral strategy is 100% effective against a given population. Because of the high mutation rate, high replication rate and large population sizes of microbial populations, the very strong selection of an anti-microbial treatment virtually ensures the evolution of resistance, whenever such resistance is possible. In many cases, an anti-microbial treatment that is less than 100% effective will shortly decline in effectiveness.

Designing New Drugs

In the past few decades some hope has come from the field of rational drug design. With the explosion of computing power and the development of new algorithms to predict molecular interactions came the expectation that new drugs could be developed *de novo* and as needed. Despite its original promises, the “rational design” of new antibiotics and antiviral drugs has turned out to be extremely difficult, and largely a process of trial and error. Furthermore, the potential for adaptation to drug resistance in microbes can quickly render these drugs obsolete.

The evolution of resistance to antisense molecules provides a case in point (Bull *et al.* 1998). Antisense RNA molecules designed to bind precisely to targets in the viral genome are rapidly rendered ineffective by evolution of the viral sequences. Antisense technology affords an easy design of new molecules against the resistant viruses, but the diversity of resistant viruses then precludes control

by any small set of inhibitors. The antisense strategy does not seem to be capable of suppressing the adaptation of viral populations to avoid inhibition.

Evolving Old Drugs

This study was designed to evaluate the feasibility of evolving drugs against an evolving pathogen. It was also hoped that, should the evolution of a drug against a resistant pathogen be possible, that this model would address important questions about the behavior of the evolutionary arms race of the drug-pathogen pair, such as possible treatment strategies that might arise from an evolving anti-pathogen drug.

The ability of a drug to evolve new effectiveness depends first on its having the analog to a genome. Additionally, of course, there must exist a mechanism for translating the genome into the molecule responsible for the anti-microbial action. The drug itself should be simple in structure, action and synthetic pathway. It might be possible to code in a genome the biosynthetic pathway of an antibiotic, but the ability of such a complex pathway to rapidly evolve new variants might be limited. Candidates for evolvable drugs will more likely be those that act on the level of nucleic acid binding (as exemplified by antisense molecules).

Several questions arise from consideration of this strategy:

1. Can nucleic-acid based inhibitors of a pathogen be developed that sufficiently block infection or growth of the pathogen?

2. Can these inhibitors be developed to allow sufficient range in action such that their effectiveness covers a wide range of pathogen response?
3. Does the “adaptability” of a pathogen to an inhibitor diminish over time, that is, can evolving inhibitors be used to direct pathogen evolution into a region of genotype space from which further adaptation is limited?

A Drug-Pathogen Arms Race: *E. coli* versus ϕ X174

The system chosen here used the bacteriophage ϕ X174 as a model virus. This virus has recently been used in several studies of adaptation (Bull *et al.* 1997; Wichman *et al.* 1999; Bull *et al.* 2000; Crill *et al.* 2000) and based on prior biochemical and structural work is very well studied (Pol and Arkel 1965; Hayashi and Hayashi 1968; Siegel *et al.* 1968; Linney *et al.* 1972; Jazwinski and Kornberg 1975; Jazwinski *et al.* 1975; Jazwinski *et al.* 1975; Jazwinski *et al.* 1975; Borrias *et al.* 1979; Hayashi *et al.* 1983; Hayashi and Hayashi 1985; Hayashi *et al.* 1989; Fane *et al.* 1993; Ekechukwu *et al.* 1995). Furthermore, its genome is easily sequenced to observe evolution.

DNA-based inhibitors of ϕ X174 have been discovered as well. Van der Avoort identified a small portion of the ϕ X174 genome that, when cloned in a plasmid, stopped infection by the virus (van der Avoort *et al.* 1982; Van der Avoort *et al.* 1983; van der Avoort *et al.* 1984). Likewise, Holzmayer *et al.* (1992) described a method to clone inhibitors of bacteriophage lambda using random libraries expressed by the host. Together, these two studies suggested that

DNA based inhibitors against ϕ X174 could be discovered and implemented. Thus a model system for an evolutionary arms race between drug and pathogen seemed feasible.

I approached the problem of designing an evolvable anti-microbial drug in the context of the bacteriophage ϕ X174. In this system the bacterial host, *E. coli* C, was transformed with a library of plasmids containing potential anti-bacteriophage sequences. Selection for bacterial growth in the presence of phage allowed me to recover sequences on the plasmid responsible for inhibition. To empirically model the evolution of pathogen resistance I selected phage that could grow on the inhibitory host, thereby “escaping” the action of the inhibitory plasmid. Table 6 shows the elements of the ϕ X174 model system and how they relate to those of a more general drug-pathogen arms race.

Table 6: Correspondence between elements of pathogen/drug evolution and a bacteriophage model system.

Human or animal host	↔	<i>E. coli</i> strain C
Pathogen	↔	Bacteriophage ϕ X174
Anti-pathogen drug	↔	Inhibitory plasmid
Resistance to drug	↔	Phage “escape” from inhibition
Evolution of drug	↔	Mutagenesis and selection of new inhibitory plasmid

METHODS

Strains and General Methods

Unless otherwise specified, the bacterial strain used throughout these experiments was the prototroph *Escherichia coli* C. The strain of ϕ X174

bacteriophage used was called ϕ XB, and was identical in sequence to the phage A used by Bull *et al.* (1997). S13 and G4 bacteriophage were the wild-type strains. Methods of host and virus propagation and storage are otherwise described in chapter 2.

PCR amplifications used Taq DNA polymerase and followed standard protocols. The annealing step took place at 50°C for 1 minute. The denaturation step was at 94°C, also for 1 minute, and the elongation step was at 72°C for 2 minutes. Before the amplification cycles were started, the reactions were heated to 94°C for 4 minutes. Amplification occurred over 35 cycles of denaturation-annealing-elongation, and was followed by an extra extension step at 72°C for 7 minutes.

Construction of Cloning Vector pUCUT

The expression vector pUC18 was modified in a series of PCR steps to create the cloning vector used in this study (pUCUT):

1. The region containing the α fragment of β -galactosidase was removed;
2. A strong transcriptional terminator was added downstream of the insertion site;
3. Three stop codons were added, one in each reading frame, downstream of the insertion site;
4. A new polylinker was inserted.

pUCUT thus retains the inducible promoter of pUC18 but lacks the α fragment of *LacZ* and so should not produce a fusion protein. Figure 13 diagrams

these steps, and gives a map of the final plasmid. The resulting cloning vector had the following properties:

1. high copy number;
2. ampicillin and tetracycline resistance;
3. a strong IPTG-regulated promoter upstream of the cloning site;
4. a strong terminator and stop codons in all three reading frames downstream of the cloning site;
5. a polylinker with EcoRI, BamHI, HindIII and EcoRV cloning sites;
6. PCR/sequencing primer binding sites flanking the polylinker.

For cloning, plasmid pUCUT was digested with the EcoRV restriction enzyme, and treated with calf intestinal alkaline phosphatase to remove terminal phosphates and decrease the incidence of intramolecular ligation. These treatments followed standard protocols, as described in Sambrook *et al.* (1989).

Cloning of ϕ X174 Sequences into pUCUT

A PCR product (extending from base 2605 to 50) from the circular ϕ X174 genome and digested with DNaseI provided a pool of genome fragments for cloning. To approximately 5 μ g DNA was mixed 10 μ L digestion buffer (500 mM Tris·HCl, pH 7.6; 100 mM CaCl₂; 1 mg/ml BSA, fraction V) and water to 100 μ L final volume. 10 μ L of this solution was transferred to a tube containing 5 μ L 50 mM EDTA, pH 8, on ice. This tube became the “0 minute” time point.

The remainder of the reaction was combined with 1.5 μ L DNase I (1 mg/ml in 0.01 N HCl), and incubated at 15°C. At predetermined time points (1, 2, 5, 10

and 30 minutes) 15 μL of the reaction was removed to a labeled tube containing 5 μL EDTA, pH8, and kept on ice. After the 30-minute time point was collected, 2 μL of each fraction was visualized on an agarose/TEA gel (0.8%) to assess the digest. Fractions that were sufficiently digested (the main band running between 100 and 500 nt) were combined, brought to 360 μL with water, extracted with phenol and phenol/chloroform and precipitated using sodium acetate (0.3 M) and cold ethanol (2.5 volumes).

The pellet containing the DNA from the combined digested fractions was dissolved in 25 μL and treated to fill in any overhangs present from the digestion, as follows. To the pellet was added 3 μL of a mixture containing 0.5 mM of each of the four deoxyribonucleotides, 3 μL 50 mM MgCl_2 and 2 μL of a 5 unit/ μL solution of the klenow fragment of DNA polymerase I. The reaction was incubated for 15 minutes at room temperature, then extracted and precipitated as described above. The final pellet was dissolved in 20 μL of water. Figure 14 illustrates the steps used in cloning.

Screening of Library for Inhibitory Activity

The library was selected for inhibition of ϕX174 . This procedure consisted of the following steps:

1. Transformation into susceptible host (*E. coli* C);
2. Plating of transformant pool onto plate containing ϕX174 and growth at 37°C overnight;
3. Scraping of resulting resistant colonies;
4. Isolation of plasmid DNA and transformation into new cells.

Electroporation was followed by 1 hr growth at 37°C in 1 ml LB broth. The growing culture of transformants was pelleted and resuspended approximately every 20 min. in LB with 2mM EDTA to reduce the incidence of infection by phage DNA that might have been carried along during the DNA prep procedure. Possible infection at this stage was a potential problem since it would have increased the selection for cellular-based resistance to phage. Step 4 was necessary to eliminate cellular-based resistance to the bacteriophage at each round of selection.

The screening process was repeated several times and at different concentrations of phage and cells. The initial steps in the selection and screening were done with the addition of 0.1 mM IPTG to the antibiotic-containing plates, in order to de-repress the Lac operator and increase expression of the insert. IPTG induction was later determined to be unnecessary for the inhibitor isolated.

Plasmids conferring phage resistance were obtained, and the inserts were sequenced. The nomenclature for resistance-conferring plasmids was a combination of the origin of the library and the number of the isolate. The *i*th isolate from a library derived from ϕ X174 was called pFi, while the *i*th isolate from S13 was called pSi.

Sequencing of Plasmid and Bacteriophage DNA

PCR primers were designed to flank the cloning site in the vector pUCUT, described below. A standard PCR cycle with a 50°C annealing temperature was used to amplify the insert. Sequencing of the insert was done from this PCR-amplified template, using either of the two PCR primers.

The phage storage solution (BE) contains EDTA and inhibits Taq polymerase activity. Phage whose genomes were to be amplified were first plated onto *E. coli* C, and 20 plaques were combined into 100 μ L LB broth. After vortexing, this suspension sat at 4°C for at least six hours, and was centrifuged to remove debris. The supernatant was kept as a “plaque prep” and used as a source of template phage DNA for PCR amplification.

Double-stranded ϕ X174 DNA was prepared by PCR amplification of phage genomic DNA in two overlapping segments, defined by the genome nucleotide positions 0-2953 and 2605-50. The PCR cycles for ϕ X incorporated a 7 minute extension step to account for the length of the product. Sequencing of the ϕ X genome was done using two sets of 12 primers each, spaced at approximately even intervals and with binding sites on opposite strands of the genome. In general, only one of these sets was used, unless the reaction resulted in sequence ambiguities, in which case the opposite strand was sequenced as well. Using the sequence analysis package DNASTar™, phage sequence data were aligned to the ancestral (ϕ XB) sequence for comparison, and changes relative to that sequence were noted.

Fitness Assay

Three methods were used to assess levels of bacterial resistance to ϕ X174. All three methods were conducted on plates and thus differ from assays used in chapter 2. These assays were (1) a quantitative estimate of phage growth rate on plates, (2) ability of a culture streak to grow on plates across a line of phage, and (3) plaque size.

In the plate fitness assay, approximately 100 plaque-forming units (pfu) were plated on log-phase host cells. This plate was grown at 37°C for exactly three hours, while another control plate, containing the same quantity of phage on a permissive host, was incubated until plaques were clearly visible.

After three hours of growth, the assay plate was soaked in approximately 5 ml borate-EDTA, and the top agar was scraped from the plate. This suspension was vortexed and allowed to stand at room temperature for 1 hour, then centrifuged to remove cells, agar and debris. The resulting solution was titered to determine the phage concentration and total phage numbers recovered from the plate.

For the three hours growth on plates, the total number of phage doublings was calculated as the base 2 logarithm of the number of phage obtained from the assay plate divided by the number of plaques on the control plate. The value for one hour of growth was one-third of the three-hour value. This method assumes that all of the phage present on the plate were extracted into solution, so it will underestimate fitness to the extent that phage are trapped in the top agar. However, any proportional loss of phage in the top agar will not affect between-phage comparisons.

This procedure gives the doubling time on plates for a bacteriophage strain grown on a particular bacterial host strain. The hosts are clearly not identical from trial to trial, which may in turn introduce variation into the fitness estimates. To control for these differences, reference strains of bacteriophage were assayed along with the strain of interest.

I also used a more rapid and qualitative assay of inhibition. This simple assay merely determined if a streak of the host cells would grow across a line of phage, on a plate of LB/agar. For this assay, a streak of a suspension of phage in LB was dried on a plate and then cross-streaked with a suspension of cells. The plate was incubated at 37°C for 8-12 hours, until cell growth in the streak was visible. In this manner, up to ten host suspensions could be assayed per plate. Three categories of susceptibility were noted. Streaks of the host cell suspension either grew unhindered across the phage streak, showed a gap in growth at the phage streak, or showed no growth from the phage streak onward. Continual growth across the cell streak indicates complete resistance, while gapped or shortened streaks show different levels of susceptibility to the phage. Figure 16 shows a schematic diagram of these outcomes.

Finally, plaque size was taken as a general qualitative index of fitness. The very low-fitness isolates had plaques that were nearly invisible, as compared to the high-fitness wild-type phage, which had plaques several millimeters in diameter.

Selection of Escape Mutants

The second step in the evolutionary arms race required a bacteriophage variant capable of growth on cells containing an inhibitory plasmid. This selection was done as a simple screen for plaques on lawns of inhibitory cells. Putative resistant phage were replated on inhibitory cells, both to purify the clone and to select further increases in resistance (as evident from plaque size).

The first resistant phage, ϕ I, was isolated by me; another 6 were found by Mara Lawniczak.

Construction of Randomly Mutagenized Library

A major goal of this study was to evolve new inhibitors, just as the virus evolved to avoid inhibition. One potential problem in evolving new inhibitory plasmids from the old one derives from the relatively low mutation rate of the *E. coli* replication machinery (Drake 1991). The variation present in a pool of 10^9 - 10^{10} inhibitory plasmids will simply be too small to adequately explore sequence space for new inhibitors, unless only a single base substitution is required. Most methods of *in vitro* random mutagenesis also result in a small proportion of mutated residues. I therefore constructed a library of synthesized inserts based on the pF1 sequence, in which each nucleotide of the insert remained unchanged with a probability of 97%, and had a 1% chance of mutating to each of the other nucleotide residues. The insert for this library was synthesized as two 100 nucleotide (nt) oligonucleotide segments. Given this pool, with 200nt each randomized with a probability of 3%, the frequency of the original wild-type sequence should equal about 0.0023. The average number of mutations among randomly selected clones should be about 6.

The strategy for synthesizing the randomized inhibitor library is diagrammed in figure 15, and primer sequences are given in table 7. Two PCR primers (v1 and v2) were used to amplify the vector sequences, along with a small region of the inhibitory sequence. The randomized oligonucleotides rf1 and rf2 were then used as PCR primers to fill in the remainder of the randomized pF1

sequence. Products from this step were purified (although not size-selected), blunt-end ligated and transformed into *E. coli* C. The transformation resulted in approximately 10^4 colonies.

Table 7: Primers used for construction of randomly mutagenized library.
Lowercase nucleotides are degenerate positions, as described in the text.

v1 5'-GAG AAG AGC CAT ACC GCT GAT TCT G -3'

v2 5'-AAA TCA AGC TTG GAT CCG AAT TCC A -3'

rf1 5'- tcc aag tat cgg caa cag ctt tat caa tac cat gaa aaa tat caa
cca cac cag aag cag cat cag tga cga cat tag aaa tat cct ttg
cag tag cgc caa tat GAG AAG AGC CAT ACC GCT GAT TCT G -3'

rf2 5'- aca att tct gga aag acg gta aag ctg atg gta ttg gct cta att
tgt cta gga aat aac cgt cag gat tga cac cct ccc aat tgt atg
ttt tca tgc ctc caa AAA TCA AGC TTG GAT CCG AAT TCC A -3'

After transforming the library into *E. coli*, I picked three clones at random, grew them in a 2 ml culture and extracted the plasmid DNA for sequencing. The remainder of the transformants were then scraped, titrated to determine the density of bacteria, and plated at a density of 2.5×10^7 , 2.5×10^9 and 2.5×10^{11} colony-forming units (cfu) on LB/Ampicillin plates with approximately 5×10^5 plaque-forming units (pfu) of either ϕX or $\phi 1$ on the surface.

As with the original isolation of the resistant sequences, the scraping/DNA-isolation/transformation procedure was repeated twice more, to result in a pool of clones highly enriched for resistance to the phage on which they were initially plated. These clones were colony-purified and resistance was assayed by plating the appropriate phage onto a log-phase culture of the

transformant and scoring for the presence or absence of plaques, or by streaking the cultures across a line of the appropriate phage.

Challenge by Related Viral Inhibitory Sequences

One rationale for using ϕ X174 as a model for evolving new drugs was that the existence of several well-characterized bacteriophage of various degrees of relatedness should allow the easy isolation of homologous sequences that may be inhibitory to new escape mutants. In essence, these related phages have possibly already evolved new inhibitory sequences, obviating the need to evolve them from the ϕ X library sequences.

The related bacteriophages I examined were S13 and G4. These differ in varying degrees from each other and from ϕ X174: the differences (expressed as pairwise percent nucleotide difference) are shown in the lower diagonal of table 8. Of particular interest is the percent pairwise difference between these phages in the region that inhibits phage growth, that is, the region spanning the 3' end of gene H and the H/A intergenic region. These differences are shown in the upper diagonal of table 8.

Table 8: Pair wise sequence differences between entire genomes (below diagonal) and between the autoinhibitory region (above diagonal) for the related bacteriophages ϕ X174, S13 and G4.

	ϕ X174	S13	G4
ϕ X174	-	3.4%	46.4%
S13	2.1%	-	49.8%
G4	33.9%	33.6%	-

I attempted to clone the regions in S13 and G4 that are homologous to the H/A intergenic region of ϕ X using the same DNase I strategy used for cloning the ϕ X inhibitory sequences. These clones were screened for resistance to the parent phage, to ϕ X174 and to the pF1 escape mutants of ϕ X174. They were also used to select escape mutants from the stocks of the parent phage.

RESULTS

Selection and Analysis of Resistance-Confering Sequences

I obtained four apparent resistance-confering clones from the initial ϕ X library. These plasmids were named pF1, pF2, pF6 and pF7. pF1 and pF2 inserts were identical (the multiple rounds of selection allows each original clone to become represented by multiple copies). Not all overlapped the H-A intergenic region; the genomic boundaries of the inserts and their orientation relative to the direction of plasmid LacZ transcription are listed in table 9.

Table 9: Boundaries and orientation of resistance-confering inserts. pCB102A is identical in sequence to the plasmid isolated by Van der Avoort.

Plasmid	Nucleotide boundaries		genes	Orientation
	start	end		
pF1, pF2	3701	3964	H,H-A	Antisense
pCB102A	3708	4201	H,H-A,A	Sense
pS4	5056	58	A, A*, B	Antisense
pS5	5176	39	A, A*, B	Antisense

Cells with plasmid pF1 were always resistant, and the insert of pF1 corresponded closely to the “incompatibility” or “reduction” region described by

Van der Avoort *et al.* Inhibition by pF6 and pF7 was not repeatable, so only pF1 was chosen for closer study. The location and orientation of the insert in plasmid pF1 is shown in figure 17. The insert from Van der Avoort's plasmid has almost exactly the same boundaries, but with the opposite orientation relative to van der Avoort *et al.* (1982).

Selection of Viral Escape Mutants

Evolution of viral resistance was sudden and complete: when plated at high titer on bacteria carrying the resistance plasmid pF1, ϕ X174 cultures gave rise to mutants that could plaque. After four selection passages on pF1-transformed *E. coli*, the population of phage produced numerous large plaques. A single isolate (ϕ 1) was plaque-purified from this culture, and became the basis of further experiments.

The ϕ 1 genome contained two substitutions: 3387 (G→A) and 3786 (G→T). Both occur within the coding region of gene H and result in a glutamate→lysine change at residue 153 and a valine→phenylalanine at residue 286, respectively. The changes do not lie within the segment of the inhibitory region.

Fitness Effect of Escape from Inhibition

Plate fitness assays were conducted for ϕ X and ϕ 1 on hosts with the wild-type plasmid pUCUT and on hosts with pF1. ϕ XB had a fitness of 8.0 on cells bearing the vector pUCUT but only 3.8 when grown on cells bearing pF1. In contrast, ϕ 1 had a fitness that was not substantially different between pUCUT and

pF1 (7.3 doublings/hr, sd=0.7 vs. 7.9 doublings/hr, sd=0.5). The recovery by this resistant virus is essentially complete (a 17-fold higher yield per plaque per hour).

Challenge by Related Viral Inhibitory Sequences

Two ϕ X174-selected clones (pS4 and pS5) were obtained from the S13 genomic library, and viral resistance was confirmed for them by a plate streak test. These plasmids also conferred resistance to S13, but not to the pF1-resistant ϕ 1 (as determined by plate streak tests). In contrast to the findings of Van der Avoort (1984), no ϕ X or ϕ 1 resistance-conferring clones were obtained from the G4-derived library, despite several attempts using different methods of cloning.

Selection of Resistant Sequences from the Randomly Mutagenized Library

The three random clones picked from the mutagenized inhibitory library were sequenced and assayed for resistance to both ϕ X and ϕ 1. None were completely resistant to either ϕ X or ϕ 1. However, the ϕ X plaques on two clones (isolates 2 and 3) were smaller than those on isolate 1, as well as being smaller than the plaques on the control plating of *E. coli* C containing the vector alone. Plaques of ϕ 1 were normal in size and number on all three clones, indicating no inhibition to this phage.

Twelve clones were randomly chosen from the pool of ϕ X-selected transformants, and all were confirmed as being resistant to ϕ X. The differences between these sequences and the starting sequence are shown in table 10. Ten clones were similarly chosen from a pool of ϕ 1-selected transformants, but none were confirmed as being resistant to ϕ 1. These isolates were not sequenced.

Table 10: Mutations in ϕ X resistant clones from randomly mutagenized library.

mutation	Clone														
	random			ϕ X174 resistant											
	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12
A3754								T			T				
G3759	A														
G3760														A	
G3762						C									
C3763											T				
A3765	C														
T3767	A														
G3782	T														
G3789								A							
G3798								A							
G3807	A														
T3827															C
A3835	T														
G3876	A														
G3888										A					
G3925	T														
T3927								A							
G3929	T														
C3937								T							
C3938								G							
C3939	G														
A3941												T			
T3945							C								
G3948			T												
T3949	A													C	
T3952			G												
A3954								T							
T3955								A							
G3956	C														
C3957										T					
C3958	T		G												
T3959								G							
C3961			T				T	T							
A3962			T												
A3963	T		T	T	T	T	T		T					T	
#mutations	14	0	5	1	1	1	3	11	2	1	2	1	0	3	1
Mean, all mutations = 3.58 (3.62)															
Mean, excluding 3963 = 1.83 (2.98)															

DISCUSSION

Evolutionary Significance of Autoinhibitory Sequences

The most successful autoinhibitory sequence from the ϕ X library (pF1) corresponded closely to that described by Van der Avoort *et al.* It completely suppressed plaque formation by wild-type phage (ϕ XB), though fitness estimates showed some phage replication. The 263-base inhibitory insert spans 27 residues of the C-terminal portion of gene H (which codes for the so-called “pilot protein” described by Jazwinski *et al.* (1975)), and nearly all of the intergenic region between genes H and A. The H/A intercistronic region is known to contain the promoter for gene A, the terminator for gene H and a gene H transcript-stabilizing element (Hayashi *et al.* 1989). It is also suspected to be the site where the genomic DNA binds to localized cellular elements during replication (van der Avoort *et al.* 1982).

The inhibitory sequences of Van der Avoort *et al.* were cloned in the sense orientation relative to the vector promoter, whereas the insert on pF1 was present in the anti-sense orientation. In pF1 inhibition of phage infection was not dependent on transcriptional activation by IPTG. Together, these data suggest that transcription of the insert is not important. Rather, the mere presence of the sequences on the plasmid is sufficient to inhibit infection. This interpretation is consistent with the mechanism proposed by Van der Avoort, in which genome-binding factors necessary for replication are sequestered or improperly localized by the plasmid DNA itself, rather than by any product of these sequences.

There are many other examples of inhibition caused by cloned viral genome sequences, including those of bacteriophage lambda (Holzmayer *et al.* 1992), HIV (Dunn *et al.* 1999) and carcinogenic retroviruses (Delaporte *et al.* 1999). The existence of a viral genomic sequence that reduces infectivity is suggestive of a mechanism to block superinfection. The product of gene H is known to bind phage genomic DNA, both at multiple sites in the capsid, and at unknown sites during replication. The protein is minimally necessary for the first two stages of replication (Jazwinski *et al.* 1975), and may be a means of localizing the phage genome to a specific intracellular domain. If the gene H product binds to the segment of the genome represented by the inhibition sequence, an excess of the sequences (either on a replicating phage genome or borne on a plasmid) would likely titrate the entering gene H protein from super-infecting genomes. (However, ϕ X DNA can be successfully transfected, so an accompanying H protein is not essential under those conditions.)

Autoinhibition and Evolutionary Versatility in ϕ X174

Notwithstanding the original motivation of the study (host-pathogen arms-races), the evolution of escape from autoinhibition can reveal something about the evolutionary versatility of the bacteriophage in a new selective environment.

ϕ X174 has been used as a model organism in many studies on adaptation to novel environments. Bull *et al.* (1997) examined the molecular changes in several independent lineages resulting from adaptation to growth at elevated temperatures and on novel hosts. They found a large degree of convergence across lineages, both in response to host-specific and temperature-specific

changes. In a chemostat study with ϕX , Wichman *et al.* (1999) found parallel substitutions between two lineages adapting to high temperature growth conditions. These parallel changes did not occur in the same order in each lineage, however, and further they did not represent the largest-effect substitutions across the course of the adaptation. The serial host-switching experiments of Crill *et al.* (2000) found a small number of sites that switched in accordance with the current host. These sites, responsible for host specificity, exhibited an asymmetrical fitness cost that did not seem to be able to be ameliorated by secondary compensating substitutions.

These studies suggest a fairly complex picture of ϕX adaptation to new environments. The high-temperature growth and host-switching experiments indicate a limited range of adaptive options available to the phage. In the strong selective environment created by autoinhibitory sequence expression, it might also be expected that a few changes of high fitness effect, if available, would be taken by the phage.

The Evolution of Viral Escape from Inhibition

Here, too, it was found that the changes found in escape mutants, all in gene H, have very high fitness effects (table 11). There are several mutations of roughly equal effect, within one general region of a DNA-binding protein. As with antisense RNA, inhibitory activity depends upon the binding of the inhibitor to a target sequence. This binding of inhibitor to a specific genomic sequence might be easily disrupted by changes in DNA-binding domains of the gene H product.

If a fitness increase depends upon the disruption of a complex interaction, the resulting fitness surface should have many peaks – there are many possible ways to effect the disruption. The number of available adaptive peaks may be limited, of course, by the constraint that the gene H protein work normally in genome entry and replication. Nevertheless, for this particular case, ϕX has many optimal solutions to the problem of inhibitory sequences.

In several of the escape mutants the substitution sites were at or near known differences between ϕX and S13 or G4. (For comparison, these substitutions are shown in table 11). This would be expected if what was required for escape was simply an alternate form of the gene H product that binds with less affinity to the ϕX -specific inhibitory sequences, but otherwise functions normally.

As shown by Van der Avoort *et al.* (1984), the ϕX inhibitory sequences do not inhibit G4. The sequences of the cloned ϕX escape variants may indicate why: G4 has amino acid substitutions at four of the sites present in those variants. G4 is essentially a ‘pre-escaped’ phage (if one can interpret each change separately).

Table 11: Nucleotide and amino acid changes in wild type and escape phages.

Nucleotide:	3340	3387	3474	3753	3755	3781	3786	3787	3788
Amino acid:	137	153	182	275		284	286		
ϕ X174	A	G	G	C	T	C	G	T	C
S13	G D→G								
G4	G D→G		A A→T	G H→Q			C V→A		
ϕ 1		A E→K					T	V→F	
ML clone 1			C A→P						
2				G H→D					
3						T S→F			
4						A S→Y			
5							A		
6							A	V→I	
CB clone 1	G D→G							V→I	

Cost of Escape

It is reasonable to assume that the changes in gene H that confer the ability to escape might come at some cost to the fitness of the escape mutant in a non-inhibiting host. There may be a small cost to escape from the pF1 autoinhibitory plasmid (7.3 doublings/hr, sd = 0.7) as compared to that on the vector control (7.9 doublings/hr, sd = 0.5), though the errors in the estimate do not allow a rejection of the hypothesis of equal fitnesses. Furthermore, the effects of the inhibitory insert on cell growth may be subtle, but might affect the apparent fitness of an infecting phage.

Selection of Sequences from Random Library

The sequence differences seen in the three unselected clones from the randomized library were somewhat consistent with expectations. Assuming a binomial distribution for mutated sites, the expected number of changes across the insert sequence was 6, with an expected variance of 5.82. In this small sample of three clones, the average number of changes was 6.33. One clone had zero changes, an event that would be expected in about 1% of samples of this size.

The fact that the wild-type clone appeared at all may be due either to a smaller than expected rate of mutagenic replacement in the oligonucleotide pool, or contamination in the cloning of full plasmids from the original PCR reaction. However, it is important to realize that these “random” clones are not actually completely random. Although they have not gone through the selection process for resistance to bacteriophage, the growth of the plasmids carrying these sequences is itself a selection. It is quite possible that a large number of mutant sequences are toxic, or inhibit cell growth. If this were the case, an over-representation of the wild-type sequence would occur in the library after cell growth.

Seven of the fifteen sequenced clones from the randomly mutagenized pool contained A to T substitutions at the position corresponding to nucleotide 3963. This prevalence is almost certainly an artifact of the library synthesis procedure. Site 3963 is the first degenerate site in the mutagenic primer rf2, and therefore the first site after the vector amplification primer v2. The PCR reaction with Taq polymerase is known to occasionally add one or more adenosine

residues to the 3' termini of the products. This reaction could explain both the multiple A→T substitutions at sites 3961, 3962 and 3963, and the T→A substitutions at 3754.

The failure to obtain second-generation inhibitors from the randomly mutagenized library might be due simply to technical factors. The complexity of the library was on the order of 10,000 clones. Although the theoretical number of sequences is 4^{200} , the limit to the library's complexity comes from the size of the PCR reaction synthesizing the library. The amount of degenerate primer in the reaction was 25 nmol, or 1.5×10^{16} molecules. This is the largest number of sequences that could possibly be examined, assuming 100% yield at every step of synthesis, cloning and bacterial transformation. In practice, of course, the actual number is many orders of magnitude smaller: the mass of 1.5×10^{16} cloned sequences would be more than 30 mg.

The transformation step in making the library is particularly vulnerable to loss of product. The library actually screened was transformed directly into *E. coli* C by electroporation. Commercially available ultra-high competent cells were also used, but yielded no transformants, either with chemically treated or electrocompetent cells. It remains to be seen whether this failure was due to some fault of the procedure or an incompatibility between the strain and the sequences being cloned.

Of course it is also possible that the library even at its most complex would not contain sequences capable of overcoming the phage escape mutants. If the escape mechanism works by way of disruption of binding between the

inhibitory sequence and the gene H product, recovery of that binding by mutations in the inhibitory sequence would be very difficult. In contrast to recovery from antisense inhibition in which the sequence need change only enough to restore Watson-Crick base pairing with the inhibitor sequence (Bull *et al.* 1998), the protein-DNA interactions in this system are much more complex.

Inhibitory Sequences from Related Phages

Inhibitory sequences from S13 worked against the parent phage and against ϕX , but not against the escape mutant $\phi 1$. S13 and ϕX are sufficiently similar that their mechanisms to block superinfection may be shared. The fact that one resistant ϕX isolate (CB clone 1) had a change that matched a difference between ϕX and S13 may indicate that there are interactions among the sites within protein H necessary for inhibition. If a D \rightarrow G change at residue 137 in gpH was sufficient to escape inhibition, S13 would also escape pF1, yet it was inhibited.

The lack of inhibitors from the G4 library was puzzling. It had been shown that the cloned G4 homolog of the H-A intergenic region confers resistance to ϕX , as well as to G4 (although the converse is not true) (van der Avoort *et al.* 1982; van der Avoort *et al.* 1984). Random isolates from the G4 library did not contain any of the H-A intergenic sequences. The failure of all attempts to clone G4 inhibitory sequences might be attributed to incompatibilities among the cell type and high copy-number plasmid used, and the inhibitory sequence. The original inhibitory sequences were cloned a low copy-number plasmid that had no transcriptional promoter. Further work will be required to resolve this problem.

Regimes of Less Stringent Selection – Evolution of Intermediate Inhibitors

The method used to clone inhibitory sequences resulted in extremely stringent selection. Because of this, the selection was most likely biased towards mechanisms of inhibition that act absolutely, that is, that block either infection or replication of the phage. There are other possible mechanisms that would have been completely missed by this approach. For example, it is possible that some sequences might confer an abortive "apoptotic" or "altruistic suicide" effect in which the infected cell dies shortly after infection. If the selection were to take place under conditions of low phage density or slowed phage growth, a micro colony of such altruistic cells would persist, as the infected cells would remove any phage encountered from the environment. Although the plate-selection regime used here should in theory have recovered altruistic inhibitors, it is possible (likely?) that phage growth during plate selection could not be suppressed enough to achieve that goal.

Figure 13: Plasmid pUCUT and its construction.

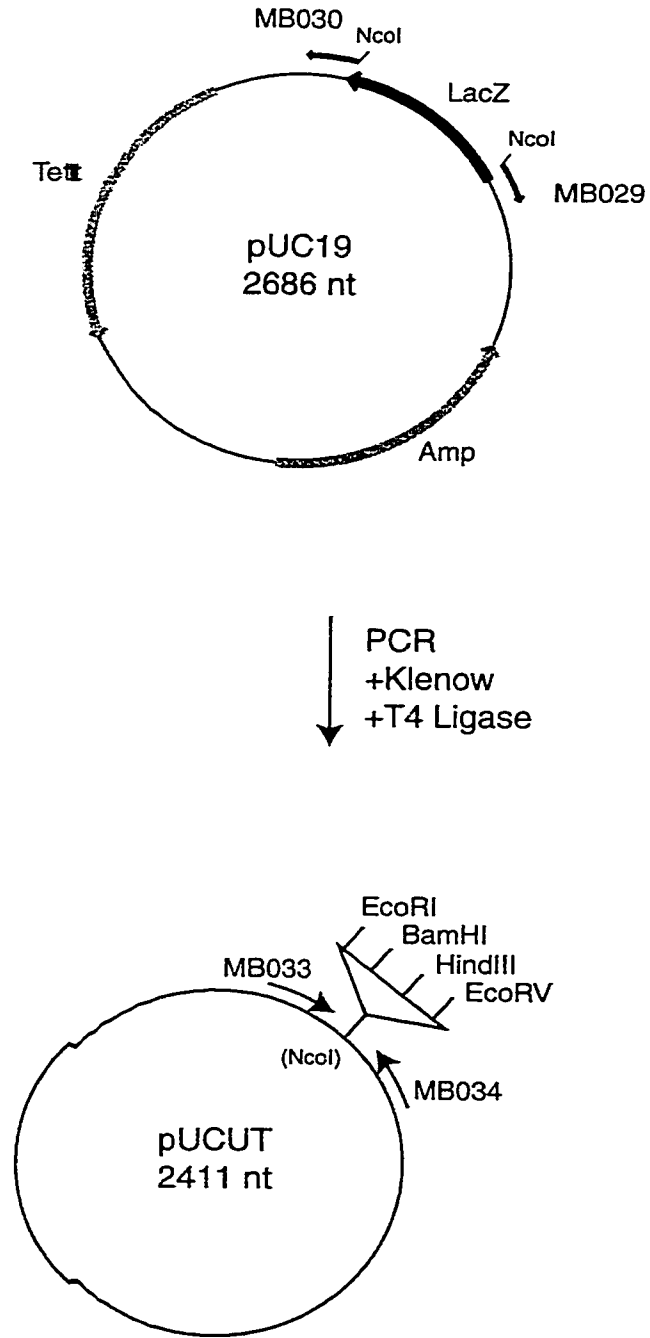


Figure 14: Cloning of autoinhibitory sequences into pUCUT.

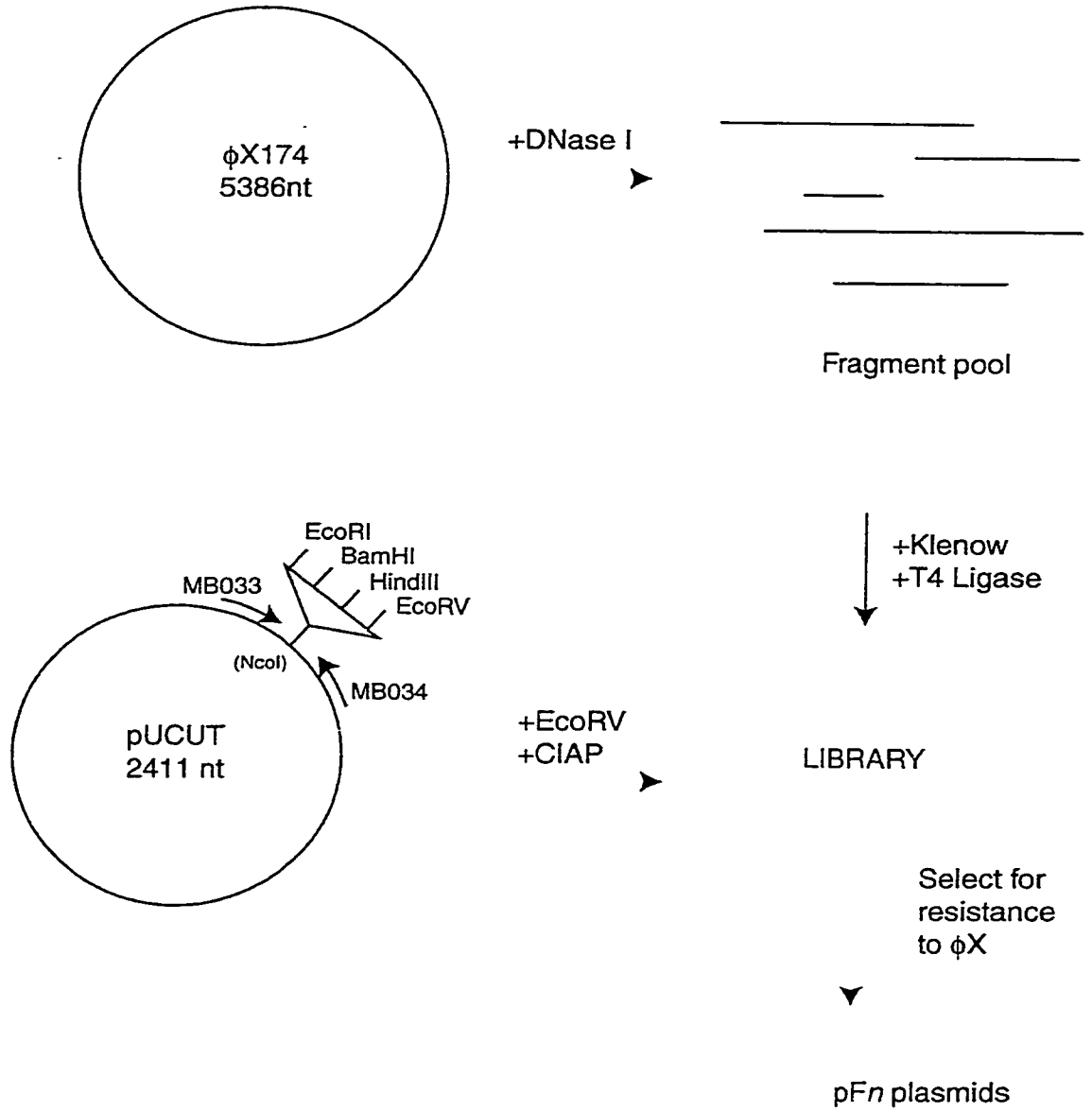


Figure 15: Strategy for synthesizing “shallow random” inhibitor library.

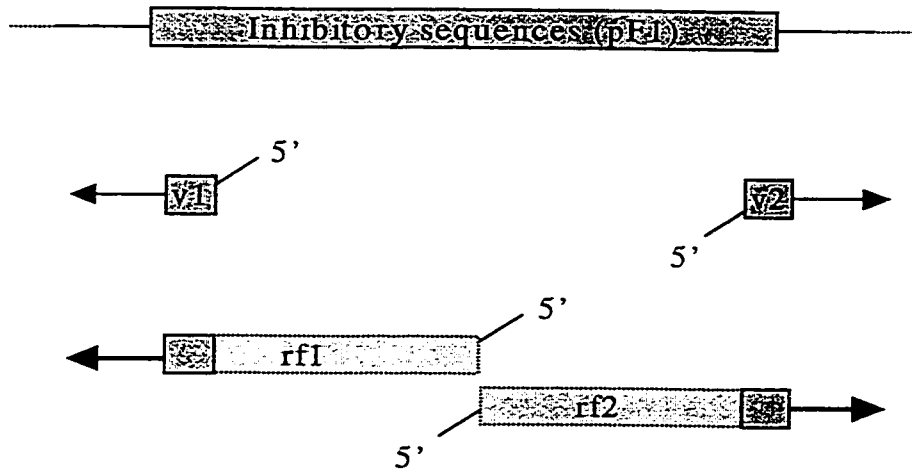


Figure 16: Streak test for bacterial resistance. Strains B, D and E are susceptible, strain A is resistant, and strain C is partly so.

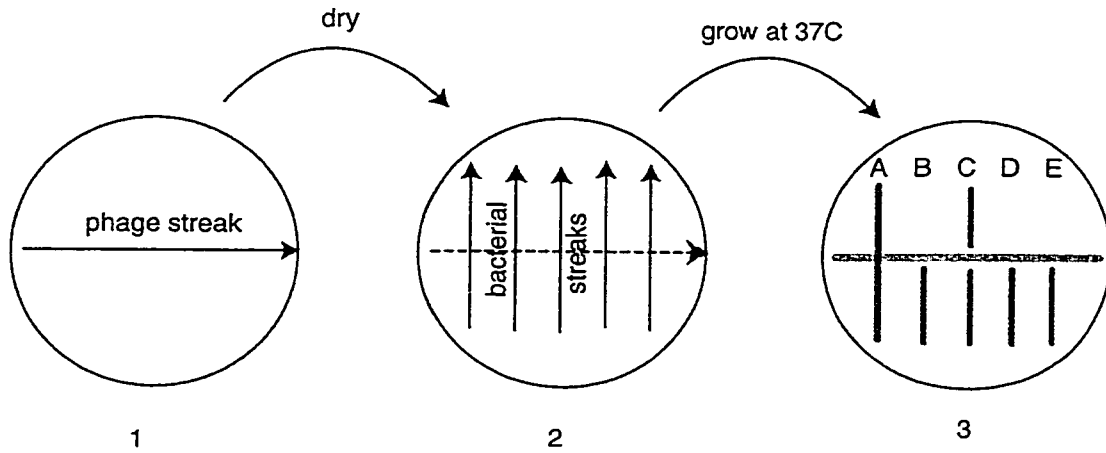
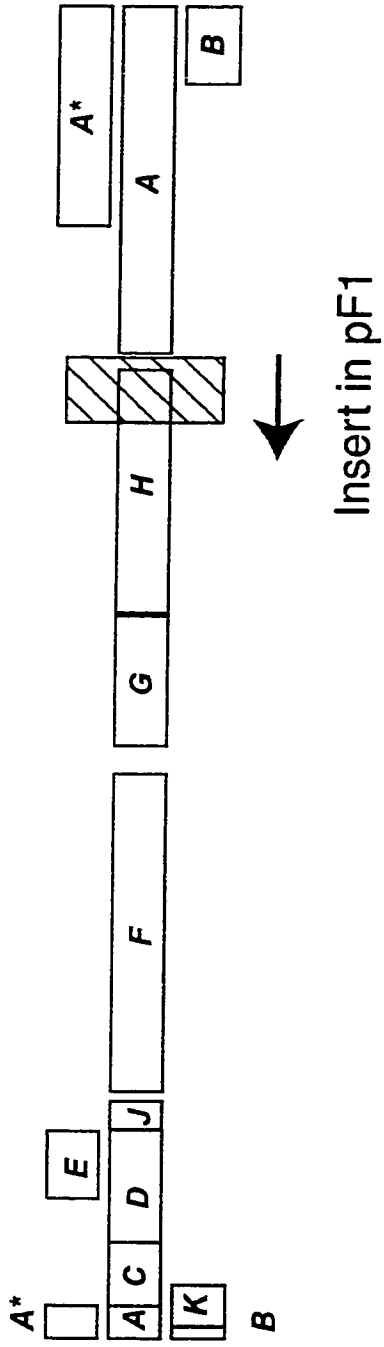


Figure 17: Location and orientation of pF1 inhibitory sequence in phage genome.

5386

0



Appendix A: Source code of program ‘asym’

“asym” simulates adaptive walks under the model given by Orr (1998), and allows modifications to some of the assumptions of the model, as described in chapter 1.

The program is written in C++ and makes use of two external libraries. Linear algebra functions are handled by the FORTRAN Basic Linear Algebra Subroutine (BLAS) library. Random number generation, random variables and sample statistics are handled Gnu Scientific Library (GSL). Both are freely available on the internet.

MAKEFILE

```
SHELL = /bin/sh

LIBHOME = /usr/local/lib

INCFLAGS = -I/usr/include/g++-2 -I/usr/local/include -I/usr/local/include/gsl
LIBFLAGS = -L./ -L/usr/local/lib/gsl
RANLIBS = -lgslrandist -lgslrng -lgslstatistics -lutils -lgslerr -lgslspecfunc
BLASLIBS = -llapack -lblas -latlas
STDLIBS = -lstdc++ -lm

$(OBJS) : $(SRC)
    $(CC) $(INCFLAGS) $(FLAGS) -c $(SRC)
SRCDIR   = ~/programs/asym
PROG     = asym
CC       = g++
LIBS     = $(RANLIBS) $(BLASLIBS) $(STDLIBS)
LINKER   = g77
#FLAGS   = -g -O
FLAGS    = -g
OBJS     = asym.o
SRC      = asym.cc

$(PROG) : $(OBJS) $(UTIL)
    $(LINKER) $(FLAGS) -o $(PROG) $(OBJS) $(LIBFLAGS) $(LIBS)

asym.o : asym.cc
    $(CC) -c asym.cc $(INCFLAGS) $(FLAGS)
```

```

“ASYM.H”
#include <string.h>
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <list>
#include <iomanip.h>
#include <fstream.h>
#include <float.h>
extern "C" {
#include <gsl/gsl_randist.h>
#include <gsl/gsl_rng.h>
#include <gsl/gsl_test.h>
#include <gsl/gsl_statistics.h>
}

#define true 1
#define false 0
#define boolean int

/* FORTRAN declarations (for BLAS) */
extern "C" {
void dcopy_(int*, double*, int*, double*, int*);
void daxpy_(int*, double*, double*, int*, double*, int*);
void dtrmv_(char*, char*, char*, int*, double*, int*, double*, int*);
void dgemv_(char*, int*, int*, double*, double*, int*, double*, int*, double*,
double*, int*);
double ddot_(int*, double*, int*, double*, int*);
void dgetrf_(int*, int*, double*, int*, int*, int*);
void dgetri_(int*, double*, int*, int*, double*, int*, int*);
}

void error(const char*);
int factorial(int);
int count(int, int);

template<class T>
class stack {
T* v;
T* p;
int sz;
public:
stack(int s) { v = p = new T[sz=s]; }
~stack() { delete[] v; }
void push(T a) { if (p - v == sz) error("stack overflow"); else *p++ = a; }
T pop() { if (p - v == 0) { error("stack underflow"); return 0.0; } else return
*--p; }
T pull() { if (p - v == 0) { error("stack underflow"); return 0.0; } else return
*++v; }
int size() const { return p-v; }
};

const double PI = 3.141592653589793238462643;
#define nMax 50
#define NMax 10000

```

```

“ASYM.CC”
#include "asym.h"

```

```

void main(int argc, char *argv[]) {

    gsl_rng_env_setup();
    gsl_rng *rng = gsl_rng_alloc(gsl_rng_ranlux);

    int n = 50, p = 0, N = 100, order = 0, sample_size = 1, count,
        mutations = 0, fixations = 0, trial = 0, i, j, k, l, sampleSize, nfix[NMax],
nben[NMax], nmut[NMax];
    unsigned long int seed;
    double trait0, epistasis, scaleProduct, effectsMean, effectsShape,
        rmOut[NMax], rnOut[NMax], roOut[NMax], rsOut[NMax], sOut[NMax], wOut[NMax],
        s, rStart, rPlus, r, abs_r, wStart, wPlus, w, w1, w2, wPot, traitStart,
traitPlus, trait,
        z0[nMax], z1[nMax], z2[nMax], disp[nMax], disp2[nMax], scale[nMax];
    char file1[20], file2[20], file3[20], file4[20], hist[20], posf1[20], symflag;
    ofstream outf, posf;

    if(argc == 1) {
        cout << "asym <effects mean> <effects shape> <asymmetry param> <seed> <n> <p>"
            << " <N> <K> <sample size> <trait> <filename>" << endl;
        exit(0);
    }

    trait0 = atof(argv[10]);
    sampleSize = atoi(argv[9]);
    order = atoi(argv[8]);
    N = atoi(argv[7]);
    p = atoi(argv[6]);
    n = atoi(argv[5]);
    seed = atol(argv[4]);
    epistasis = atof(argv[3]);
    effectsShape = atof(argv[2]);
    effectsMean = atof(argv[1]);

    if(epistasis == 0.0) symflag = 's';
    else symflag = 'a';

    sprintf(file1, "%s.in", argv[11]);

    outf.open(file1);
    outf << "Parameters for experiment: " << argv[11] << endl
        << "effects_mean = " << effectsMean << endl
        << "effects_shape = " << effectsShape << endl
        << "sample size = " << sampleSize << endl
        << "asymmetry = " << epistasis << endl
        << "distance of trait origin from optimum = " << trait0 << endl
        << "number of dimensions (n) = " << n << endl
        << "number of neutral dimensions (p) = " << p << endl
        << "order of saved mutations (K) = " << order << endl
        << "sample size (N) = " << N << endl
        << "symmetry in pleiotropy? " << (symflag == 's' ? "yes" : "no") << endl;
    outf << endl << " seed for random number generator = " << seed << endl;
    outf.close();

    gsl_rng_set(rng, seed);

    sprintf(file1, "%s.m.dat", argv[11]);
    sprintf(file2, "%s.n.dat", argv[11]);
    sprintf(file3, "%s.o.dat", argv[11]);
    sprintf(file4, "%s.s.dat", argv[11]);
    sprintf(posf1, "%s.pos", argv[11]);
}

```

```

posf.open(posf1);

nmut[trial] = 0;
nfix[trial] = 0;
nben[trial] = 0;

wPlus = exp(-0.5);
cout << endl
    << "
                beneficial          total beneficial"
    << endl
    << "trial  mutations tried  mutations tried  fixes      mutations      "
    << "  total fixations  distance to opt    fitness    " << endl
    << "-----"
    << "-----"
    << "-----" << endl
    << setw(6) << trial + 1 << " "
    << setw(15) << mutations << " "
    << setw(15) << nmut[trial] + 1 << " "
    << setw(6) << nfix[trial] << " "
    << setw(15) << wPlus << endl;
// start trials -- N total
trial = 0;
while(trial < N) {

// find matrix describing mutation bias
if(symflag != 's') {
    scaleProduct = 0.0;
    for(j = 0; j < n; j++) {
        scale[j] = gsl_ran_gamma(rng, epistasis, 1.0 / epistasis);
        if(fabs(scale[j]) < 1e-50) scale[j] = 0.0;
        scaleProduct += scale[j];
    }
    scaleProduct /= n;
    for(j = 0; j < n; j++) {
        scale[j] /= scaleProduct;
    }
} else for(j = 0; j < n; j++) scale[j] = 1.0;
posf << "axes mean and variance:\t" << gsl_stats_mean(scale, n) << '\t'
    << gsl_stats_est_variance(scale, n) << endl;

count = 0;
mutations = 0;
nmut[trial] = 0;
nben[trial] = 0;
nfix[trial] = 0;

if(order == 0 || (order > 0 && trial == 0)) {
    sprintf(hist, "%s.%d.dat", argv[11], trial + 1);
    outf.open(hist);
}

// find starting point -- fitness should be proportional to 0.607 (= exp(-1/2))
z0[0] = scale[0];
for(j = 1; j < n; j++) z0[j] = 0.0;
w = exp(-0.5);
wStart = w;
rStart = z0[0];
posf << trial << '\t' << nmut[trial] << '\t' << nfix[trial] << ":\n";
    for(k = 0; k < n; k++) posf << ' ' << z0[k];
    posf << endl;
}

```

```

// don't bother going further if not going to keep sample point anyway,
// if f > 0.9, or if done
while(w < 0.960653066 && (nmut[trial] < 100000000) &&
      (order == 0 || nfix[trial] + 1 <= order)) {

    // set effect of mutation to minimum
    wPot = 0.0;
    for(j = 0; j < sampleSize; j++) {

        // generate mutation and find potential new location
        w1 = 0.0;
        abs_r = 0.0;
        if(effectsShape == 1.0) r = gsl_ran_exponential(rng, effectsMean);
        else r = gsl_ran_gamma(rng, effectsShape, effectsMean / effectsShape);
        if(fabs(r) < 1e-50) r = 0.0;
        for(k = 0; k < n; k++) {
            disp[k] = gsl_ran_gaussian(rng, 1.0);
            if(fabs(disp[k]) < 1e-50) disp[k] = 0.0;
            abs_r += pow(disp[k], 2.0);
        }
        abs_r = r / sqrt(abs_r);
        for(k = 0; k < n; k++) {
            z1[k] = z0[k] + (disp[k] * abs_r);
            w1 += pow(z1[k], 2.0)/scale[k];
        }
        if(w1 > 1300) w1 = 1300;
        w1 = exp(-w1/2.0);
        if(w1 > wPot) {
            wPot = w1;
            for(k = 0; k < n; k++) {
                disp2[k] = disp[k];
                z2[k] = z1[k];
                rPlus = r;
            }
        }
    } // for(j = 0; j < sampleSize; j++)

    if((++mutations * sampleSize) % 10000 == 0) {
cout      << argv[11] << " "
          << setw(6) << trial + 1 << " "
          << setw(15) << mutations << " "
          << setw(15) << nmut[trial] + 1 << " "
          << setw(6) << nfix[trial] << " "
          << setw(6) << w << " "
          << setw(18) << wPot << endl;
    }

    ++nmut[trial];
    // we've found the best of <sampleSize> mutations:
    // proceed with screen for fixation
    if(wPot > w) { // a beneficial mutation
        ++nben[trial];
        // favorable: find selection coefficient,
        // calculate prob of fixation and seive for fix'n
        s = wPot/w - 1.0;
        if(gsl_ran_flat(rng, 0.0, 1.0) <= 1.0 - exp(-2.0 * s)) {
            // we have a winner: store all scaled effects for later plotting
            ++nfix[trial];
            for(k = 0; k < n; k++) z0[k] = z2[k];
            if(count == NMax) break;
            //      rmOut[count] = rPlus/(trait * sqrt(n));
            //      roOut[count] = rPlus/(distance * sqrt(n));
        }
    }
}

```



```

    outf << trial + 1 << '\t' << nmut[trial] << '\t' << nben[trial]
        << '\t' << nfix[trial] << '\t'
        << w << '\t' << wPot << '\t' << rPlus/sqrt(n) << '\t' << s
        << '\t' << wPot - w << '\t' << (wPot - w)/w << endl;
    posf << trial << '\t' << nmut[trial] << '\t' << nfix[trial] << ":\n";
    for(k = 0; k < n; k++) posf << ' ' << z0[k];
    posf << endl << endl;
    w = wPot;
}
}
} // while(w < 0.960653066 && fixations < NMax)
trial++;
outf.close();
} // while(trial < N);
posf.close();
}

```

Appendix B: Source code of program ‘fixx’

“fixx” simulates adaptive walks through Fisher’s trait-space but allowing the presence of multiple genotypes and clonal competition.

The program is written in C++ and the libraries described in appendix A.

MAKEFILE

```
SHELL = /bin/sh

LIBHOME = /usr/local/lib

INCFLAGS = -I/usr/include/g++-2 -I/usr/local/include -I/usr/local/include/gsl
LIBFLAGS = -L./ -L/usr/local/lib/gsl
RANLIBS = -lgslrandist -lgslrng -lgslstatistics
BLASLIBS = -llapack -lblas -latlas
STDLIBS = -lstdc++ -lm

$(OBJS) : $(SRC)
$(CC) $(INCFLAGS) $(FLAGS) -c $(SRC)
SRCDIR = ~/programs/fixx
PROG = fixx
CC = g++
LIBS = $(RANLIBS) $(BLASLIBS) $(STDLIBS)
LINKER = g77
#FLAGS = -g -O
FLAGS = -g
OBJS = fixx.o
SRC = fixx.cc

$(PROG) : $(OBJS) $(UTIL)
$(LINKER) $(FLAGS) -o $(PROG) $(OBJS) $(LIBFLAGS) $(LIBS)

fixx.o : fixx.cc
$(CC) -c fixx.cc $(INCFLAGS) $(FLAGS)
```

“FIXX.H”

```
#include <string.h>
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <set>
#include <iomanip.h>
#include <fstream.h>
#include <float.h>
extern "C" {
#include <gsl/gsl_heapsort.h>
#include <gsl/gsl_randist.h>
#include <gsl/gsl_rng.h>
#include <gsl/gsl_vector.h>
```

```

#include <gsl/gsl_blas.h>
#include <gsl/gsl_test.h>
#include <gsl/gsl_statistics.h>
}

#define true 1
#define false 0
#define boolean int

const double PI = 3.141592653589793238462643;
const double NSCALE = 0.39894228; // = 1/sqrt(2*PI)
const double SCALE = 3.0;
#define nMax 50
#define NMax 10000

typedef double(*PF)(double, double, gsl_vector*);

double gaussian(double scale, double width, gsl_vector* data) {
    double ddot;
    gsl_blas_ddot(data, data, &ddot);
    return (scale * exp(-(ddot/(2.0 * width * width))));
};

class genome;

class mutation {
    gsl_vector *x, *x0;
    set<genome*> isIn;
    double size, s, length;
    static int ndims;
    static PF pwf; // pointer to fitness function
    static gsl_rng *rng;
    static double alpha, beta;
public:
    mutation();
    mutation(gsl_vector*);
    mutation(gsl_vector*, gsl_vector*);
    ~mutation() { gsl_vector_free(x0); gsl_vector_free(x); }
    static void set_pwf(PF pf) { pwf = pf; }
    static void set_rng(gsl_rng* r) { rng = r; }
    static void set_params(double a, double b) { alpha = a; beta = b; }
    static void set_ndims(int i) { ndims = i; }
    void Start(gsl_vector* v) { gsl_vector_memcpy(x0, v); }
    gsl_vector* Start() { return x0; }
    gsl_vector* X() { return x; }
    double Set(gsl_vector*);
    double Set(gsl_vector*, gsl_vector*);
    double S() { return s; }
    double S(gsl_vector*);
    double S(double d) { s = d; return s; }
    double Size() { return size; }
    double Size(double d) { size = d; return size; }
    double Length();
    int IsIn();
    boolean IsIn(genome* g) { if(isIn.find(g) != isIn.end()) return true; }
    void Add(genome* g) { isIn.insert(g); }
    void Remove(genome* g) { isIn.erase(g); }
};

class genome {
private:
    gsl_vector *x;

```

```

    set<mutation*> has;
    double distance, size, w, s;
    static PF pwf; // pointer to fitness function
    static int ndims;
public:
    genome(genome* g) {
        size = 1.0;
        has = *(g->sHas());
        x = gsl_vector_alloc(ndims);
        gsl_vector_memcpy(x, g->X());
    }
    genome() {
        has.clear();
        x = gsl_vector_calloc(ndims);
        size = 0.0;
        w = 0.0;
        s = 0.0;
    }
    genome(gsl_vector* v) {
        has.clear();
        x = gsl_vector_alloc(ndims);
        gsl_vector_memcpy(x, v);
        size = 0.0;
        w = (*pwf)(SCALE, 1.0, x);
        s = 0.0;
    }
    ~genome() { gsl_vector_free(x); }
    void Print() {
        cout << endl;
        for(int i = 0; i < ndims; i++)
            cout << gsl_vector_get(x, i) << '\t';
        cout << endl;
    }
    static void set_pwf(PF pf) { pwf = pf; }
    static void set_ndims(int i) { ndims = i; }
    double W();
    double W(double d) { w = d; return w; }
    double S(double);
    double Size() { return size; }
    double Size(double d) { size = d; return size; }
    double Mutate(mutation*);
    double Spawn(mutation*);
    int Has();
    double Has(mutation* m) {
        if(has.find(m) != has.end())
            return size;
        else
            return 0.0;
    }
    double Distance() {
        distance = 0.0;
        gsl_blas_ddot(x, x, &distance);
        distance = sqrt(distance);
        return distance;
    }
    int Insert(mutation* m) { has.insert(m); return has.size(); }
    int CopyTo(genome* g) {
        if(!has.empty()) {
            set<mutation*>::iterator mptr;
            for(mptr = has.begin(); mptr != has.end(); mptr++)
                g->Insert((*mptr));
        }
    }
}

```

```

        return has.size();
    }
    set<mutation*>* sHas() { return &has; }
    gsl_vector* X() { return x; }
    gsl_vector* Add(mutation* m) {
        gsl_vector_add(x, m->X());
        has.insert(m);
        m->Add(this);
        return x;
    }
};

mutation::mutation() {
    isIn.clear();
    length = 0.0;
    s = 0.0;
    x = gsl_vector_calloc(ndims);
    x0 = gsl_vector_calloc(ndims);
};

mutation::mutation(gsl_vector* st) {
    double w0, w1, r, scale;
    isIn.clear();
    gsl_vector *sum;
    sum = gsl_vector_calloc(ndims);
    length = 0.0;
    s = 0.0;
    x = gsl_vector_calloc(ndims);
    x0 = gsl_vector_calloc(ndims);
    gsl_vector_memcpy(x0, st);
    if(alpha == 1.0) r = gsl_ran_exponential(rng, alpha * beta);
    else r = gsl_ran_gamma(rng, alpha, beta);
    for(int i = 0; i < ndims; i++) gsl_vector_set(x, i, gsl_ran_gaussian(rng, 1.0));
    gsl_blas_ddot(x, x, &scale);
    gsl_vector_scale(x, r/sqrt(scale));
    gsl_blas_ddot(x, x, &length);
    length = sqrt(length);
    w0 = (*pwf)(SCALE, 1.0, x0);
    gsl_vector_add(sum, x0);
    gsl_vector_add(sum, x);
    w1 = (*pwf)(SCALE, 1.0, sum);
    s = w1/w0 - 1.0;
    gsl_vector_free(sum);
};

mutation::mutation(gsl_vector* st, gsl_vector* weights) {
    isIn.clear();
    length = 0.0;
};

double mutation::Set(gsl_vector* st) {
    gsl_vector_memcpy(x, st);
    length = 0.0;
    gsl_blas_ddot(x, x, &length);
    length = sqrt(length);
    return length;
};

double mutation::Set(gsl_vector* st, gsl_vector* en) {
    gsl_vector_memcpy(x0, st);
    gsl_vector_memcpy(x, en);
    length = 0.0;
};

```

```

    gsl_blas_ddot(x, x, &length);
    length = sqrt(length);
    return length;
};

double mutation::S(gsl_vector* start) {
    double w0, w1;
    gsl_vector *sum;
    sum = gsl_vector_calloc(ndims);
    if(s == 0.0) {
        gsl_vector_memcpy(x0, start);
        w0 = (*pwf)(SCALE, 1.0, x0);
        gsl_vector_add(sum, x0);
        gsl_vector_add(sum, x);
        w1 = (*pwf)(SCALE, 1.0, sum);
        s = w0/w1 - 1.0;
    }
    gsl_vector_free(sum);
    return s;
};

double mutation::Length(void) {
    if(length == 0.0) {
        gsl_blas_ddot(x, x, &length);
        length = sqrt(length);
    }
    return length;
};

int mutation::IsIn() {
    set<genome*>::iterator gp = isIn.begin();
    if(isIn.empty()) cout << "none" << endl;
    else
        do {
            cout << (*gp) << endl;
            gp++;
        } while(gp != isIn.end());
    return isIn.size();
};

double genome::Mutate(mutation* m) {
    m->Add(this);
    has.insert(m);
    m->Start(x);
    gsl_vector_add(x, m->X());
    w = (*pwf)(SCALE, 1.0, x);
    return w;
};

double genome::W() {
    w = (*pwf)(SCALE, 1.0, x);
    return w;
};

int genome::Has() {
    return has.size();
};

```

```

“FIXX.CC”
#include "fixx.h"

PF mutation::pwf = &gaussian;
gsl_rng* mutation::rng = NULL;
double mutation::alpha = 1.0;
double mutation::beta = 1.0;
int mutation::ndims = 50;
PF genome::pwf = &gaussian;
int genome::ndims = 50;

void main(int argc, char *argv[]) {

    gsl_rng_env_setup();
    gsl_rng *rng = gsl_rng_alloc(gsl_rng_ranlux);
    gsl_vector *start, *origin, *zero;

    double alpha, beta, N = 1.0e10, N_prime, mu = 1.0e-6,
        goal = SCALE * 0.960653066, w_bar, s, tLoss, count;
    unsigned long int seed, nmutations = 0L;
    int ndims = 50, nfix[NMax], nben[NMax], nmut[NMax], trial, ntrials = 100,
        i, j, k, l, m, generation, log = 100;
    boolean adapted;
    mutation *mtest;
    genome *g, *gtest, *deadgenome;
    set<mutation*> mutations, newmutations, losses;
    set<mutation*>::iterator mptr;
    set<genome*> population, offspring, boneyard;
    set<genome*>::iterator gptra;
    char file1[20], file2[20], file3[20], file4[20],
        hist[20], posf1[20], symflag;
    ofstream outf1, outf2, posf;

    if(argc == 1) {
        cout << "fixx <file> <alpha> <beta> <seed> <dims> <trials> "
            << "<N> <mu>" << endl;
        exit(0);
    }

    log = atoi(argv[9]);
    mu = atof(argv[8]);
    N = atof(argv[7]);
    ntrials = atoi(argv[6]);
    ndims = atoi(argv[5]);
    seed = atol(argv[4]);
    beta = atof(argv[3]);
    alpha = atof(argv[2]);
    sprintf(file1, "%s.in", argv[1]);

    outf1.open(file1);
    outf1 << "Parameters for experiment: " << argv[1] << endl
        << "alpha = " << alpha << endl
        << "beta = " << beta << endl
        << "mu = " << mu << endl
        << "number of dimensions (n) = " << ndims << endl
        << "Population Size (N) = " << N << endl
        << "seed for random number generator = " << seed << endl;
    outf1.close();

    cout << "Starting experiment " << argv[1] << " with the following parameters:"
        << endl << "alpha = " << alpha << "; "

```

```

        << "beta = " << beta << "; "
        << "mu = " << mu << "; "
        << "number of dimensions (n) = " << ndims << "; "
        << "Population Size (N) = " << N << endl
        << "seed for random number generator = " << seed << endl;
if(log) cout << "logging to " << argv[1] << ".<trial>.dat every "
        << log << " generations." << endl;
else cout << "Not logging generation-by-generation data" << endl;

gsl_rng_set(rng, seed);
mutation::set_params(alpha, beta);
mutation::set_rng(rng);
mutation::set_ndims(ndims);
genome::set_ndims(ndims);

start = gsl_vector_calloc(ndims);

gsl_vector_add_constant(start, 1.0/sqrt((double) ndims));
g = new genome(start);

sprintf(file1, "%s.dat", argv[1]);
outf1.open(file1);

for(trial = 0; trial < ntrials; trial++) {

    g = new genome(start);
    g->Size(N);
    sprintf(file2, "%s.%d.dat", argv[1], trial + 1);
    outf2.open(file2);

    adapted = false;
    nmut[trial] = 0;
    nben[trial] = 0;
    nfix[trial] = 0;
    generation = 0;
    population.clear();
    mutations.clear();
    offspring.clear();
    newmutations.clear();
    boneyard.clear();
    losses.clear();
    population.insert(g);

    while(!adapted) {

        // expand population by fitness, find w_bar
        w_bar = 0.0;
        N_prime = 0.0;
        gp_ptr = population.begin();
        while(gp_ptr != population.end()) {
            N_prime += (*gp_ptr)->Size() * (*gp_ptr)->W();
            w_bar += (*gp_ptr)->Size() * (*gp_ptr)->W();
            gp_ptr++;
        }
        w_bar /= N_prime;
        N_prime = N/N_prime;

        gp_ptr = population.begin();
        while(gp_ptr != population.end()) {

            // normalize genome population so that total population = N

```



```

(*gpPtr)->Size((*gpPtr)->Size() * N_prime);

// calculate selection coefficient
s = ((*gpPtr)->W()/w_bar) - 1.0;

// threshold for stochastic/deterministic behavior
if((*gpPtr)->Size() < N * 0.001) tLoss = 1.0 / s;
else tLoss = N * 0.001;

// kill genome if s = 0 and of very small size
if(s < 0.0 && (*gpPtr)->Size() < N * 0.001) {
    //delete (*gpPtr);
    // population.erase(gpPtr);
    boneyard.insert((*gpPtr));
}
// if below threshold then test for loss
if((*gpPtr)->Size() < tLoss) {

    // lost
    if(gsl_ran_flat(rng, 0.0, 1.0) < exp(-2.0 * s)) {
        // delete (*gpPtr);
        // population.erase(gpPtr);
        boneyard.insert((*gpPtr));

    // not lost: at threshold
    } else
        (*gpPtr)->Size(tLoss);

// not below threshold: add mutants
} else {

    // number of mutant is binomially distributed
    // (double mutants don't happen)
    nmutations = gsl_ran_binomial(rng, mu,
        (unsigned long int)(floor((*gpPtr)->Size())));

    // generate <nmutations> new mutants to the current genome
    for(i = 0; i < nmutations; i++) {
        ++nmut[trial];
        mtest = new mutation((*gpPtr)->X());
        gtest = new genome((*gpPtr));
        gtest->Mutate(mtest);
        if(gtest->W() < (*gpPtr)->W()) {
            delete mtest;
            delete gtest;
        } else {
            ++nben[trial];
            s = (gtest->W()/w_bar) - 1.0;
            if(gsl_ran_flat(rng, 0.0, 1.0) < exp(-2.0 * s)) {
                delete mtest;
                delete gtest;
            } else {
                ++nfix[trial];
                gtest->Size(1.0 / s);
                offspring.insert(gtest);
                newmutations.insert(mtest);
            }
        }
    }
}
gpPtr++;
} // while(gpPtr != population.end())

```

```

if(!offspring.empty()) {
    gptr = offspring.begin();
    while (gptr != offspring.end()) {
        population.insert(*gptr);
        offspring.erase(gptr);
        ++gptr;
    }
}
if(!boneyard.empty()) {
    gptr = boneyard.begin();
    while (gptr != boneyard.end()) {
        population.erase(*gptr);
        delete (*gptr);
        boneyard.erase(gptr);
        ++gptr;
    }
}
if(!newmutations.empty()) {
    mptr = newmutations.begin();
    while(mptr != newmutations.end()) {
        mutations.insert(*mptr);
        newmutations.erase(*mptr);
        ++mptr;
    }
}
if(!mutations.empty()) {
    mptr = mutations.begin();
    while(mptr != mutations.end()) {
        count = 0.0;
        gptr = population.begin();
        while(gptr != population.end()) {
            count += (*gptr)->Has(*mptr);
            ++gptr;
        }
        if(count == 0.0) {
            losses.insert(*mptr);
        }
        ++mptr;
    }
}
if(!losses.empty()) {
    mptr = losses.begin();
    while(mptr != losses.end()) {
        mutations.erase(*mptr);
        delete(*mptr);
        losses.erase(mptr);
        ++mptr;
    }
}

if(generation % log == 0 && log)
    outf2 << trial + 1 << '\t' << generation + 1 << '\t' << w_bar << '\t'
        << nmut[trial] << '\t' << nben[trial] << '\t'
        << nfix[trial] << '\t' << population.size() << '\t'
        << mutations.size() << endl;
if(generation % 1000 == 0)
    cout << "." << flush;

boneyard.clear();
losses.clear();
offspring.clear();

```

```

    newmutations.clear();
    ++generation;
    if(w_bar > goal) adapted = true;
}
k = 0;
N_prime = 0.0;
for(gpctr = population.begin(); gpctr != population.end(); ++gpctr) {
    N_prime += (*gpctr)->Size();
}

cout << endl << trial + 1 << '\t' << nmut[trial] << '\t'
    << nben[trial] << '\t' << nfix[trial] << endl;

k = 0;
mptr = mutations.begin();
outf1 << "#" << trial + 1 << '\t' << nmut[trial] << '\t'
    << nben[trial] << '\t' << nfix[trial] << endl;
while(mptr != mutations.end()) {
    count = 0.0;
    gpctr = population.begin();
    while(gpctr != population.end()) {
        count += (*gpctr)->Has((*mptr));
        ++gpctr;
    }
    count /= N_prime;
    // if(count > 0.15)
    outf1 << trial + 1 << '\t' << count << '\t'
        << (*mptr)->Length() << '\t' << (*mptr)->S() << endl;
    delete(*mptr);
    ++mptr;
}
while(gpctr != population.end()) {
    delete (*gpctr);
    ++gpctr;
}
outf2.close();
}
outf1.close();
}

```

Literature Cited

- Anderson, R. M. (1998). "Tuberculosis: old problems and new approaches." Proc Natl Acad Sci U S A **95**(23): 13352-4.
- Anderson, R. M. (1999). "The pandemic of antibiotic resistance [news]." Nat Med **5**(2): 147-9.
- Anderson, R. M. and R. M. May (1992). "Understanding the AIDS pandemic." Sci Am **266**(5): 58-61, 64-6.
- Austin, D. J., K. G. Kristinsson, et al. (1999). "The relationship between the volume of antimicrobial consumption in human communities and the frequency of resistance [see comments]." Proc Natl Acad Sci U S A **96**(3): 1152-6.
- Bjorkman, J., I. Nagaev, et al. (2000). "Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance." Science (Washington D C) **287**(5457): 1479-1482.
- Borrias, W. E., M. Hagenaar, et al. (1979). "Functional relationship between bacteriophages G4 and phi X174." J Virol **31**(2): 288-98.
- Bull, J. J., M. R. Badgett, et al. (2000). "Big-benefit mutations in a bacteriophage inhibited with heat." Mol Biol Evol **17**(6): 942-50.
- Bull, J. J., M. R. Badgett, et al. (1997). "Exceptional convergent evolution in a virus." Genetics **147**(4): 1497-507.

- Bull, J. J., A. Jacobson, et al. (1998). "Viral escape from antisense RNA." Mol Microbiol **28**(4): 835-46.
- Burch, C. L. and L. Chao (1999). "Evolution by small steps and rugged landscapes in the RNA virus phi6." Genetics **151**(3): 921-7.
- Chao, L., T. Tran, et al. (1992). "Muller's Ratchet and the Advantage of Sex in the RNA Virus Phi-6." Evolution **46**(2): 289-299.
- Crill, W. D., H. A. Wichman, et al. (2000). "Evolutionary reversals during viral adaptation to alternating hosts." Genetics **154**(1): 27-37.
- Delaporte, C., L. Gros, et al. (1999). "Selection of genetic suppressor elements conferring resistance to DNA topoisomerase II inhibitors." Anticancer Molecules: Structure, Function, and Design **886**: 187-190.
- Domingo, E., E. Baranowski, et al. (1998). "Quasispecies structure and persistence of RNA viruses." Emerg Infect Dis **4**(4): 521-7.
- Domingo, E., C. Escarmis, et al. (1998). "Population dynamics in the evolution of RNA viruses." Adv Exp Med Biol **440**: 721-7.
- Domingo, E., C. Escarmis, et al. (1996). "Basic concepts in RNA virus evolution." Faseb J **10**(8): 859-64.
- Drake, J. W. (1991). "A constant rate of spontaneous mutation in DNA-based microbes." Proc Natl Acad Sci U S A **88**: 7160-7164.
- Drake, J. W. (1994). "Rates of spontaneous mutation among RNA viruses." Proc Natl Acad Sci U S A **90**: 4171-4175.

- Dunn, S. J., S. W. Park, et al. (1999). "Isolation of efficient antivirals: genetic suppressor elements against HIV-1." Gene Ther **6**(1): 130-7.
- Ekechukwu, M. C., O. D. J., et al. (1995). "Host and phi-X-174 mutations affecting the morphogenesis or stabilization of the 50S complex, a single-stranded DNA synthesizing intermediate." Genetics **140**(4): 1167-1174.
- Elena, S. F., M. Davila, et al. (1998). "Evolutionary dynamics of fitness recovery from the debilitating effects of Muller's ratchet." Evolution **52**(2): 309-314.
- Escarmis, C., M. Davila, et al. (1996). "Genetic lesions associated with Muller's ratchet in an RNA virus." J Mol Biol **264**(2): 255-67.
- Escarmis, C., M. Davila, et al. (1999). "Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet." J Mol Biol **285**(2): 495-505.
- Fane, B. A., S. Shien, et al. (1993). "Second-site suppressors of a cold-sensitive external scaffolding protein of bacteriophage ϕ X174." Genetics **134**: 1003-1011.
- Fisher, R. A. (1930). The Genetical Theory of Natural Selection. London, Oxford University Press.
- Gerrish, P. J. and R. E. Lenski (1998). "The fate of competing beneficial mutations in an asexual population." Genetica **103**(1-6): 127-44.

- Hayashi, M. N. and M. Hayashi (1968). "The stability of native DNA-RNA complexes during in vivo phiX-174 transcription." Proc Natl Acad Sci U S A **61**(3): 1107-14.
- Hayashi, M. N. and M. Hayashi (1985). "Cloned DNA sequences that determine mRNA stability of bacteriophage phi X174 in vivo are functional." Nucleic Acids Res **13**(16): 5937-48.
- Hayashi, M. N., M. Hayashi, et al. (1983). "Role for the J-F intercistronic region of bacteriophages phi X174 and G4 in stability of mRNA." J Virol **48**(1): 186-96.
- Hayashi, M. N., R. Yaghami, et al. (1989). "mRNA stabilizing signals encoded in the genome of the bacteriophage phi x174." Mol Gen Genet **216**(2-3): 364-71.
- Holzmayr, T. A., D. G. Pestov, et al. (1992). "Isolation of dominant negative mutants and inhibitory antisense RNA sequences by expression selection of random DNA fragments." Nucleic Acids Res **20**(4): 711-7.
- Jazwinski, S. M. and A. Kornberg (1975). "DNA replication in vitro starting with an intact phiX174 phage." Proc Natl Acad Sci U S A **72**(10): 3863-7.
- Jazwinski, S. M., A. A. Lindberg, et al. (1975). "The gene H spike protein of bacteriophages phiX174 and S13. I. Functions in phage-receptor recognition and in transfection." Virology **66**(1): 283-93.

- Jazwinski, S. M., A. A. Lindberg, et al. (1975). "The lipopolysaccharide receptor for bacteriophage phiX174 and S13." Virology **66**(1): 268-82.
- Jazwinski, S. M., R. Marco, et al. (1975). "The gene H spike protein of bacteriophages phiX174 and S13. II. Relation to synthesis of the parenteral replicative form." Virology **66**(1): 294-305.
- Kimura, M. (1983). The Neutral Theory of Molecular Evolution. Cambridge, Cambridge University Press.
- Lande, R. (1983). "The response to selection on major and minor mutations affecting a metrical trait." Heredity **50**(1): 47-65.
- Lederberg, J. (1998). "The future of infectious diseases." J Urban Health **75**(3): 463-70.
- Lederberg, J. (2000). "Infectious history." Science **288**(5464): 287-93.
- Levin, B. R., V. Perrot, et al. (2000). "Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria." Genetics **154**: 985-997.
- Linney, E. A., M. N. Hayashi, et al. (1972). "Gene A of X174. 1. Isolation and identification of its products." Virology **50**(2): 381-7.
- Lynch, M., J. Conery, et al. (1995). "Mutation accumulation and the extinction of small populations." Am Nat **146**(4): 489-518.
- Maniatis, T., E. F. Fritsch, et al. (1982). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

- Moore, F. B., D. E. Rozen, et al. (2000). "Pervasive compensatory adaptation in *Escherichia coli*." Proc R Soc Lond B Biol Sci **267**(1442): 515-22.
- Muller, H. J. (1932). "Some genetic aspects of sex." Am Nat **8**: 118-138.
- Novella, I. S., E. A. Duarte, et al. (1995). "Exponential increases of RNA virus fitness during large population transmissions." Proceedings of the National Academy of Sciences USA **92**: 5841-5844.
- Novella, I. S., S. F. Elena, et al. (1995). "Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness." J. Virol. **69**(5): 2869-2872.
- Orr, H. A. (1998). "The population genetics of adaptation: the distribution of factors fixed during adaptive evolution." Evolution **52**(4): 935-949.
- Orr, H. A. (1999). "The evolutionary genetics of adaptation: a simulation study." Genet Res **74**(3): 207-14.
- Orr, H. A. (2000). "Adaptation and the cost of complexity." Evolution **54**(1): 13-20.
- Otto, S. P. and M. C. Whitlock (1997). "The probability of fixation in populations of changing size." Genetics **146**(2): 723-33.
- Paquin, C. and J. Adams (1983). "Relative fitness can decrease in evolving asexual populations of *S. cerevisiae*." Nature **306**: 368-371.

- Pol, J. H. v. d. and G. A. v. Arkel (1965). "The inactivating and mutagenic effect of hydroxylamine on bacteriophage phi x-174." Mutat Res 2(5): 466-9.
- Sambrook, J., T. Maniatis, et al. (1989). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.
- Schuster, P. and K. Sigmund (1989). "Fixation probabilities for advantageous mutants: a note on multiplication and sampling." Math Biosci 95(1): 37-51.
- Schuster, P. and J. Swetina (1988). "Stationary mutant distributions and evolutionary optimization." Bull Math Biol 50(6): 635-60.
- Siegel, J. E., M. N. Hayashi, et al. (1968). "Phi-X-174 coat protein mutants affecting DNA synthesis." Biochem Biophys Res Commun 31(5): 774-8.
- Steinhauer, D. A. and J. J. Holland (1987). "Rapid evolution of RNA viruses." Ann. Rev. Microbiol. 41: 409-433.
- Stephan, W., L. Chao, et al. (1993). "The advance of Muller's ratchet in a haploid asexual population: Approximate solutions based on diffusion theory." Genetical Research 61(3): 225-231.
- Swetina, J. and P. Schuster (1982). "Self-replication with errors. A model for polynucleotide replication." Biophys Chem 16(4): 329-45.

- Tessman, I. (1968). "Mutagenic treatment of double- and single-stranded DNA phages T4 and S13 with hydroxylamine." Virology **35**(2): 330-3.
- Van der Avoort, H. G., R. Teertstra, et al. (1983). "Genes and regulatory sequences of bacteriophage phi X174." Biochim Biophys Acta **741**(1): 94-102.
- van der Avoort, H. G., G. A. van Arkel, et al. (1982). "Cloned bacteriophage phi X174 DNA sequence interferes with synthesis of the complementary strand of infecting bacteriophage phi X174." J Virol **42**(1): 1-11.
- van der Avoort, H. G., A. van der Ende, et al. (1984). "Regions of incompatibility in single-stranded DNA bacteriophages phi X174 and G4." J Virol **50**(2): 533-40.
- Wichman, H. A., M. R. Badgett, et al. (1999). "Different trajectories of parallel evolution during viral adaptation." Science **285**(5426): 422-4.

Vita

Matthew Jonas Brauer was born in San Diego, California on 15 January 1967, the son of Bryan B. Brauer and Beth E. Jonasson. He grew up in Oakland California and entered the University of California at Berkeley in the summer of 1984. While at Berkeley, he worked in the labs of the Chiron Corporation. He graduated in December 1988 with a Bachelor of Arts degree in Biochemistry, and continued working at Chiron for the next two years. In 1990, he left biotechnology and attended the summer course in coastal field ecology at Bodega Bay. He spent the next several years in seasonal field biology jobs in Alaska and California. In August 1994, Brauer moved to Austin to join his new wife, Jennifer Murphy, in the Graduate School at the University of Texas. He completed a Master's of Science in Statistics in the summer of 1998. Brauer is beginning postdoctoral work in the lab of David Botstein in the Department of Genetics at the Stanford University School of Medicine.

His first child, Iris Fiona, was born 11 October 2000, roughly coincident with the completion of this dissertation.

Permanent address: 2627 Lawton St.
San Francisco, CA 94122

This dissertation was typed by the author.