



This is to certify that the

dissertation entitled

DYNAMICS OF MUTATION AND SELECTION
IN ASEXUAL POPULATIONS

presented by

Philip J. Gerrish

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Zoology

Richard E. Lenski
Major professor

Date April 2, 1998

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>

DYNAMICS OF MUTATION AND SELECTION
IN ASEXUAL POPULATIONS

By

Philip J. Gerrish

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Zoology

1998

ABSTRACT

DYNAMICS OF MUTATION AND SELECTION IN ASEXUAL POPULATIONS

By

Philip J. Gerrish

The dynamics of asexual populations are characterized by strong associations between alleles at different loci due to complete linkage. One consequence of complete linkage is the phenomenon known as hitchhiking, whereby a specified mutant allele is driven to high frequency or even fixation by a beneficial mutation to which it is linked. If the specified mutant allele confers a change in the mutation rate of the organism, then the fixation of that allele affects the mean mutation rate of the population. Observations of high mutation rates in evolving *E. coli* populations are consistent with the hypothesis that mutator alleles were fixed in these populations by hitchhiking. Theoretical exploration of the hitchhiking mechanism reveals its general significance for asexual evolution. I conclude that (i) mutation rates found in asexual populations are more likely to be determined by sporadic hitchhiking than by evolutionary "fine-tuning" as previous theoretical models would suggest, and (ii) there exists a most probable increase in mutation rate due to hitchhiking that is both positive and finite and its value is typically significant.

As
national
station
alternat
dional
paramete
transien
which an
the case
solve th
fitness
rate, 1
of avail
limited.
adaptive
punctuat
stasis r
environm
populati
rate of
rate and
rate of
with sig
maintain

Another consequence of complete linkage is the phenomenon called "clonal interference", whereby the progression of one beneficial mutation to fixation is hindered or even prevented by competition with alternative beneficial mutations. From theoretical exploration of clonal interference, I derive several fundamental population-genetic parameters, as well as the probability that a beneficial mutation will transiently achieve polymorphic frequency. After treating the case in which an unlimited number of beneficial mutations are available, I treat the case in which that number is limited. From these developments, I solve the inverse problem of estimating the following parameters from fitness data of evolving *E. coli* populations: (i) beneficial mutation rate, (ii) the distribution of mutational effects, and (iii) the number of available beneficial mutations for the case in which this number is limited. Salient conclusions from both theoretical treatments are (i) adaptive evolution of asexual populations is characteristically punctuated with short bursts of rapid change followed by long periods of stasis regardless of population size or mutation rate, (ii) in identical environments, the trajectory of adaptive evolution of large, parallel populations is highly repeatable due to clonal interference, (iii) the rate of fitness improvement is an increasing function of both mutation rate and population size, but the function is decelerating so that the rate of adaptive evolution is constrained by a "speed limit", and (iv) with significant probability, clonal interference may transiently maintain fitness variants at polymorphic frequencies.

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

ACKNOWLEDGMENTS

First and foremost, I thank my committee members: I thank my advisor, Dr. Richard Lenski, for admirable instruction, guidance and support as well as his vital role in the conception of ideas presented here, I thank Dr. Paul Sniegowski for his crystal clear explanations of population genetics as well as his vital role in the conception of ideas presented here, I thank Dr. V. Mandrekar for superior instruction in stochastic processes, I thank Dr. Judy Mongold and Dr. Don Hall for helpful comments and conversations. I especially thank Danny Rozen for important conceptual contributions, Drs. Alex White and Alejandra Sorto for help with mathematical and statistical analyses. I am grateful to Dr. Francois Taddei for sharing a manuscript prior to publication and to Dr. Cliff Zeyl for permission to cite unpublished data. I thank Dr. Brendan Bohannon, Lynette Ekunwe and Phyllis Frank for technical assistance, Drs. Tom Cebula, Fred Adler, Frank Stewart, Warren Ewens, Andrew Leigh Brown, Santiago Elena, Mike Travisano, John Gerrish, Larry Segerlind, Arjan DeVisser, and Paco Moore for helpful discussions and comments. Thanks to Drs. H. Maki and J.E. LeClerc for plasmids. Thanks to my present supervisor, Dr. Marcia Kalish, for patience and support. Thanks to Lucy Gerrish for moral support and good humor, Irene Gerrish for playing dinosaur, and Dr. Theophilus Okosun for plenty of beer.

LIST OF

LIST OF

C

C

C

INTRODU

F

C

C

C

C

Chapter

Ac

In

Me

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF MATHEMATICAL SYMBOLS	xiii
Chapter 2	xiii
Chapter 3	xiv
Chapter 4	xvi
INTRODUCTION	1
Fundamental issues and roadmap to the dissertation	1
Chapter 1	2
Chapter 2	6
Chapter 3	8
Chapter 4	9
Chapter 1: EVOLUTION OF HIGH MUTATION RATES IN EXPERIMENTAL POPULATIONS OF <i>ESCHERICHIA COLI</i>	11
Abstract	11
Introduction and Findings	12
Methods	14
Experimental System	14
Mutation rate measurements	15
Time of Origin and Persistence of Mutator Phenotypes	16
Complementation Tests	17
Analysis of Fluctuation Test Data	17

Discussion	18
Chapter 2: THE FATE OF COMPETING BENEFICIAL MUTATIONS	
IN AN ASEXUAL POPULATION	25
Abstract	25
Introduction	26
Clonal interference and fixation	28
Clonal interference among beneficial mutations	28
Some definitions	28
The expected number of interfering mutations	29
Fixation probability of a beneficial mutation	33
Expected rate of substitution	35
Expected selection coefficient of successful mutations	36
Effect of clonal interference on rate of fitness increase	37
Estimation of parameters: an empirical example	38
Transiently common mutations	42
Clonal interference - a general model	42
Probability of transiently polymorphic beneficial mutations	48
Probability of transiently common mutations: the leapfrog	52
Discussion	54
Summary of results	54
Assumptions of the models	56
Model validation by simulation	62
Inclusion of the double mutant	63
Implications for the evolution of	

Chapter 1

Ab

Int

Pro

Fi

Fi

Di

reproductive strategies	64
Implications for the general nature of adaptive evolution	67
Evidence for transiently common beneficial mutations in microbial populations	69
A suggestion for further research	72
Chapter 3: THE ORDER OF FIXATION OF BENEFICIAL MUTATIONS IN ASEXUAL POPULATIONS	84
Abstract	84
Introduction	85
Probabilities of fixation orderings	88
Order of fixation of two beneficial mutations	88
Probability that a superior mutation is the first of many	94
Probability that fixation ordering is from largest to smallest s	95
General solution for any ordering of fixations	96
Fitness trajectories I: parallel populations share a common set of available beneficial mutations	98
Expectation and variance	99
Reducing the number of parameters	103
Fitness trajectories II: parallel populations share a common distribution of beneficial mutational effects	106
The non-ordered region	107
The ordered region	110
Discussion	115
Assumptions of the models	115
Estimating parameters	118
Ordered substitutions in nature, with an application to HIV	122

Chapter 4: HITCHHIKING OF DELETERIOUS MUTATIONS	
WITH SPECIAL ATTENTION TO MUTATOR ALLELES	135
Abstract	135
Introduction	136
Theoretical developments	138
The hitchhiking rate	138
An effective mutation-selection balance	140
Probability of fixation of double-mutations, $Pr\{fix\}$.	142
Converting per-capita rate to populational probability	148
An application	151
Hitchhiking rate of mutator mutations	151
Estimating rate of mutation to mutator for evolving <i>E. coli</i> populations	154
Discussion	157
Assumptions	157
Comparison with simulation	159
Effect of population size on hitchhiking rate	162
Effect of selective disadvantage on hitchhiking rate .	163
Effect of mutator strength on hitchhiking rate	164
APPENDIX 1.1: FLUCTUATION TEST ANALYSIS PROGRAM	178
Theory	178
Obtaining and using the program	182
APPENDIX 2.1: PROBABILITY OF SURVIVING DRIFT	183
APPENDIX 2.2: n -GENOTYPE LOGISTIC SYSTEM WITH MUTATION	185
General solution	185
Application of boundary conditions due to mutation	187

No.

APPENDIX

APPENDIX

APPENDIX

APPENDIX

APPENDIX

APPENDIX

LIST OF

Notation for the 3-genotype case	188
APPENDIX 2.3: EXPECTED NUMBER OF CANDIDATE REPLICATIONS	190
APPENDIX 2.4: FUNCTIONS EMPLOYING THE RECTANGULAR DISTRIBUTION	194
APPENDIX 3.1: ADAPTIVE SUBSTITUTIONS IN ASEXUAL POPULATIONS ARE ACCURATELY MODELED AS INSTANTANEOUS REPLACEMENTS	197
APPENDIX 3.2: ANALYTICAL INTEGRATION OF EQUATIONS	200
APPENDIX 4.1: EFFECTIVE MUTATION-SELECTION BALANCE	204
APPENDIX 4.2: EFFECTIVE POPULATION NUMBER UNDER A SERIAL TRANSFER REGIME	208
LIST OF REFERENCES	218

LIST OF FIGURES

Figure 1.1. Rates of mutation in 12 Ara ⁻ and Ara ⁺ experimental populations (A-1 to A-6 and A+1 to A+6) at 10000 generations and in their common ancestors REL606 (Ara ⁻) and REL607 (Ara ⁺).	22
Figure 1.2. Time of appearance and evolutionary persistence of mutator phenotypes in experimental populations Ara ⁻ 2, Ara ⁻ 4, and Ara ⁺ 3.	23
Figure 1.3. Mutation rates to nalidixic acid resistance in 10,000-generation isolates from populations Ara ⁻ 2, Ara ⁻ 4 and Ara ⁺ 3 and ancestor REL606 transformed with plasmids bearing wild-type alleles of seven known general mutator loci.	24
Figure 2.1. The probability that a given beneficial mutation with selection coefficient, s , achieves fixation.	74
Figure 2.2. The probability of fixation of an arbitrarily chosen beneficial mutation is a decreasing function of both beneficial mutation rate, μ , and population size, N .	75
Figure 2.3. The substitution rate of a population is an increasing function of its beneficial mutation rate	76
Figure 2.4. The expected selection coefficient of substitutions, $\langle s(\alpha, \mu, N) \rangle$, is an increasing function of population size, N .	77
Figure 2.5. The rate of fitness improvement is an increasing function of both population size, N , and beneficial mutation rate, μ .	78

Figure
b
f
v

Figure
p

Figure
d

Figure
b

Figure
ch
ba
s

Figure
is
f
H
16

Figure
to
po
ov
x

Figure
po

Figure

Figure

Figure
tr
E
1

Figure 2.6. The probability that an arbitrarily chosen beneficial mutation transiently achieves polymorphic frequency is plotted against log population size for various beneficial mutation rates.	79
Figure 2.7. The leapfrog phenomenon illustrated phylogenetically.	80
Figure 2.8. The leapfrog phenomenon illustrated dynamically.	81
Figure 2.9. A simulation of competition among numerous beneficial mutations.	82
Figure 2.10. The probability of fixation of an arbitrarily chosen beneficial mutation is plotted against the beneficial mutation rate, μ , for various population sizes.	83
Figure 3.1. Probability that a superior mutation ($s_s = 0.1$) is fixed before an inferior mutation ($s_I = 0.08$) as a function of both population number, N (for which $\mu_s = \mu_I = 3 \times 10^{-9}$), and overall beneficial mutation rate, μ (for which $N = 3.3 \times 10^7$).	128
Figure 3.2. Probability of fixation ordering from largest to smallest selective advantage as a function of both population number, N (for which $\mu = 6 \times 10^{-9}$), and overall beneficial mutation rate, μ (for which $N = 3.3 \times 10^7$).	129
Figure 3.3. Evolutionary trajectories as a function of population size.	130
Figure 3.3 (cont.)	131
Figure 3.3 (cont.)	132
Figure 3.4. Least-squares fit of expected fitness trajectory to fitness data from evolving <i>E. coli</i> population, the Ara ⁻ 1 line from Lenski & Travisano (1994).	133

Figure 3.5. Estimated probability (solid line) and maximum probability (dotted line) that the inferior AZT-resistance mutation, K70R, is fixed before the superior mutation, T215Y/F, as a function of effective population number.	134
Figure 4.1. The dynamics of mutation-selection balance.	170
Figure 4.2. Panel A shows the per-capita hitchhiking rate of a specified deleterious mutation as a function of population number, N . Panel B shows the corresponding probability that a population fixes the deleterious mutation by hitchhiking within a time interval of 10,000 generations.	172
Figure 4.3. Panel A shows the per-capita hitchhiking rate of a specified deleterious mutation as a function of its selective disadvantage, s_d . Panel B shows the corresponding probability that a population fixes the deleterious mutation by hitchhiking within a time interval of 10,000 generations.	173
Figure 4.4. Panel A shows the per-capita hitchhiking rate of a mutator allele as a function of its strength, m . Panel B shows the corresponding probability that a population fixes the mutator allele by hitchhiking within a time interval of 10,000 generations. . .	174
Figure 4.5. The per-capita recruitment rate of double-mutations (beneficial mutations on mutator background) as a function of mutator strength, m	175
Figure 4.6. Probability of fixation of a double-mutation (beneficial mutation on mutator background) as a function of mutator strength, m	176
Figure 4.7. Frequency of a deleterious mutation during a substitutional event on both wildtype and beneficial backgrounds.	177

Chapter

$x(t)$

$y(t)$

$z(t)$

u

v

s

t

q

$u(s)$

λ

Prefix

σ

ω

LIST OF MATHEMATICAL SYMBOLS

Chapter 2:

$x(t)$	-	number of wildtype individuals at time t
$y(t)$	-	number of individuals carrying the beneficial mutation in question at time t
$z(t)$	-	number of individuals carrying an alternative beneficial mutation at time t
μ	-	beneficial mutation rate
N	-	total population number
s	-	selection coefficient
t_f	-	time to fixation
α	-	exponential parameter for distribution of beneficial mutational effects
$\pi(s)$	-	probability that a beneficial mutation of selective advantage s survives drift (referred to as the "survival function")
λ	-	expected number of "interfering mutations" as defined in the text
$Pr\{fix\}$	-	probability of fixation
σ	-	rate of adaptive substitution
$\langle . \rangle$	-	expected value

6

5

4

3

2

Chapter

R

The cas

I

S

Pr(I/S)

Pr(S/I)

2

- f - frequency
- R - number of "candidate replications" as defined in the text
- ψ - expected number of superior mutations that would prevent a given beneficial mutation from attaining frequency f .
- t_z - time at which the appearance and survival of a superior mutation of selective advantage s_z ensures that the beneficial mutation in question attains a frequency of exactly f .
- γ - expected number of superior mutations preventing the fixation of a beneficial mutation that has achieved a frequency of at least f .

Chapter 3:

- K - the linear constant in the survival function, such that $\pi(s) \approx Ks$.

The case of only two available beneficial mutations:

- I - denotes inferior beneficial mutation
- S - denotes superior beneficial mutation
- $Pr\{I;S\}$ - probability that the inferior mutation is fixed before the superior mutation
- $Pr\{S;I\}$ - probability that the superior mutation is fixed before the inferior mutation
- r_I - rate of appearance and survival of the inferior mutation

4

The ca

M

Pr 172

2

3/3/

$\Phi(\zeta)$

2

EIT

W

WIT

EIT

WIT

2

r_s - rate of appearance and survival of the superior mutation

The case of many available beneficial mutations:

M - number of available beneficial mutations

$Pr\{1;2,3,\dots,M\}$ - probability that the largest-effect mutation is fixed before all others

r_j - rate of appearance of the beneficial mutation with the j^{th} largest selective advantage

$\zeta\{j\}$ - permuting function mapping place in the fixation ordering onto rank by selective advantage

$\phi(\zeta)$ - probability of the fixation ordering specified by the permuting function ζ

τ_j - time between the $j-1^{\text{th}}$ fixation and the appearance of the j^{th} beneficial mutation to be fixed

$E(T_i|\zeta)$ - expected time until the i^{th} fixation, given fixation ordering ζ

w_i - population fitness after the i^{th} fixation, given fixation ordering ζ

$w(t|\zeta)$ - population fitness at time t , given fixation ordering ζ

$E(w|t)$ - expected population fitness at time t

$\text{Var}(w|t)$ - variance of population fitness among replicate populations at time t

I_j - the set of all available beneficial mutations both inferior to and subsequent to the j^{th} beneficial mutation fixed

S_1

H_1

S_2

Chapter

ψ

$\dot{\psi}$

θ

\ddot{u}_B

\ddot{u}_B'

\ddot{u}_D

S_2

Prefix

h

$x(t)$

$y(t)$

- S_j - the set of all available beneficial mutations both superior to and subsequent to the j^{th} beneficial mutation fixed
- $\#I_j$ - cardinality of set I_j
- $\#S_j$ - cardinality of set S_j

Chapter 4:

- ϕ - mutation-selection balance
- $\hat{\phi}$ - equilibrium mutation-selection balance
- ϕ_e - "effective mutation-selection balance" as defined in the text
- μ_B - overall beneficial mutation rate on the wildtype background
- μ'_B - overall beneficial mutation rate on the specified deleterious background
- μ_D - rate of mutation from wildtype to the specified deleterious mutation
- s_D - selective disadvantage of the specified deleterious mutation
- $Pr\{fix\}$ - probability of fixation
- h - per-capita hitchhiking rate of the specified deleterious mutation
- $x(t)$ - number of wildtype individuals at time t
- $y(t)$ - number of individuals carrying only the specified deleterious mutation at time t

- $z(t)$ - number of individuals carrying both
 beneficial and deleterious mutations at time t

- m - factor by which a specified mutator allele
 elevates the general mutation rate, e.g.
 $\mu'_B = m\mu_B$.

- μ_δ - overall deleterious mutation rate of the
 wildtype

Fundame

Ad

genetic

different

select

independ

variant

displa

at a s

other

value

genetic

Wayne

speci

fixati

linked

INTRODUCTION

Fundamental issues and roadmap to the dissertation

Adaptive evolution occurs by the action of selection on genetic variants in a population. In sexual populations, different fitness variants may recombine, thereby allowing selection to act at different loci with a certain degree of independence. In asexual populations, however, fitness variants compete and the fittest variant eventually displaces all others. Thus the fate of a particular allele at a specified locus is determined as much by selection at other loci to which it is linked as by its own selective value. One consequence of asexuality (or, more generally, genetic linkage) is a phenomenon known as hitchhiking (Maynard Smith & Haigh, 1974; Berg, 1995), whereby a specified mutant allele is driven to high frequency or even fixation by a successful beneficial mutation to which it is linked. A second consequence of asexuality is the

phenomena

progress

hinder

benefit

7

hitchhike

of use

mutate

hitchhike

2 and

under

benefit

in case

Chapter

Chapter

Chapter

Chapter

Chapter

Chapter

Chapter

Chapter

Chapter

Chapter

phenomenon called "clonal interference", whereby the progression of one beneficial mutation to fixation is hindered or even prevented by competition with alternative beneficial mutations.

This dissertation explores the effects of both hitchhiking and clonal interference on the adaptive dynamics of asexual populations. Chapter 1 is an empirical study of mutation rate evolution in bacterial populations and invokes hitchhiking as an explanation for observed trends. Chapters 2 and 3 are theoretical explorations of clonal interference under different assumptions about the availability of beneficial mutations. Chapter 4 ties the theory developed in chapters 2 and 3 to the empirical observations of Chapter 1.

Chapter 1

Mutations may arise in the genes responsible for accurate DNA synthesis, proofreading or repair. An individual is said to carry a mutator allele if any of these genes is impaired or disabled by such a mutation; a mutator allele therefore confers an elevated mutation rate. An

individual

function

thus, a

rate.

recurre

the mu

select

laborat

To

mutati

with t

as a p

should

availa

strong

availa

delete

conver

1970;

predic

consta

individual is said to carry an antimutator allele if the function of any of these genes is enhanced by a mutation; thus, an antimutator allele confers a diminished mutation rate. Because mutator and antimutator alleles are recurrently produced in a population, their fate and hence the mutation rate of the population may be determined by selection. Chapter 1 examines mutation rate evolution in laboratory populations of *E. coli* (Sniegowski et al., 1997).

This study began as a test of the hypothesis that mutation rates in a constant environment should decrease with time. This hypothesis is grounded in the logic that, as a population adapts to a constant environment, there should be (i) weaker selection for mutator alleles as the available beneficial mutations are expended and (ii) stronger selection for antimutator alleles as the remaining available mutations become almost exclusively neutral or deleterious. This line of reasoning is in accordance with conventional theory (Kimura, 1960; Kimura, 1967; Leigh, 1970; Leigh, 1973; Gillespie, 1981; Ishii et al., 1989) and predicts that evolution should reduce mutation rates in a constant environment.

To

perform

estimated

that has

entirely

1994 .

is not

complex

Sarkar,

1992; S

analyt

estima

as the

a comp

this p

are g

H

E. col

system

Surpr

rates

the an

To test this hypothesis, Sniegowski et al. (1997) performed fluctuation assays (Luria & Delbruck, 1943) to estimate the mutation rates of twelve *E. coli* populations that had been evolving independently in a constant environment for 10,000 generations (Lenski & Travisano, 1994). Estimation of mutation rates from fluctuation assays is notoriously difficult because of its mathematical complexity (Lea & Coulson, 1949; Stewart et al., 1990; Sarkar, 1991). Building on recent developments (Ma et al., 1992; Stewart, 1994; Jaegger & Sarkar, 1995), I derive the analytical Newton series for the maximum likelihood estimator of mutation rates from fluctuation assays, as well as the analytical variance, and have incorporated these into a computer program. Instructions for obtaining and using this program, as well as the novel mathematical derivations, are given in Appendix 1.1.

Having estimated mutation rates for each of the twelve *E. coli* populations, none of the twelve populations had a systematically lower mutation rate than the ancestor. Surprisingly, three of the twelve populations had mutation rates that were about two orders of magnitude *higher* than the ancestral rate. This result was cause for rejection of

the

conf

exam

conf

assu

this

evol

thei

insto

popul

where

mutat

provi

for a

and s

result

popula

studies

the original hypothesis and for critical review of conventional mutation rate evolution theory. In re-examining this theory, Paul Sniegowski (personal communication) proposed that a major flaw was the implicit assumption of infinite population size. A consequence of this flaw is the unsubstantiated prediction that adaptive evolution should "fine-tune" mutation rates. To explain their observations, Sniegowski et al. (1997) hypothesized instead that either (i) mutation rates in asexual populations may be determined by chance hitchhiking events, whereby mutator alleles are driven to fixation by beneficial mutations that they produce, or (ii) mutator alleles may provide a direct fitness benefit if reduced fidelity allows for an increased replication rate. Chapters 1 and 4 explore and support the plausibility of (i). While preliminary results suggest rejection of (ii) for the *E. coli* populations, the validity of this rests on future empirical studies.

Chapter

Ch

adaptive

among b

asexual

benefit

interf

to fix

& Kano

called

phenom

popula

commu

have a

and on

benef

even

compe

Such

prese

is s

Chapter 2

Chapter 2 is a theoretical exploration of the asexual adaptive dynamics that result from competitive interactions among beneficial mutations (Gerrish & Lenski, 1998). In asexual populations, clones that carry alternative beneficial mutations compete with one another and, thereby, interfere with the expected progression of a given mutation to fixation (Fisher, 1930; Muller, 1932; Muller, 1964; Crow & Kimura, 1965; Haigh, 1978). This phenomenon has been called "clonal interference". Intuitively, it seemed this phenomenon would significantly affect fundamental population-genetic parameters. Richard Lenski (personal communication) speculated that this phenomenon might also have a curious effect on phylogenetic and fitness dynamics and on the relationship between these. He reasoned that a beneficial mutation may rise to polymorphic frequency, or even become the majority genotype, before being competitively displaced by a superior, alternative mutation. Such a phenomenon, which he termed the "leapfrog", would present strange phylogenetic dynamics in which ancestor *ab* is succeeded by mutant *Ab*, and *Ab* is then succeeded not by

AB but

related

the first

the app

mutated

under a

occur a

reasons

En

analytical

parameters

and ra

probab

polymers

Employ

solvent

mutated

of mut

AB but by *aB* (the two successive mutants being more closely related to the ancestor than they are to each other). From the fitness standpoint, the leapfrog phenomenon would give the appearance of two fixations when, in fact, only one mutation is actually fixed. The question remained, however, under what conditions the leapfrog phenomenon was likely to occur and whether such conditions were biologically reasonable.

From theoretical exploration of clonal interference, I analytically derive several fundamental population-genetic parameters such as fixation probability, substitution rate and rate of fitness increase. Also, I derive the probability that a beneficial mutation transiently achieves polymorphic frequency or majority status (the "leapfrog"). Employing fitness data from an evolving *E. coli* population, I solve the inverse problem to estimate (i) the beneficial mutation rate in this population and (ii) the distribution of mutational effects.

chapte

A

popula

envir

are n

benef

envir

mutat

be co

dynam

i.e.,

fini

the

fini

popu

comp

such

spec

equa

reg

Chapter 3

An assumption made throughout Chapter 2 is that the population in question is evolving in a slowly changing environment, such that opportunities for further improvement are never depleted. In other words, the number of available beneficial mutations is infinite. In a constant environment, however, the set of available beneficial mutations is eventually depleted. This set must therefore be considered finite. Chapter 3 explores the adaptive dynamics of asexual populations in a constant environment, i.e., when the number of available beneficial mutations is finite.

Given a finite set of available beneficial mutations, the course of adaptive evolution is determined by both the timing and ordering of fixations. In large, asexual populations, the ordering of fixations is affected by competition among beneficial mutations. Taking into account such clonal interference, I derive the probability of any specified fixation ordering. From this, I derive dynamic equations for fitness expectation and variance among replicate populations, given (i) a particular set of

benefit

distrib

an ev

to de

popul

effec

muta

Chap

mode

exp

ai,

fix

Ben

de

rat

ste

ben

ste

de

beneficial mutations, or given (ii) only a common distribution of mutational effects. Using fitness data from an evolving *E. coli* population, I solve the inverse problem to determine (i) the beneficial mutation rate in this population, (ii) the distribution of beneficial mutational effects, and (iii) the number of available beneficial mutations.

Chapter 4

Chapter 4 employs the theory developed in Chapter 2 to model the hitchhiking mechanism that was proposed as an explanation for the observations of Chapter 1 (Sniegowski et al., 1997). In order for any allele to hitchhike to fixation, it must first be linked to a beneficial mutation. Beneficial mutations appear in linkage with a specified deleterious mutation at a certain rate. As long as this rate is greater than zero, the subpopulation carrying the specified deleterious mutation will eventually produce a beneficial mutation. If this beneficial mutation is subsequently fixed in the population, then the specified deleterious mutation to which it is linked will hitchhike to

fixation. The likelihood that the beneficial mutation will be fixed is reduced, however, by (i) the selective disadvantage of the deleterious mutation to which it is linked, (ii) genetic drift, and (iii) clonal interference from both wildtype and mutant subpopulations. Taking all of this into account, I derive the probability that a particular deleterious allele will hitchhike to fixation within a given time period. The probability that a *mutator* allele hitchhikes to fixation is then presented as a special case. Results corroborate the observations of Chapter 1 and suggest that hitchhiking is of general importance to the evolution of mutation rates in asexual populations.

Abstr

spont

1991)

rates

1989,

limit

dell

Gabs

effe

(Mac

spont

long

comp

(Ta

ass

199

1

199

Get

Chapter 1

EVOLUTION OF HIGH MUTATION RATES IN EXPERIMENTAL POPULATIONS OF *ESCHERICHIA COLI*¹

Abstract

Most mutations are likely to be deleterious, and hence the spontaneous mutation rate is generally held at a very low value (Drake, 1991). Nonetheless, evolutionary theory predicts that high mutation rates can evolve under certain circumstances (Leigh, 1970; Ishii et al., 1989; Taddei et al., 1997). Empirical observations have heretofore been limited to short-term studies of the fates of mutator strains deliberately introduced into laboratory populations of E. coli (Cox & Gibson, 1974; Chao & Cox, 1983; Trobner & Piechocki, 1984) and to the effects of intense selective events on mutator frequencies in E. coli (Mao et al., 1997). Here we report a new phenomenon: the rise of spontaneously originated mutators in populations of E. coli undergoing long-term adaptation to a novel environment. Our results corroborate computer simulations of mutator evolution in adapting clonal populations (Taddei et al., 1997) and may shed light on recent observations that associate high mutation rates with emerging pathogens (LeClerc et al., 1996) and with certain cancers (Modrich et al., 1995).

1

This chapter was written originally as a paper (Sniegowski et al., 1997). The "we" in this chapter refers to P.D. Sniegowski, P.J. Gerrish, and R.E. Lenski.

Introduct

lens

establish

dional st

directly

founded f

their onl

environme

essential

provided

Substanti

certain c

been docu

Travisand

1996; El

We r

strain a

had been

populati

three po

displaye

Introduction and Findings

Lenski and collaborators (Lenski et al., 1991) established twelve replicate experimental populations of a clonal strain of *E. coli* in order to study evolution directly in the laboratory. Because these populations were founded from a single ancestral clone, mutation provided their only source of genetic variation. The glucose-limited environment in which these populations were propagated was essentially novel at the outset of the experiment and provided considerable scope for adaptive evolution. Substantial evolutionary increases in fitness and changes in certain other phenotypic features in these populations have been documented elsewhere (Lenski et al., 1991; Lenski & Travisano, 1994; Vasi et al., 1994; Travisano & Lenski, 1996; Elena et al., 1996).

We measured mutation rates in the common ancestral strain and in the twelve experimental populations after they had been evolving for 10,000 generations. A majority of the populations retained the ancestral mutation rate; however, three populations (designated Ara⁻2, Ara⁻4 and Ara⁺3) displayed mutation rates that were between one and two

orders of

Figure 1

phenotype

Ara-4 and

phenotype

experiment

all subs

mutators

indicate

frequency

To

phenotype

from pop

the ance

wild-type

ancestra

Ara-4 and

alleles

pathway

the ance

by the p

In the A

orders of magnitude higher than those in the ancestor (Figure 1.1). Figure 1.2 illustrates when mutator phenotypes arose during the evolution of populations Ara⁻², Ara⁻⁴ and Ara⁺³, as determined by screening for the mutator phenotype in isolates stored periodically throughout the experiment. Once a lineage displayed a mutator phenotype, all subsequent isolates from that lineage also tested as mutators through 10,000 generations. This observation indicates that mutators rose to and remained at high frequencies in these populations.

To ascertain a genetic basis for the observed mutator phenotypes, we transformed 10,000-generation clonal isolates from populations Ara⁻², Ara⁻⁴ and Ara⁺³ and an isolate from the ancestral strain REL606 with multicopy plasmids bearing wild-type alleles of seven known general mutator loci. The ancestral mutation rate was fully restored in the Ara⁻², Ara⁻⁴ and Ara⁺³ strains only by the presence of wild-type alleles of genes in the methyl-directed mismatch repair pathway (Modrich, 1991) (Figure 1.3). The mutation rate in the ancestral strain, REL606, was not significantly affected by the plasmids (with a single exception; see Figure 1.3). In the Ara⁺³ strain the *mutS*⁺ allele alone restored the

ancestra

ancestra

data als

recent s

between.

in one m

other in

Methods

Experiments

The

of E. C.

Wederb

from RE

REL606

populat

diustio

1 Brown

Populat

cells a

ancestral mutation rate. In the Ara⁻2 and Ara⁻4 strains the ancestral rate was completely restored by *uvrD*⁺, and the data also indicated a partial effect of *mutL*⁺. In fact, recent studies have suggested a mechanistic interaction between the *mutL* and *uvrD* gene products, such that a defect in one may be complemented by increased production of the other in some cases (P. Modrich, personal communication).

Methods

Experimental System

The common ancestral strain was REL606, an Ara⁻ clone of *E. coli* B obtained by B. R. Levin from S. Lederberg (Lederberg, 1966). A spontaneous Ara⁺ revertant was selected from REL606 and designated REL607. Six clones each from REL606 and REL607 were used to found the twelve experimental populations, which were propagated at 37°C by daily 100-fold dilutions into 10 ml of fresh Davis minimal medium (Carlton & Brown, 1981) supplemented with glucose at 25 µg/ml. Population sizes fluctuated between approximately 5 X 10⁶ cells after dilution and 5 X 10⁸ cells at stationary phase.

isolates

generati

the expa

Mutation

Fig

a given

all str

testing,

the orig

physiol

populat

strain v

for each

measure

aliquot

culture

plating

discern

final p

random

Isolates from each population were stored at -80°C at 100-generation intervals during the first 2500 generations of the experiment and at 500-generation intervals thereafter.

Mutation rate measurements

Fluctuation tests (Luria & Delbruck, 1943) to estimate a given rate of mutation were conducted simultaneously on all strains to provide controlled comparison. Prior to testing, strains were revived and regrown for three days in the original experimental medium to reestablish the physiological conditions that had prevailed in the evolving populations. In general, 24 or more cultures of a given strain were grown from inocula of approximately 1,000 cells for each fluctuation test. For a given mutation rate measurement, all strains to be compared were grown in aliquots from the same batch of Davis minimal medium. Cultures were grown to stationary phase before selective plating. (An additional 24 hours in stationary phase had no discernable effect upon the mutation rate in pilot studies.) Final population sizes were estimated by growing and randomly harvesting three extra cultures for each strain and

measuring

the mean

mutants

suppleme

enumerat

µg/ml of

plating

time of

M.

generat

experim

medium

assayed

Expecte

ancestr

which

>95% C

measuring cell densities using a Coulter particle counter; the means were used in mutation rate calculations. Ara⁺ mutants were enumerated on Davis minimal agar medium supplemented with arabinose at 10 µg/ml; Nal^r mutants were enumerated on LB agar medium (Miller, 1992) containing 20 µg/ml of nalidixic acid; T5^r mutants were enumerated by plating cultures with excess phage T5 in soft agar.

Time of Origin and Persistence of Mutator Phenotypes

Mutator screens were conducted at approximately 500 generation intervals over the 10,000 generations of the experiment. Overnight cultures in 5 ml of Davis minimal medium (supplemented with glucose at 1,000 µg/ml) were assayed for Nal^r mutant numbers at stationary phase. Expected distributions of mutants, based on mutator and ancestral mutation rates, were used to generate criteria by which isolates could be scored as mutator or nonmutator with >95% confidence.

complem

St

wildtyp

mutS p

dnaE (p

al., 1

source

Scholz

Bessma

from S

condu

Propa

where

cultu

Anal

Dele

to c

fin

Complementation Tests

Strains were transformed with plasmids carrying wildtype alleles for *mutH* (plasmid pGW1899), *mutL* (pGW1842), *mutS* (pGW1811), *uvrD* (pGT26), *mutT* (pSK25), *dnaQ* (pMM5) and *dnaE* (pMK9) according to a standard protocol (Sambrook et al., 1989). The plasmid labels correspond to the following sources: pGW from Pang et al. (1985), pGT from Taucher-Scholz & Hoffman-Berling (1983), pSK from Bhatnagar & Bessman (1988), pMM from Horiuchi et al. (1981), and pMK from Schaaper & Cornacchio (1992). Fluctuation tests were conducted as described above, except that all strains were propagated in LB medium (containing 60 µg/ml of ampicillin where the strain was plasmid-bearing) and five parallel cultures were used per fluctuation test.

Analysis of Fluctuation Test Data

I wrote a computer program employing a recent Luria-Delbrück distribution-generating algorithm (Ma et al., 1992) to calculate maximum likelihood mutation rates from fluctuation test data. Instructions for obtaining and using

this prop

confidenc

Figure 1

(1994) . 7

rates 11

theoretic

in \mathbb{M} , as

mutation

Discuss

B.

are 11

an. evo

mutatio

assoc

select

select

mutat

in ch

these

1992

this program are given in Appendix 1.1. Approximate 95% confidence intervals for the mutation rates illustrated in Figure 1.1 were calculated using the formulae of Stewart (1994). Approximate 95% confidence limits for the mutation rates illustrated in Figure 1.3 are based upon the theoretical variance of the maximum likelihood estimate of $\ln m$, assuming normality, where m is the expected number of mutations per culture (see Appendix 1.1).

Discussion

Because most mutations are deleterious, mutator alleles are likely to have negative average effects on fitness. In an evolving clonal population, however, a deleterious mutator can rise to high frequency (hitchhike) in association with an adaptive mutation provided that the selective cost of the mutator does not outweigh the selective benefit of the adaptive mutation. Hitchhiking of mutators with adaptive mutations was demonstrated previously in chemostat populations of *E. coli* (Chao & Cox, 1983), but these results left open the question of whether such events would be likely to occur in natural populations: When a

mutator

number

always

populat

However,

realis

a bene

I

in nat

mutati

high

Compu

analy

unpub

occas

popu

adap

as t

may

com

the

and

mutator was introduced above a relatively high threshold number relative to the wild type, the mutator population always acquired an adaptive mutation before the wildtype population did and the mutator rose to high frequency. However, when the mutator was introduced in lower and more realistic numbers, the wild type population always acquired a beneficial mutation first and displaced the mutator.

In the evolution experiment we have described here, as in natural populations, mutator alleles must have arisen by mutation. This raises the question of how mutators reached high frequencies in three of our twelve populations. Computer simulations (Taddei et al., 1997) and stochastic analytical modeling (Gerrish, Sniegowski and Mandrekar, unpublished; see Chapter 4) suggest that rare mutators may occasionally hitchhike to high frequencies in finite asexual populations as a consequence of chance associations with adaptive mutations. We view such chance hitchhiking events as the likely explanation for our results. Sufficient time may not have existed to observe them in earlier studies of competition between mutator and wild-type strains (Cox & Gibson, 1974; Chao & Cox, 1983; Trobner & Piechocki, 1984), which were carried out for only a few tens or hundreds of

generat

study.

that ce

fitness

mutator

neutral

WHD de

conditi

A

al. (1

all, f

experim

predic

higher

retain

effect

mutato

show b

over c

appro

1997,

mutabi

generations, as opposed to the 10,000 generations in our study. Although we cannot formally rule out the possibility that certain mutations in mismatch repair might increase fitness directly, it is much more likely that the evolved mutators we observed are deleterious or at best effectively neutral. Indeed, known mutator alleles of *mutS*, *mutL* and *uvrD* do not increase cell fitness under our experimental conditions (C. Zeyl, unpublished data).

A prediction of the simulations conducted by Taddei et al. (1997) is that mutators will hitchhike in some, but not all, finite asexual populations undergoing adaptation. Our experimental observations are consistent with this prediction. A further prediction is that fitness will be higher on average in mutator populations than in those that retain the wild type mutation rate. However, this fitness effect is very small and subtle: approximately half of the mutator populations in the Taddei et al. simulations did not show higher fitness, and the observed increases in fitness over contemporaneous wild type populations were slight (approximately 1%) in the remainder (see Taddei et al., 1997, their Figure 3B). We tested for a relationship between mutability and fitness in our experimental populations in

se

For

re

re

av

re

re

av

re

re

re

re

re

re

re

re

re

re

re

re

re

re

several ways, none of which gave statistical significance. For example, we tested for a correlation between mutability ranks (as calculated from data obtained in screens for mutator phenotypes: see Methods) and relative fitness ranks averaged across the duration of the experiment in all twelve populations. The results were inconclusive ($r = 0.28$, $n = 12$, one-tailed $P = 0.19$). It is clear from our experiments and the simulations of Taddei et al. that increased mutation rate and not increased fitness is the more striking consequence of the hitchhiking process.

Our finding that asexual populations can evolve high mutation rates in a relatively benign environment may shed light on recent observations associating mismatch repair mutators with certain cancers (Modrich, 1995) and with pathogenicity in *E. coli* and *Salmonella* (LeClerc et al., 1996). As in our experimental populations, high mutation rates may evolve as a stochastic byproduct of adaptation in clonal tumour lineages and in populations of asexual pathogens (Nowell, 1974; Moxon et al., 1994). The potential for faster evolution once high mutation rates have evolved may have important health implications.

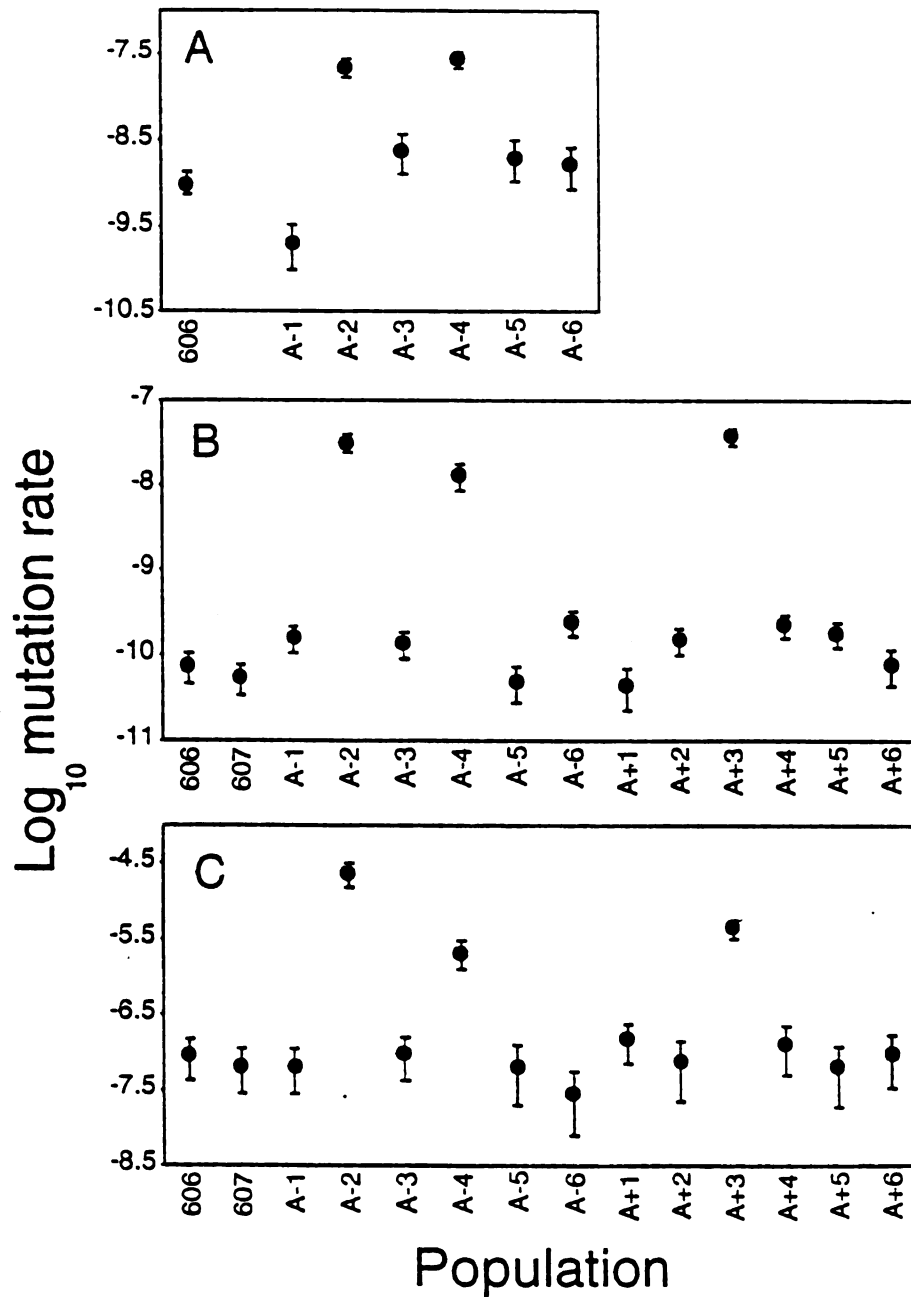


Figure 1.1. Rates of mutation in 12 Ara⁻ and Ara⁺ experimental populations (A-1 to A-6 and A+1 to A+6) at 10000 generations and in their common ancestors REL606 (Ara⁻) and REL607 (Ara⁺). Error bars give approximate 95% confidence intervals. (A) Reversion to Ara⁺ in isolates from the seven Ara⁻ populations. (B) Mutation to nalidixic acid resistance. (C) Mutation to bacteriophage T5 resistance.

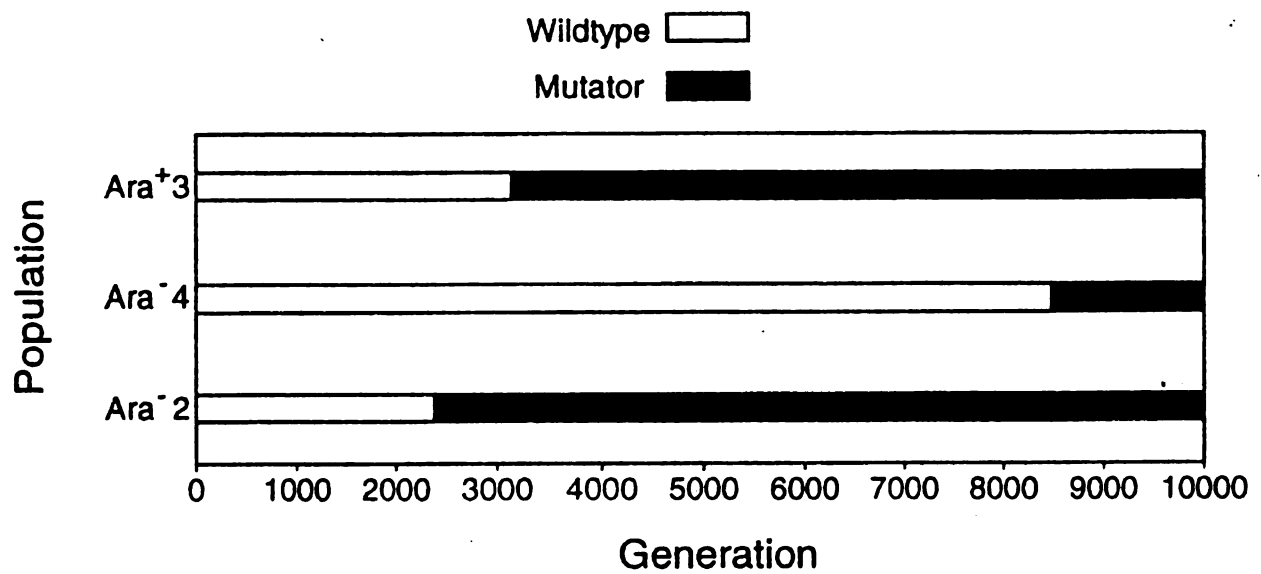


Figure 1.2. Time of appearance and evolutionary persistence of mutator phenotypes in experimental populations Ara⁻2, Ara⁻4, and Ara⁺3.

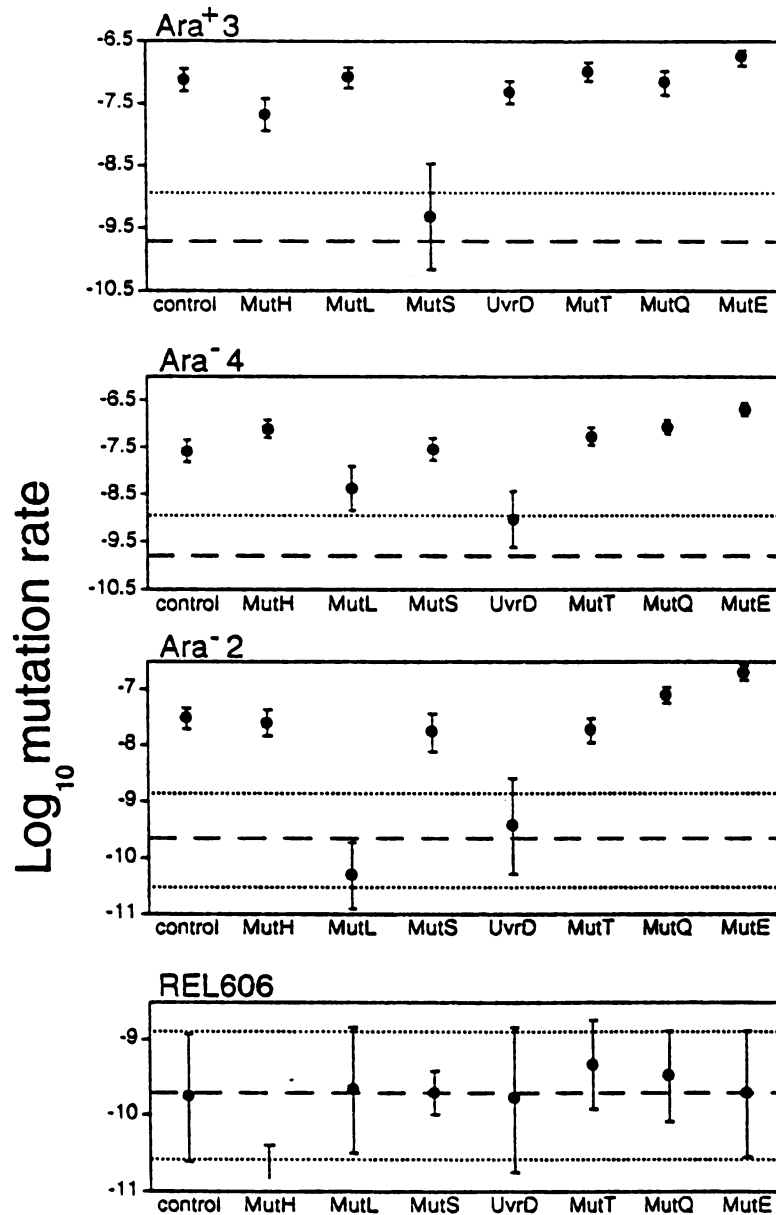


Figure 1.3. Mutation rates to nalidixic acid resistance in 10,000-generation isolates from populations Ara⁻2, Ara⁻4 and Ara⁺3 and ancestor REL606 transformed with plasmids bearing wild-type alleles of seven known general mutator loci. Controls shown are plasmid-free. Error bars give approximate 95% confidence limits. Only an upper confidence limit is shown for the case in which no mutants were obtained in any culture. For visual comparison, the dashed horizontal line and dotted horizontal lines in each panel illustrate the mutation rate and approximate 95% confidence limits measured in the ancestral control.

1
1
e
e
1
s
s
P
f
t
a
m
p

Chapter 2

THE FATE OF COMPETING BENEFICIAL MUTATIONS IN AN ASEXUAL POPULATION¹

Abstract

*In sexual populations, beneficial mutations that occur in different lineages may be recombined into a single lineage. In asexual populations, however, clones that carry such alternative beneficial mutations compete with one another and, thereby, interfere with the expected progression of a given mutation to fixation. From theoretical exploration of such "clonal interference", we have derived (1) a fixation probability for beneficial mutations, (2) an expected substitution rate, (3) an expected coefficient of selection for realized substitutions, (4) an expected rate of fitness increase, (5) the probability that a beneficial mutation transiently achieves polymorphic frequency ($\geq 1\%$), and (6) the probability that a beneficial mutation transiently achieves majority status. Based on (2) and (3), we were able to estimate the beneficial mutation rate and the distribution of mutational effects from changes in mean fitness in an evolving *E. coli* population.*

¹

This chapter was written originally as a paper (Gerrish & Lenski, 1998). The "we" in this chapter refers to P.J. Gerrish and R.E. Lenski.

Intro

occu

mutat

incor

(Fish

The t

be co

mutat

(Lens

frequ

ances

occu

can r

negat

howev

Such

sprea

"clon

impor

asex.

Introduction

Asexual populations adapt to their environment by the occurrence and subsequent rise in frequency of beneficial mutations. Without recombination, a population must incorporate beneficial mutations in a sequential manner (Fisher, 1930; Muller, 1932, 1964; Crow & Kimura, 1965). The time required for fixation of a beneficial mutation may be considerable if the population is large; however, the mutation remains at low frequency for much of this time (Lenski et al., 1991). While the mutation is at low frequency, another beneficial mutation may arise on the ancestral background. If two such beneficial mutations occur in a sexual population, then the two novel genotypes can recombine to form a fitter double-mutant (assuming no negative gene interactions). In an asexual population, however, these two novel genotypes compete with one another. Such competition between beneficial mutations slows the spread of, and may even eliminate, the first mutation. Such "clonal interference" between beneficial mutations has many important consequences for the dynamics of evolution in asexual populations.

The idea that progression of a beneficial mutation to fixation may be impeded by competing beneficial mutations was articulated by Muller (1932, 1964) in the context of discussions on the evolutionary advantage of sex. Almost in passing, a brief theoretical treatment was later given by Haigh (1978), in which he proposed a discrete-time model of competing beneficial mutations. Employing a different approach, we give a full theoretical treatment of the phenomenon of competing beneficial mutations and its consequences.

The body of this paper is presented in two main parts. In the first part, a probability of fixation is derived which incorporates the effect of competition between beneficial mutations, and some consequences of this derivation are then explored. The dynamics of fixation are such that a relatively simple derivation suffices. In the second part, the probability is derived that a beneficial mutation achieves a frequency greater than or equal to some specified frequency, f . From this, the probability that a beneficial mutation becomes transiently polymorphic ($0.01 < f < 1$) or transiently common ($0.5 < f < 1$) is derived. The derivations in the second part require treatment of the

3.

a.

C.

C.

S.

O.

N.

C.

X.

g.

t.

g.

d.

a.

z.

z.

z.

z.

dynamics of a three-genotype system; hence, the derivations are more complex than those of the first part.

Clonal interference and fixation

Clonal interference among beneficial mutations

Some definitions. We refer to the common progenitor of one or more mutants as the *ancestor*; the ancestral genotype, which is haploid, is denoted by ab , and the number of carriers of the ab genotype present at time t is denoted by $x(t)$. A mutant that carries a *beneficial mutation* has genotype Ab and has $y(t)$ carriers at time t . Another mutant that carries an *alternative mutation*, also beneficial, has genotype aB and has $z(t)$ carriers at time t . When discussing a beneficial mutation that is followed by the appearance of one or more alternative mutations, the first beneficial mutation shall often be described in retrospect as the *original mutation*. If the original mutation is followed by a superior mutation, then there is a significant probability that the original mutation will be eliminated.

the

the

all

the

he

su

th

th

di

in

the

1

a

a

g

g

7

7

a

5

2

This phenomenon, whereby the fate of an original beneficial mutation is altered by the appearance of a superior alternative mutation, shall be called *clonal interference*.

The expected number of interfering mutations. We derive here the expected number of alternative mutations that are superior to a beneficial mutation and hence interfere with the progression of that mutation to fixation. Assuming that the number of such *interfering mutations* is Poisson distributed, we determine the probability that no interfering mutation occurs by calculating the zero-class. This probability is an important factor in determining the likelihood of fixation of a beneficial mutation.

Suppose a population is homogeneous until the appearance of a beneficial mutation, at which time the population consists of two fitness variants: the ancestral genotype and the beneficial mutant. Let the beneficial mutant appear in the ancestral population at time $t = 0$. The beneficial mutant, being competitively superior to the ancestor, slowly displaces the latter until finally reaching fixation at some time, t_f . Therefore, the time interval during which the beneficial mutation is present but not yet

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

fixed is $(0, t_f)$. The expected number of interfering mutations is simply the expected number of mutations occurring in the interval $(0, t_f)$ that are competitively superior to the original beneficial mutation.

If total population size, N , is constant, then the dynamics of the two genotypes are logistic (see, for example, Crow & Kimura, 1970). Let s denote the difference in Malthusian parameters between ancestor ab and beneficial mutant Ab . We have chosen the letter s because, under logistic growth, a difference in Malthusian parameters is equivalent to a selection coefficient when the unit of time is generations. Let μ denote the beneficial mutation rate per capita, per generation. Define t_f as the time to virtual fixation of the mutant subpopulation, i.e., $y(t_f) = N-1$. The expected number of further beneficial mutations between the time of appearance of the original mutation and its fixation is

$$\mu \int_0^{t_f} x(t) dt = \frac{\mu}{s} N \ln N . \quad (1)$$

Of these beneficial mutations produced, we now calculate the fraction that interfere with the growth of the original mutation. We assume the effects of beneficial mutations to be exponentially distributed (c.f. Kimura, 1970; see *Discussion*). Hence, the probability density for s is $\alpha e^{-\alpha s}$, where α characterizes the distribution of mutational effects and may be determined from empirical data (see *Estimation of parameters*).

In the first few generations of growth, a beneficial mutation may be lost due to random sampling events, or drift. We employ the general notation, $\pi(s)$, to denote the probability that a beneficial mutation is not lost by drift in these first few generations. (While dependence on s alone is indicated by our notation, this probability may in some cases be a function of other parameters as well.) All further derivations employ the general notation, $\pi(s)$, whereas all computations implement the approximation, $\pi(s) \approx 4s$, which is derived in Appendix 2.1 for the special case of bacterial populations. We emphasize that our analytical results are general for asexual species; implementing these results, however, depends on first

2

5

as

W

1

di

W

P

(1

S

as

2

3

7

as

7

1

finding an appropriate function, $\pi(s)$, for the particular species under study. A viral species, for example, might be assumed to have a Poisson distribution of offspring, in which case $\pi(s) \approx 2s$ when population size is constant (Haldane, 1927). Expressions for probabilities of surviving drift in fluctuating populations are given in Otto and Whitlock (1997).

In light of the two preceding paragraphs, the probability that an arbitrarily chosen beneficial mutation (i) has a selection coefficient greater than s and (ii)

survives drift is $\int_s^\infty \pi(u) \alpha e^{-\alpha u} du$. All further derivations

assume that $\pi(u)$ is linearly related to u , in which case this integral is equivalent to $e^{-\alpha s} \pi(s + \frac{1}{\alpha})$.

Because loss by drift occurs in the first few generations, whereas loss by clonal interference is more probable in later generations, we can make the simplifying assumption that these two processes are independent. Therefore, the expected number of mutations occurring in the interval $(0, t_f)$ that are superior to a given beneficial

mutation with selective advantage, s , and that survive the effects of drift is

$$\lambda(s, \alpha, \mu, N) = \frac{\mu}{s} N \ln(N) e^{-\alpha s} \pi(s + \frac{1}{\alpha}) . \quad (2)$$

This is the expected number of interfering mutations.

Fixation probability of a beneficial mutation

If a beneficial mutation survives the first few generations and is not lost by drift, its fixation is still far from ensured. In fact, fixation of a beneficial mutation may be very unlikely as a consequence of the presence of, and competition with, alternative mutations.

A beneficial mutation is fixed only if (i) it survives drift, and (ii) no superior mutation appears and survives drift in the time interval required for fixation. Given selective advantage s , the probability that a beneficial mutation will not be lost by either drift or clonal interference is simply the product,

$$Pr\{fix|s, \alpha, \mu, N\} = \pi(s) e^{-\lambda(s, \alpha, \mu, N)} . \quad (3)$$

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

This probability is plotted against s in Figure 2.1.

Finally, the probability density for the condition that a given beneficial mutation confers a selective advantage s is $\alpha e^{-\alpha s}$. The probability, therefore, that an arbitrarily chosen beneficial mutation will become fixed in a population is

$$Pr\{fix|\alpha, \mu, N\} = \alpha \int_0^{\infty} \pi(s) e^{-\lambda(s, \alpha, \mu, N) - \alpha s} ds . \quad (4)$$

In Figure 2.2, the fixation probability given by equation (4) is plotted for different combinations of α and N . Note the dramatic decrease in fixation probability with increasing beneficial mutation rate and with increasing N . The above calculations will become more informative and useful when the fixation probability is converted to an expected substitution rate of beneficial mutations in a population.

The expected rate of substitution

We now make a simple, intuitive, but erroneous calculation in order to demonstrate that the results of this section may counter a seemingly reasonable train of thought. The total number of beneficial mutations produced per generation by a population is equal to the beneficial mutation rate times the number of individuals in the population, μN . Suppose that a fraction $4\bar{s}$ survive drift. If one made the assumption that a certain fixed fraction, β , of these beneficial mutations go to fixation, then the rate of substitution would be $4\bar{s}\beta\mu N$. Put differently, the rate of substitution might be presumed to be a linear function of either mutation rate or population size. In this section, however, we show that when clonal interference is taken into account and the population is large, mutation rate and population size have surprisingly small effects on substitution rate.

With the fixation probability given by equation (4), the expected substitution rate of beneficial mutations is given by

$$\langle \sigma(\alpha, \mu, N) \rangle = \mu N \Pr\{fix|\alpha, \mu, N\} , \quad (5)$$

where $\langle \cdot \rangle$ denotes the expected value. As shown in Figure 2.3, a very large change in beneficial mutation rate (several orders of magnitude) has little effect on the substitution rate of the population, especially when the population is large. This constraint may be thought of as a "law of diminishing returns," where the investment is the number of beneficial mutations produced by a population and the returns are adaptive substitutions.

The expected selection coefficient of successful mutations

Figure 2.1 showed that there is some critical value of s below which the probability of fixation of a beneficial mutation is essentially zero. A beneficial mutation whose selective advantage is small is not likely to become fixed because it must compete with many superior mutations. On the other hand, a beneficial mutation whose advantage is large is less likely to be produced. Hence, there must be some intermediate selection coefficient that balances the fixation advantage of large s with the more frequent occurrence of small s . This balance corresponds to the expected selection coefficient of successful mutations.

•••

:

5

•

()

3

0

5

Let $p(s) = K \pi(s) e^{-\lambda(s, \alpha, \mu, N) - \alpha s}$, where K is a normalizing constant such that $\int_0^{\infty} p(s) ds = 1$. Then $p(s)$ is the probability density that a beneficial mutation of selective advantage s will be (i) produced and (ii) fixed. Therefore, the expected value for the selection coefficient of successful mutations is

$$\langle s(\alpha, \mu, N) \rangle = \int_0^{\infty} s p(s) ds . \quad (6)$$

Figure 2.4 reveals that this expectation is essentially constant for $\mu N < 0.01$ and increases approximately linearly with the log of population size when $\mu N > 0.1$.

Effect of clonal interference on rate of fitness increase

At this point, sufficient information has been provided to determine how clonal interference between beneficial mutations affects the rate of adaptive evolution. Having derived (i) the rate at which substitutions occur and (ii) the expected selective advantage conferred by substitutions,

we do

simply

Equa

diff

MAX

Inde

Math

of

Int

Est

con

ten

the

the

the

we now calculate the expected rate of fitness increase simply as the product of (i) and (ii):

$$\begin{aligned} \frac{dw}{dt} &= \langle \sigma(\alpha, \mu, N) \rangle \langle s(\alpha, \mu, N) \rangle \\ &= \alpha \mu N \int_0^{\infty} s \pi(s) e^{-\lambda(s, \alpha, \mu, N) - \alpha s} ds \quad . \end{aligned} \tag{7}$$

Equation (7) is plotted against $\text{Log}_{10} N$ in Figure 2.5 for different mutation rates. It appears that $\frac{dw}{dt}$ approaches a maximum value for increasing N . The same is true for μ . Indeed, that a maximum value exists can be shown mathematically. The implication is that there exists a sort of "speed limit" for asexual evolution imposed by clonal interference.

Estimation of parameters: an empirical example

The previous developments show some characteristic consequences of clonal interference; yet, these developments remain at the level of sweeping generalities until we find the region of parameter space in which biological reality lies. We demonstrate here that the parameters α and μ may, in fact, be estimated empirically.

Equations (5) and (6) govern the expected rate of substitution and the expected selection coefficient of substitutions, respectively, both being functions of α , μ , and N . If N is known, then the resulting two equations contain two unknowns and are linearly independent:

$$\begin{aligned} \langle s(\alpha, \mu) \rangle - s_{obs} &= 0 \\ \langle \sigma(\alpha, \mu) \rangle - \sigma_{obs} &= 0 \end{aligned} \quad (8)$$

The parameters α and μ may, therefore, be determined from this pair of equations given observed values for the substitution rate, σ_{obs} , and the selection coefficient of substitutions, s_{obs} .

It is possible to obtain such values by tracking the fitness trajectory of an evolving population (Lenski et al., 1991; Lenski & Travisano, 1994). The average time between periodic selection events gives the reciprocal of the substitution rate estimate; the average fitness increase caused by periodic selection events gives an estimate for the selection coefficient of substitutions.

As an example, we estimate α and μ using the fitness trajectory observed for an evolving *Escherichia coli*

population (Lenski et al., 1991; Lenski & Travisano, 1994). This example serves two purposes: (i) it demonstrates the estimation procedure, and (ii) it puts us in the "biological ball-park" of parameter space. Lenski and colleagues serially propagated several *E. coli* populations for ten thousand generations of binary fission in a constant environment. (A particularly nice feature of working with bacteria is that samples of the evolving populations may be frozen and later "resurrected" for comparison with samples from earlier or later times. In this way, one may track the evolution of populations over time by competing the evolved populations against the ancestor to estimate their relative fitness.) That calculation of generation number implies a discrete-time formulation of population growth, whereas the mathematics in this paper employ a continuous-time formulation. In the following estimation of parameters, we adjust the number of generations by a factor of $\ln 2$ (≈ 0.693) to reflect this difference. During the first 2000 generations of binary fission (≈ 1400 natural generations), they intensively assayed fitness for one population (Lenski & Travisano, 1994). The observed fitness trajectory was characteristically punctuated with sudden

fitness increases followed by long periods of stasis. This general pattern is in accordance with the results of previous sections: that due to clonal interference, the substitution of a beneficial mutation is a rare, isolated event, and that the fitness increases due to substitutions are large. Based on three sudden fitness increases during ~1400 natural generations (Lenski & Travisano, 1994), the average substitution rate is estimated as $\sigma_{obs} = 0.002$ substitutions per generation; the average fitness increase resulting from a substitution is $s_{obs} = 0.1$. The effective population size with respect to the substitution of beneficial mutations, and given the serial transfer regime, was determined to be 3.3×10^7 (Lenski et al., 1991).

We have estimated parameters α and μ from these data by finding the point of intersection between the solution curves of equations (8). The solution for this system of equations is $\alpha = 35$ and $\mu = 2.0 \times 10^{-9}$ beneficial mutations per replication. Given that the genomic mutation rate of *E. coli* is approximately 3×10^{-3} mutations per replication (Drake, 1991), one can infer that the proportion of mutations that are beneficial is roughly one in a million.

We emphasize that these estimates depend on (i) the assumption of an exponential distribution of beneficial mutational effects, and (ii) the assumption that α and μ remain constant even as mean fitness increases. The empirical fitness trajectories referred to in this section show a decreasing rate of increase, suggesting that assumption (ii) is false if the environment is constant. (See *Assumptions of the models.*)

Transiently common mutations

Clonal interference - a general model

Suppose that, while one beneficial mutation grows in number, a second beneficial mutation appears that is superior to the first. The population is now composed of three genotypes of interest: the ancestor and two competing beneficial mutations. If the first beneficial mutation is not close to fixation, then its growth is unaffected by the growth of the second, superior mutation until the latter has become sufficiently abundant to affect the mean fitness of

the population noticeably. When the superior mutation attains sufficient number, the growth of the original mutation is retarded until, at some point, it reaches a maximum frequency and then begins to decline. We are interested in the probability that the frequency at which this maximum occurs is greater than or equal to some frequency, f . To determine the probability that any particular beneficial mutation achieves a frequency of at least f , we begin by computing the time, t_z , at which a superior mutation with selective advantage s_z must have appeared to insure that the original mutation achieves a maximum frequency of exactly f . Then we calculate the probability that no such superior mutation occurs in the interval $(0, t_z)$; this is the probability that the original mutation achieves a maximum frequency of at least f . (Note that t_z is itself a function of the selective advantage, s_z , of a given superior mutation.) To facilitate presentation of this development, we introduce the term *candidate replication* to refer to any replication event which, if it were to produce a superior mutation, would prevent the original mutation from attaining frequency f .

Consider a three-genotype system with ancestor x , original beneficial mutant y , and alternative superior mutant z ; the deterministic solution for the dynamics of such a system is derived in Appendix 2.2. The time, t_{\max} , at which beneficial mutation, y , reaches maximum number is a function of the time of occurrence, t_z , of an alternative mutation, z , which is superior to y , i.e., $t_{\max} = t_{\max}(t_z)$.

The time, t_z , is that which satisfies $y(t_{\max}(t_z)) = fN$,

where $t_{\max}(t_z)$ is such that $\left. \frac{dy}{dt} \right|_{t_{\max}(t_z)} = 0$. If the superior

mutation, z , were to occur before time t_z , then the original mutation, y , would not achieve frequency f . We can, therefore, calculate the probability that no superior mutation occurs in the interval $(0, t_z)$ by determining the expected number of such mutations in this interval and assuming that they are Poisson distributed.

The first step in determining the expected number of superior mutations interfering with the original mutation is to calculate how many candidate replications take place, i.e., the number of ancestral replications in the interval

$(0, t_z)$, or $R = \int_0^{t_z} x(t) dt$. But t_z is a function of the

selective advantage, s_z , of the superior mutation. R is closely approximated by evaluating $t_z(s_z)$ at the expected value for s_z conditional on it being greater than s_y , i.e.,

$$\hat{t}_z = t_z(\langle s_z | s_z > s_y \rangle), \text{ where } \langle s_z | s_z > s_y \rangle = s_y + \frac{1}{\alpha} \text{ is the expected}$$

selection coefficient of a superior mutation.

The expected number of beneficial mutations in the interval $(0, \hat{t}_z)$ is μR , where μ is the per-replication rate at which beneficial mutations are produced. Of these μR beneficial mutations, only the fraction $e^{-\alpha s_y}$ will be competitively superior to y , the original beneficial mutation. And of these $\mu R e^{-\alpha s_y}$ superior mutations, only another fraction $\pi(s_y + \frac{1}{\alpha})$ will survive drift. Therefore, the expected number of beneficial mutations that occur in the interval $(0, \hat{t}_z)$, that are superior to y , and that

survive drift is $\psi = \mu R e^{-\alpha s_y} \pi(s_y + \frac{1}{\alpha})$. Because this expectation is a function of s_y but not s_z , we simplify our notation at this point by letting $s = s_y$. The analytical solution for R , the number of candidate replications, is derived in Appendix 2.3. The resulting expected number of superior mutations that would prevent a given beneficial mutation from attaining frequency f is

$$\psi(s, \alpha, \mu, N, f) = \mu N \ln(N/\chi) e^{-\alpha s} \pi(s + \frac{1}{\alpha}) , \quad (9)$$

where $\chi = 1 + \frac{1}{\alpha s + 1} \left((\frac{1}{f} - 1) (\alpha s N)^{\frac{\alpha s}{\alpha s + 1}} - \alpha s \right)$. Thus, the

probability that a given beneficial mutation achieves a maximum frequency of at least f is

$$\pi(s) e^{-\psi(s, \alpha, \mu, N, f)} , \quad (10)$$

where the effect of drift is incorporated by $\pi(s)$.

It is important to point out that equation (9) incorporates an approximation that is essentially an equality for $f < 0.95$, but which introduces significant

error for $f > 0.99$. (See Appendix 2.3 for details.) A technical difficulty with equation (10) is that there is no guarantee that $\psi(s, \alpha, \mu, N, f)$ is non-negative, whereas a fundamental assumption of the Poisson process is that the Poisson parameter be non-negative. To remedy this problem, we impose the condition,

$$\psi(s, \alpha, \mu, N, f) = \max\{\psi(s, \alpha, \mu, N, f), 0\}.$$

Otherwise, a negative Poisson parameter may arise if superior mutation z must appear *before* original beneficial mutation y to insure that the latter attains maximum frequency f , i.e., \hat{t}_z is negative. In this case, the probability that a given beneficial mutation achieves a maximum frequency of at least f is equal to one, because an assumption of our analysis is that the superior mutation z does not appear before original mutation y . We have shown that this assumption does not introduce much error (see *Assumptions of the models*).

Probability of transiently polymorphic beneficial mutations

In this section, our objective is to determine with what probability one might expect a beneficial mutation to rise *temporarily* to polymorphic frequency. We define polymorphic frequency as any frequency greater than or equal to 0.01. In the *Clonal interference and fixation* section, we were only concerned with whether or not a beneficial mutation became fixed in a population, i.e., whether or not $f \geq \frac{N-1}{N}$. Now, we examine the probability that the frequency, f , of a beneficial mutation exceeds 0.01 yet never reaches $\frac{N-1}{N}$. This is the probability that a mutation will be transiently polymorphic.

Given that a beneficial mutation survives drift, the probability that it will achieve polymorphic frequency is $e^{-\Psi(s, \alpha, \mu, N, 0.01)}$. Given that the same mutation achieves polymorphic frequency, the probability that it does not reach fixation is computed as the probability that at least one superior mutation appears in the interval (\hat{t}_z, t_f) . The expected number of superior mutations appearing in this interval is:

$$\gamma(s, \alpha, \mu, N, f) = \frac{\mu}{s} N \ln(\chi) e^{-\alpha s} \pi(s + \frac{1}{\alpha}) , \quad (11)$$

where χ is as defined in equation (9). Therefore, given that a mutation with selective advantage, s , has achieved polymorphic frequency, the probability that it does not reach fixation is $1 - e^{-\gamma(s, \alpha, \mu, N, 0.01)}$. The probability that a mutation will be transiently polymorphic is the product of (i) the probability that the mutation survives drift, (ii) the probability that the mutation achieves polymorphic frequency given (i), and (iii) the probability that the mutation does not reach fixation given (ii). Therefore, the probability that any arbitrarily chosen beneficial mutation transiently achieves polymorphic frequency is

$$Pr\{poly|\alpha, \mu, N\} = \alpha \int_0^{\infty} \pi(s) e^{-\psi(s, \alpha, \mu, N, 0.01) - \alpha s} \left(1 - e^{-\gamma(s, \mu, \alpha, N, 0.01)} \right) ds . \quad (12)$$

This equation is plotted in Figure 2.6. Given a certain population size, there is an intermediate value of the beneficial mutation rate at which the probability is greatest that an arbitrarily chosen beneficial mutation will

transiently achieve polymorphic frequency. Likewise, given a certain beneficial mutation rate, there is an intermediate population size that maximizes the probability that an arbitrary beneficial mutation will be transiently polymorphic. This result seems reasonable, because an increased recruitment rate of beneficial mutations, μN , increases the probability that a superior mutation occurs before a given beneficial mutation can reach polymorphic frequency (i.e., increases clonal interference). By lowering μN , on the other hand, one reduces the probability that a superior mutation occurs later, hence increasing the probability that a beneficial mutation, which has already achieved polymorphic frequency, will go to fixation (i.e., is not transient).

Given the parameters estimated previously for an evolving *E. coli* population ($\alpha = 35$, $\mu = 2.0 \times 10^{-9}$, $N = 3.3 \times 10^7$), the probability that an arbitrarily chosen beneficial mutation becomes transiently polymorphic is approximately 0.034. With $\mu N \approx 0.07$, a beneficial mutation would have occurred every 15 generations or so. Of these, about 1 in 30 would become transiently polymorphic. Hence, one would expect about three transient polymorphisms

($f > 0.01$) in 1400 natural generations. This number is roughly comparable to the number of periodic selection events that were observed. This correspondence suggests that each beneficial mutation that went to fixation displaced not only its "parent" genotype but also a "sibling" genotype that had achieved some success.

Surprisingly, these estimates do not rely heavily on the assumption that beneficial mutations are exponentially distributed. Calculations based on an alternative rectangular distribution show that the probability that a beneficial mutation transiently achieves polymorphic frequency is approximately 0.05. Thus, by assuming a rectangular distribution, one might expect about four transient polymorphisms in 1400 natural generations. The fact that assuming such very different distributions results in less than a two-fold difference in estimates suggests that these results are fairly robust. The section *Assumptions of the models* gives a more complete discussion of this test of robustness.

Probability of transiently common mutations: the leapfrog

In a slight variation of the previous section, we will now examine the probability that a beneficial mutation achieves a frequency of 0.5 but is not fixed. We devote a separate section to this special case because of the strange dynamics it would present to an observer of a population in which it occurred. In this case, a mutant Ab reaches majority status before being supplanted by a superior mutant aB , where both mutants are derived directly from the same ancestor ab . At the genetic level, this appears as a "leapfrog" episode in which (i) Ab replaces ab as the most common genotype and thereafter aB replaces Ab as the most common genotype, even though (ii) aB is more closely related to ab than to Ab (Figure 2.7). If one were to sample this population at times t_1 , t_2 , and t_3 , as indicated in Figure 2.8, then one would observe that the sample from t_3 is more closely related at the genetic level to the sample taken at t_1 than to that taken at t_2 .

Following the derivation of equation (12), the probability that an arbitrarily chosen beneficial mutation transiently achieves a frequency of 0.5 or more is

$$Pr\{leapfrog|\alpha, \mu, N\} = \alpha \int_0^{\infty} \pi(s) e^{-\psi(s, \alpha, \mu, N, 0.5) - \alpha s} \left(1 - e^{-\gamma(s, \mu, \alpha, N, 0.5)} \right) ds . \quad (13)$$

Using the parameters previously estimated from an evolving *E. coli* population, and following the same logic as described at the end of the previous subsection, a beneficial mutation would occur every 15 generations or so. About one in every 55 of these mutations would be subject to the "leapfrog" effect, which should thus occur every 800 generations or so. Therefore, it is quite possible that some of the three periodic selection events observed during the 1400 natural generations experiment were complicated by this effect. Whether empirical data would resolve the leapfrog as one or two periodic selection events would depend, in part, on how close in time the relevant genotypes became numerically dominant. If a leapfrog was resolved as two distinct periodic selection events, then the descendants after 1400 natural generations should differ from the founding ancestral genotype by fewer than the three beneficial mutations that would be expected under the

presumption that each periodic selection event was caused by the sequential substitution of an additional mutation (Figure 2.8). Figure 2.9a shows a numerical simulation of the *E. coli* populations using the empirically estimated parameters. The resulting trajectory for mean fitness, shown in Figure 2.9b, illustrates that a single leapfrog episode may indeed give the appearance of two periodic selection events. But rather than implying the successive fixation of two beneficial mutations, only a single substitution has actually occurred.

Discussion

Summary of results

Competition between clones that carry different beneficial mutations may be very important for the evolutionary dynamics of asexual populations. The prevalence of such "clonal interference" among beneficial mutations increases dramatically with population size and with mutation rate. The following points summarize some of the most salient consequences of clonal interference. (1)

The fixation probability of a given beneficial mutation is a decreasing function of both population size and mutation rate. (2) Substitutions appear as discrete, rare events, no matter how frequently beneficial mutations arise. If a beneficial mutation is to overcome clonal interference and become fixed, then it must confer a substantial selective advantage. The advantage that is required for a reasonable probability of fixation is an increasing function of population size and mutation rate. (3) The rate of fitness increase is an increasing function of both population size and mutation rate, but it is only weakly dependent on these parameters when their product is not small. (4) Using observable trajectories for the mean fitness of evolving asexual populations, it is possible to estimate both the beneficial mutation rate and the distribution of beneficial mutational effects. We obtained such estimates for an evolving laboratory population of *Escherichia coli*. (5) Beneficial mutations that become transiently abundant, but which do not go to fixation, may be quite common in asexual populations. (6) Some of these transient polymorphisms may give rise to a "leapfrog" effect, in which the majority genotype at some point in time is less closely related to

the immediately preceding majority genotype than to an earlier genotype. Parameter estimates obtained for the evolving laboratory population of *E. coli* are consistent with this effect being an important feature of asexual dynamical systems.

Assumptions of the models

The models presented here assume that the general form of the distribution of beneficial mutational effects is that of an exponential distribution. Kimura (1979) employs the more general gamma distribution to describe the distribution of *deleterious* mutational effects. Elena et al. (1997) have shown that a compound gamma-rectangular distribution fits well to experimental data from transposon-induced mutations in *E. coli*. Intuitively, the exponential distribution seems a good choice for beneficial mutational effects, because it is reasonable to suppose that there are many more beneficial mutations of small effect than of large effect. Fisher (1930) reasoned that most mutations of large effect are deleterious as a geometrical consequence of the high dimensionality of fitness landscapes. He argued that the

ratio of deleterious to beneficial mutations increases with mutational effect (i.e., phenotypic difference between mutant and non-mutant), because a large radius in phenotypic space is very likely to circumscribe potential improvements, whereas a small radius stands a better chance of being tangent to an improvement. That this effect increases with the dimensionality of the fitness landscape is an easily demonstrable fact of geometry. A convincing argument for the use of the exponential distribution in particular comes from extreme value theory (see Gillespie, 1991, p. 262).

Suppose that M fitness alleles are present in a population such that $W_{[1]} > W_{[2]} > W_{[3]} > \dots > W_{[M]}$ (where W denotes fitness). If the population is in dynamic equilibrium, then the fittest of these M alleles greatly outnumbers the other $M-1$ alleles, which are held at some low frequency by mutation-selection balance. A fitness mutation results in a genotype whose fitness is drawn at random from some unknown parent distribution. Now, imagine that a novel fitness mutation appears that is *beneficial* (i.e., fitter than the current fittest genotype). If we denote the fitness of this

mutation by $W_{[0]}$, then $W_{[0]} > W_{[1]}$ and the selection coefficient of this novel mutation is $s = \frac{W_{[0]}}{W_{[1]}} - 1$. Gillespie (1991)

shows that this s is exponentially distributed in the limit as $M \rightarrow \infty$, regardless of the shape of the unknown parent distribution. (The reason for this result has to do with the fact that $W_{[0]}$ and $W_{[1]}$ are the two largest fitnesses; they are extreme values of the parent distribution.) In other words, in the limit of infinite fitness alleles, the distribution of s is necessarily exponential.

To evaluate the sensitivity of our analysis to the assumption that selection coefficients are exponentially distributed, we replaced the exponential density with a

rectangular density, i.e., $p(s) = \begin{cases} 1/s_{\max} & , \quad s < s_{\max} \\ 0 & , \quad s > s_{\max} \end{cases}$. See

Appendix 2.4 for results of these derivations. Following the logic employed in *Estimation of parameters*, we have estimated the parameters s_{\max} and μ to be 0.12 and 5×10^{-10} , respectively, for an evolving laboratory population of *E. coli*. Note that the beneficial mutation rate estimate, μ , is of the same order of magnitude as was obtained assuming exponentially distributed selection coefficients. Figure 2.10 further shows that replacing the exponential with a

rectangular density changes the resulting fixation probabilities only slightly. The slight discrepancy when μN is very small, such that clonal interference is unimportant, arises because the average selection coefficient (and hence $4s$) is slightly higher for the rectangular than for the exponential distribution. The probabilities of transient polymorphisms (either $f > 0.01$ or $f > 0.5$) are consistently higher when a rectangular density is assumed, although the discrepancies are small. In view of these results, our analyses appear to be reasonably robust with respect to the form of the distribution of selection coefficients.

A second assumption of our analyses is that neither the beneficial mutation rate nor the distribution of selection coefficients changes over time. But in a constant environment, a population becomes better adapted with time, leaving progressively less room for further improvement. It is likely that a well adapted population has (i) a lower overall rate of beneficial mutation, (ii) a smaller average effect of beneficial mutations, or both. Consequently, $\mu = \mu(w)$ may be a decreasing function of fitness, while $\alpha = \alpha(w)$ may increase with fitness. These parameters are therefore constant only when w is constant. This condition may be met

in an environment that changes just fast enough to counter adaptation of a population.

A third assumption made in these models is that the progress of a given beneficial mutation is unaffected by the presence of *inferior* beneficial mutations. By definition, inferior beneficial mutations cannot themselves competitively exclude a given beneficial mutation. However, the selective advantage of a given beneficial mutation will be lower relative to these inferior beneficial mutations than relative to the ancestral genotype, and so inferior beneficial mutations may prolong the time that is required for fixation of a given beneficial mutation. As a consequence, there may be a longer interval in which a superior beneficial mutation could appear that would prevent fixation of the original beneficial mutation. To address this possible complication, the probabilistic models were made fully dynamic by considering all beneficial mutations since the most recent substitution. When the dust settled, the results were essentially unchanged from those that we have presented. For example, the fixation probability of an arbitrarily chosen beneficial mutation was changed by a

factor of $\frac{\alpha}{\alpha+1}$ which is inconsequential since α is generally large.

An assumption made in estimating the parameters α and μ from observed fitness trajectories is that the sudden jumps observed in these trajectories are, in fact, fixation events. Based on the results of *The leapfrog*, however, this assumption is questionable; if a leapfrog event were to occur, then it would give the appearance of two such fixation events (Figure 2.9). Thus, of the three observed jumps in fitness during 1400 natural generations (Lenski & Travisano, 1994), for example, perhaps only two were actual fixations and one was the result of a leapfrog episode. If this were the case, then our estimates of α and μ would be incorrect. To evaluate the degree to which these estimates may be in error, we changed the assumption that observed fitness jumps represent fixations and assumed instead that these jumps represent beneficial mutations that achieved a frequency of $f > 0.5$. To that end, we employed the derivations of *Clonal interference - a general model*. This change of assumption did not appreciably affect the estimates ($\alpha = 29$, $\mu = 1.6 \times 10^{-9}$), indicating that our

initial assumption, at least in this case, did not introduce much error.

Model validation by simulation

A general result of the section, *Clonal interference and fixation*, is that trajectories of population mean fitness are characteristically punctuated, with sudden jumps followed by long periods of stasis, regardless of the mutation recruitment rate, μN . To assess this general prediction qualitatively, we simulated the occurrence of, and competition among, many different beneficial mutations whose selection coefficients were drawn at random from an exponential distribution. Figure 2.9 demonstrates that, despite fierce competition among numerous beneficial mutations ($\mu N = 0.1$), the population mean fitness is not appreciably affected until a fitness variant achieves high frequency. (These results also lend support to the assumption that mutations inferior to the currently dominant variant play a negligible role in clonal interference.)

To test the models quantitatively, we ran repeated simulations. The probabilistic predictions for (i) the

probability of fixation, (ii) the expected fitness increase conferred by a substitution, (iii) the expected substitution rate, (iv) the probability of transiently achieving polymorphic frequency, all agreed well with a large number of fully stochastic simulations. We emphasize that these simulations allow for the more realistic situation in which a mutant may acquire further beneficial mutations at any time after its appearance.

Inclusion of the double mutant

To this point, we have emphasized competition between three genotypes, the progenitor (ab) and two mutants that carry different beneficial mutations (Ab and aB). However, a fourth genotype should eventually appear that has both beneficial mutations (AB). If the effects of the two beneficial mutations on fitness are additive, then the double mutant will eventually take over the population. A full treatment of the dynamics involving this fourth genotype is beyond the scope of this paper. For now, we address only one particular issue. If a leapfrog event is to be manifest, then genotypes Ab and aB must each achieve

majority status before *AB* does; otherwise, the dynamics will appear as a sequential substitution (Figure 2.7). The probability of occurrence of the leapfrog must, therefore, incorporate the probability that sequential substitution does not occur. We have conservatively estimated this

probability as $\exp\left\{-\frac{\pi(s)\mu}{e} \int_0^t y(t) dt\right\}$; this factor was

incorporated into the integrand of equation (13) and found to have a negligible effect (probabilities were reduced by no more than five percent for a wide range of parameters). Therefore, we neglected this factor in our earlier developments in order to keep things as simple as possible.

Implications for the evolution of reproductive strategies

Muller (1964) briefly alludes to the concept of clonal interference while making a case for the evolutionary advantage of sex. Muller argued that adaptive evolution of asexual populations is inefficient, because the fraction of beneficial mutations that are lost due to competition with

alternative beneficial mutations may be substantial in a large population. Recombination would remedy such inefficiency, which suggested an evolutionary advantage for sex. This argument was restated and explored analytically by Crow & Kimura (1965), to which Maynard Smith (1968) responded by pointing out that Muller's original argument relied on the erroneous assumption that mutations were unique events, such that each could occur only once. In a counter-example, Maynard Smith demonstrated that models of sexual and asexual systems yielded the same rate of adaptive evolution when mutations were treated as recurrent events. For a nice summary of this controversy and further developments on this topic, see Felsenstein (1974, 1988). Much recent work has focused on how fixation probabilities are affected by variance in fitness at background loci and the degree of linkage to these loci (Barton, 1993, 1995; Keightley, 1991; Pamilo et al., 1987; Peck et al., 1997). Barton (1994) derived the conditional probability of fixation of a beneficial mutation given that a single substitution occurs or that substitutions occur at a given rate. He explored the dynamics of this probability under varying degrees of recombination. We believe that the

models presented here may contribute to understanding the evolution of sex by giving an explicit expression for the unconditional probability of fixation of a beneficial mutation, in the limit as recombination rate goes to zero.

Another part of a population's reproductive strategy, namely its mutation rate, may also be affected by the clonal interference phenomenon. Much work has been done to determine whether and how natural selection may adjust mutation rates. A high mutation rate may confer an evolutionary advantage, for example, if it increases the rate of substitution of beneficial alleles. This advantage, however, must overcome the disadvantage of a parallel increase in deleterious mutations. Leigh (1970) demonstrated theoretically that elevated mutation rates can evolve in asexual populations that experience oscillating selection on some locus. Since then, much work has supported the notion that evolutionary elevation of mutation rates is at least possible, and perhaps likely, in changing environments (Gillespie, 1981; Ishii et al., 1989). In light of the developments presented in this paper, however, it seems that the strength of selection to elevate mutation rates (above some minimal value set by the physiological

cost of fidelity) may be smaller than the established theory would indicate, especially when populations are large. As we have shown, an increase in mutation rate hardly changes the rate of adaptation of large populations because of clonal interference (Figure 2.5). To gain an appreciable increase in the rate of adaptation for a large population would, therefore, require a disproportionate increase in mutation rate. Such a large increase in mutation rate, however, would undoubtedly have a detrimental effect due to the greatly increased production of deleterious alleles. Consequently, it seems reasonable to suggest that selection for elevated mutation rates should be weak in large populations.

Implications for the general nature of adaptive evolution

Three especially interesting consequences of the results obtained here concern the general nature of adaptive evolution in asexual populations. The first is that one should expect the trajectory for mean fitness of any asexual population to be punctuated with short bursts of rapid, significant increase followed by long periods of stasis,

regardless of the size of the population or its mutation rate. This result contradicts the intuitive, but erroneous, view that discrete bouts of periodic selection (in which individual mutations sweep to fixation) should overlap, thus giving the appearance of continuity, when the mutation recruitment rate, μN , is sufficiently high. A second intriguing implication is that there exists a "speed limit" on the rate of adaptive evolution in asexual populations. As shown in Figure 2.5, the rate of improvement in a population's mean fitness decelerates with increasing μ and N . This result reflects intensified clonal interference as well as the longer time required for selection to proceed to fixation in large populations. A third important consequence is closely related to the second: the rate of adaptive evolution is clearly not always limited by mutation rate. In fact, because of clonal interference, the rate of adaptive evolution is only weakly dependent on mutation rate and population size unless μN is small ($\mu N < 0.1$ for $\alpha = 35$).

Evidence for transiently common beneficial mutations in microbial populations

One of the intriguing consequences of asexuality is that beneficial mutations may become quite common temporarily but eventually go extinct as superior mutations arise (Figures 2.8 & 2.9). In principle, it should be possible to find evidence for this effect in natural populations of asexual organisms. A complication arises, however, in that a beneficial mutation may also become transiently common, but then disappear, if the environment changes so that the mutation is no longer favored.

For example, Holmes et al. (1992) followed the molecular evolution of a population of the human immunodeficiency virus (HIV) within a single infected patient, and their data show several instances of transiently common mutations. In particular, they monitored changes in the RNA sequence encoding the third hypervariable loop of gp120 (V3) throughout the asymptomatic phase of infection (7 years) of a single hemophiliac patient. All 12 viral sequences that were obtained immediately after infection were identical, and these were denoted as sequence

A. In year three, a set of related sequences, denoted C1 through C5, were numerically dominant (11 of 15), but in year seven they and their descendants comprised only a small fraction of the population (2 of 13). By contrast, sequence E1 was present only as a small minority (1 of 15) after three years, but after seven years it and its descendants were numerically dominant (10 of 13). Within the E1 clade, a subset of derived sequences, denoted E2 through E5, became numerically dominant after five years (19 of 23). However, neither they nor their descendants were represented by even a single sequence in years six (0 of 15) and seven (0 of 13). Thus, it is evident that certain mutations became transiently common, only to decline subsequently in frequency. Moreover, the data show the "leapfrog" effect in which the majority type at one point in time is not descended from the majority type that immediately preceded it. Holmes et al. (1992, p. 4838) recognized the importance of these dynamics when they said that changes in the viral population, "instead of being the sequential replacement of one antigenically distinct variant by another, may involve a complex interaction between the different, and competing, evolutionary lineages present in the plasma."

The one caveat to our interpretation of these data, however, is that the host immune system responds to the viruses, and so the HIV population is evolving in a changing environment. Thus, for example, sequences E2 through E5 may not have been driven extinct solely by an intrinsically superior mutation, but instead they may have become selectively disadvantaged after they were targeted by the immune system. This scenario is supported by the fact that the V3 loop is a principal target of the immune system. But even with the added complication of the changing immune environment, asexuality can have important dynamical consequences for HIV and other pathogens. In particular, the "leapfrog" effect necessarily increases the genetic distance between successive majority types (Figures 2.7 & 2.8), and so it may actually facilitate a pathogen's evasion of the host immune response.

An unambiguous demonstration of the "leapfrog" effect will require data from an asexual organism living in a constant environment. To that end, we are now using molecular methods to determine the phylogeny of clones sampled over time from an experimental population of *E. coli*, as it evolved for thousands of generations in a

defined laboratory environment (Lenski et al., 1991; Lenski & Travisano, 1994; Elena et al., 1996). If the "leapfrog" phenomenon is important, then we expect to see a clade become numerically dominant, only to be driven extinct by the emergence of another, even more successful clade that is derived from its ancestral base (rather than from the formerly dominant clade).

A suggestion for further research

Clonal interference is not the only dynamic that inhibits the progression of beneficial mutations to fixation in an asexual population. A similar inhibition may be caused by Muller's ratchet (Muller, 1964; Haigh, 1978), in which deleterious mutations will tend to accumulate in small asexual populations. As shown by Manning & Thompson (1984) and by Peck (1994), the fate of a beneficial mutation is determined as much by the selective disadvantage of any deleterious mutations with which it is linked as by its own selective advantage. In asexual organisms, the entire genome in which a beneficial mutation occurs will remain linked to that mutation and will hitchhike to fixation if

that is the fate of the mutation. Therefore, a beneficial mutation that spreads to fixation presents a severe population bottleneck in which only a single genome is sampled, thus exacerbating the effect of Muller's ratchet. Consequently, a beneficial mutation may only be considered advantageous if its benefit more than compensates for the drastic reduction in effective population size caused by its fixation and the associated acceleration of Muller's ratchet.

Haigh (1978) modeled the effect of deleterious mutations on population fitness; Manning & Thompson (1984) and Peck (1994) modeled the effect of deleterious mutations on the fate of beneficial mutations; and the models presented here provide a quantitative account of how beneficial mutations affect one another. A logical next step would be to integrate these models into a single theoretical treatment of mutation in asexual populations. Such a synthesis would be a valuable contribution toward a general understanding of evolutionary dynamics in asexual systems.

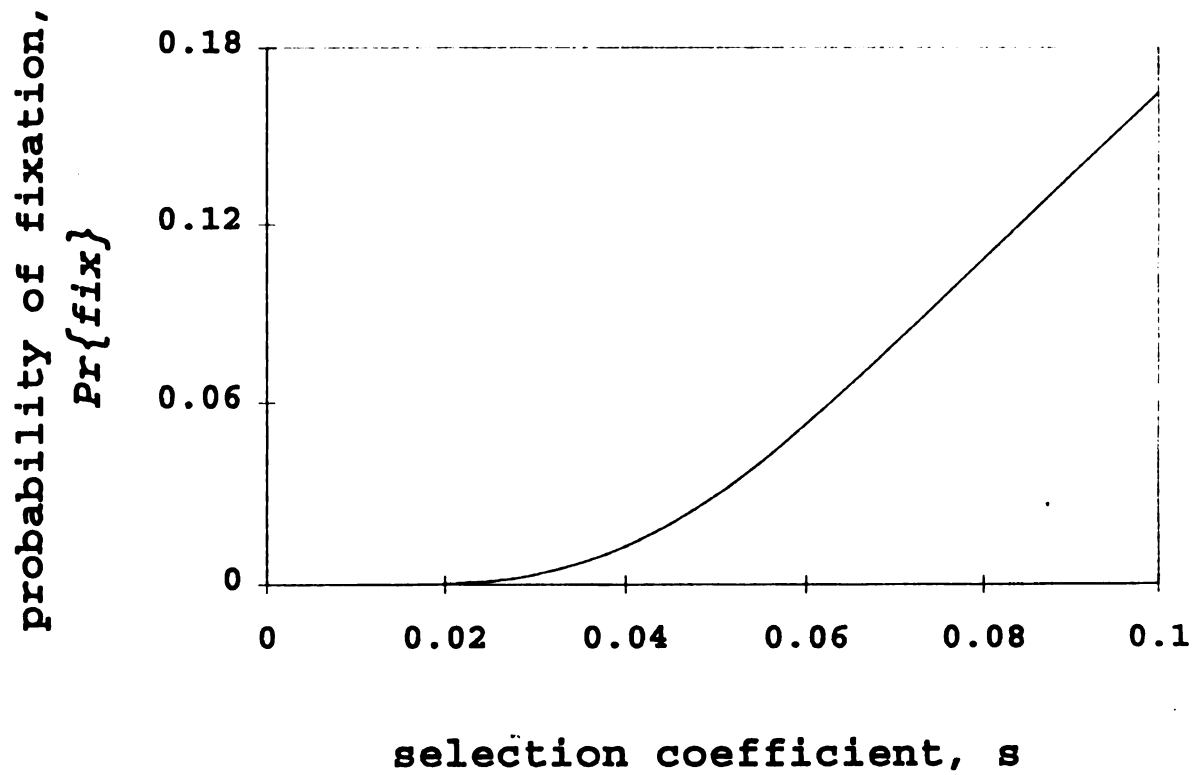


Figure 2.1. The probability that a given beneficial mutation with selection coefficient, s , achieves fixation. Equation (3) with $\alpha = 35$, $\mu = 2.0 \times 10^{-9}$, and $N = 3.3 \times 10^7$.

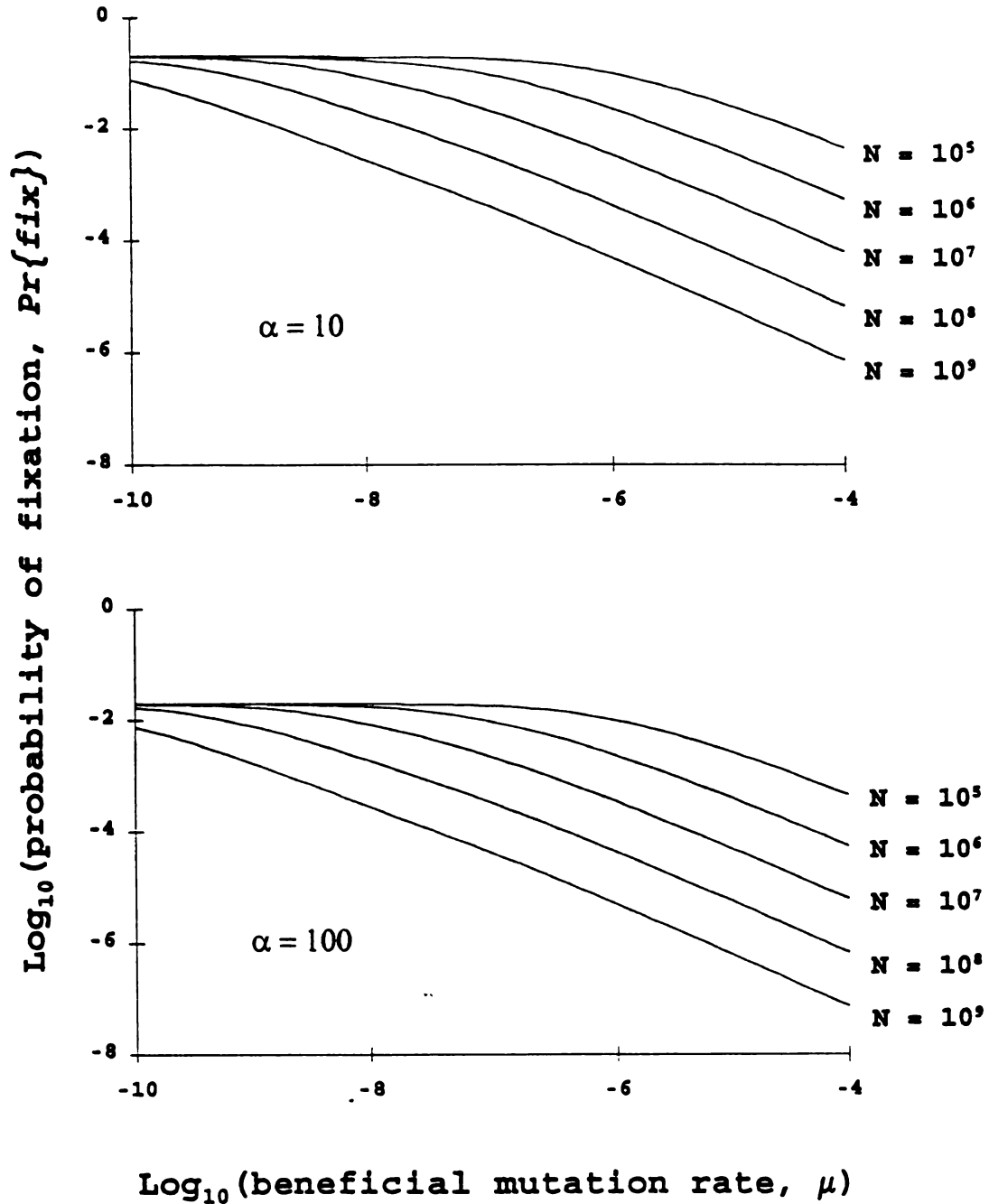


Figure 2.2. The probability of fixation of an arbitrarily chosen beneficial mutation is a decreasing function of both beneficial mutation rate, μ , and population size, N (Equation (4)). The exponential parameter, α , appears only to shift the curves down vertically without changing their shape.

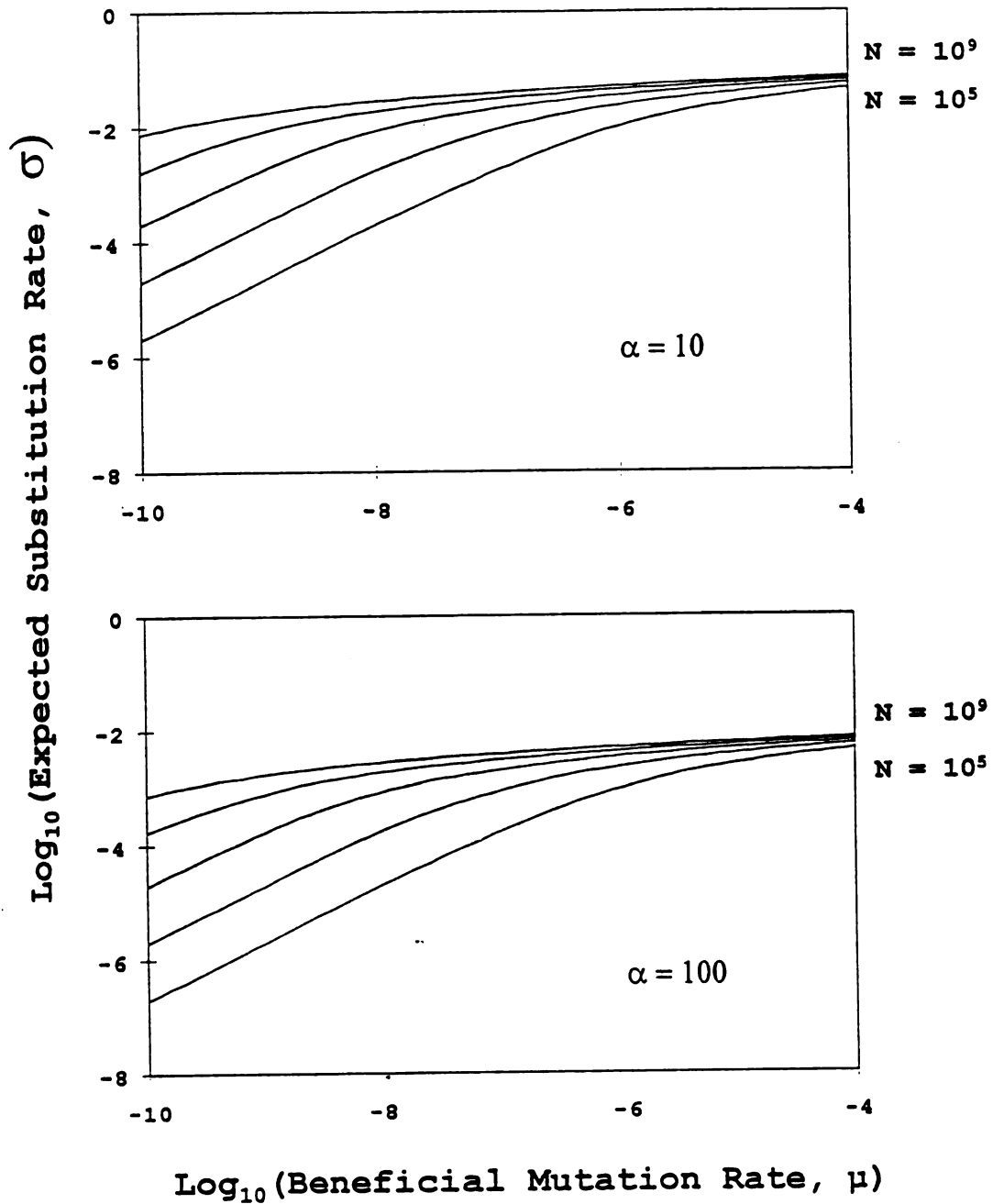


Figure 2.3. The substitution rate of a population is an increasing function of its beneficial mutation rate (Equation (5)). When the population size is large, however, a large change in beneficial mutation rate hardly affects the substitution rate.

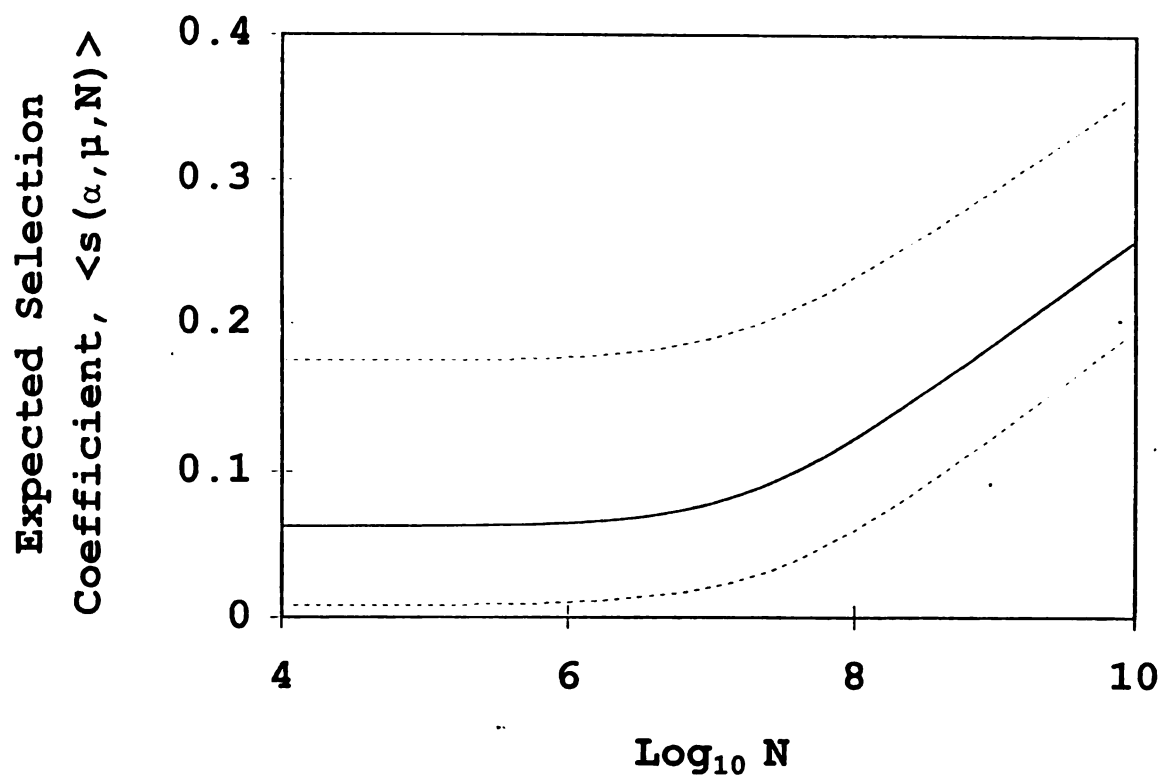


Figure 2.4. The expected selection coefficient of substitutions, $\langle s(\alpha, \mu, N) \rangle$, is an increasing function of population size, N . Equation (6) with $\alpha = 35$, and $\mu = 2.0 \times 10^{-9}$. The solid line indicates the expected value; the dashed lines indicate numerically determined 95% predictive confidence limits.

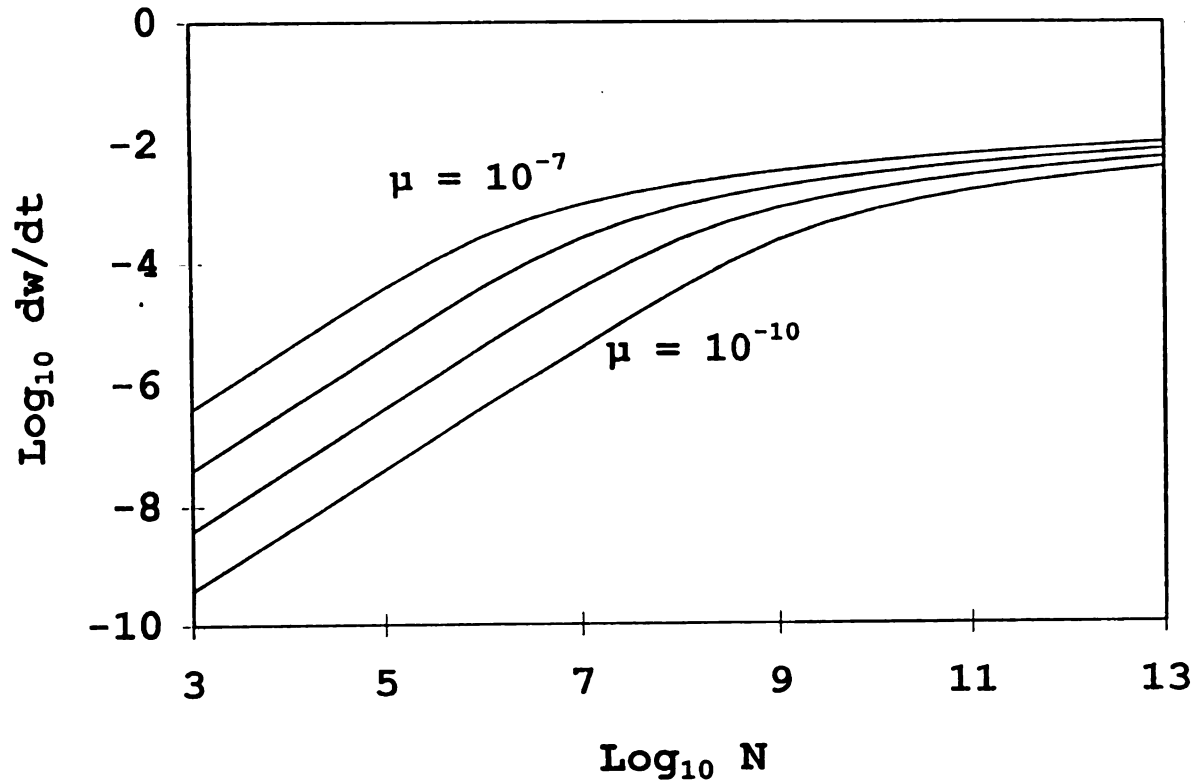


Figure 2.5. The rate of fitness improvement, $\frac{dw}{dt}$, is an increasing function of both population size, N , and beneficial mutation rate, μ . Equation (7) with $\alpha = 35$. This rate of increase decelerates substantially, however, due to increased clonal interference when $\mu N > 0.1$ (i.e., when more than one beneficial mutation is produced on average every ten generations).

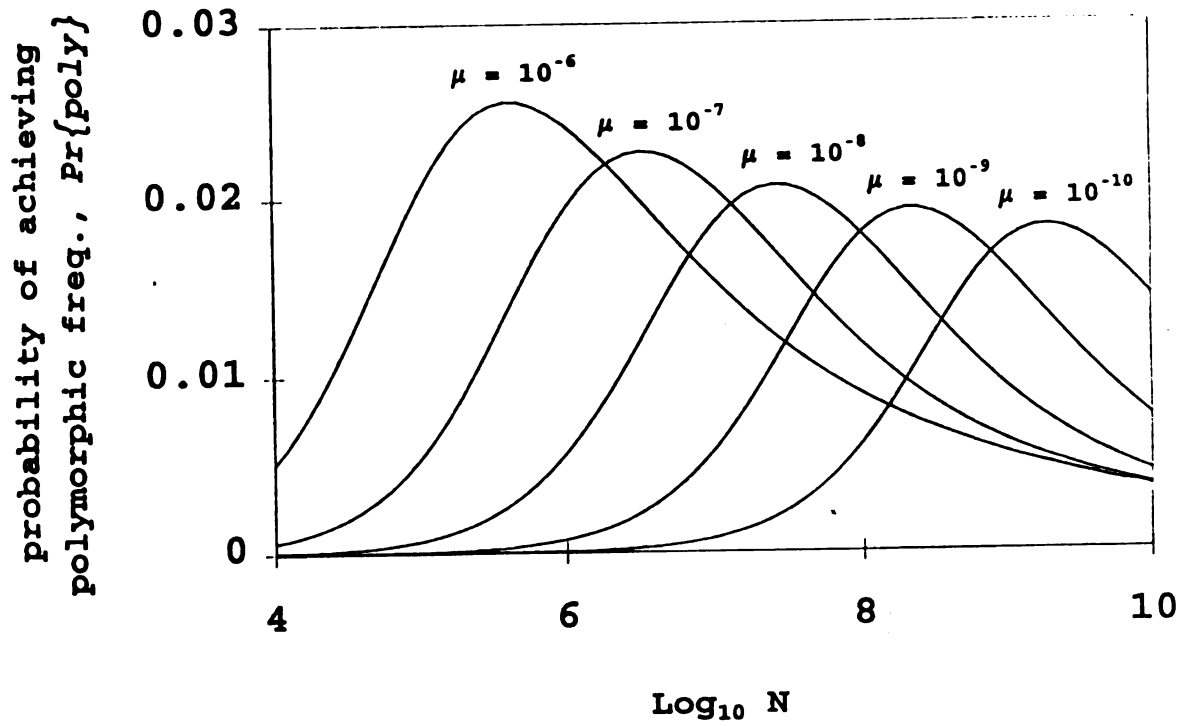


Figure 2.6. The probability that an arbitrarily chosen beneficial mutation transiently achieves polymorphic frequency is plotted against log population size for various beneficial mutation rates. Equation (12) with $\alpha = 35$. For a given beneficial mutation rate, there is an intermediate population size at which the probability of achieving polymorphic frequency is a maximum.

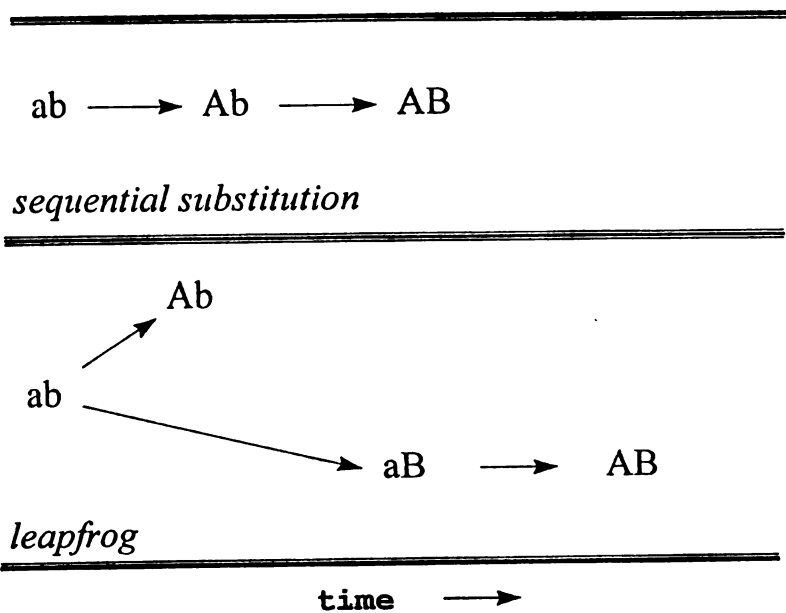


Figure 2.7. The leapfrog phenomenon illustrated phylogenetically. The phylogeny of majority genotypes is compared with that of sequential substitution.

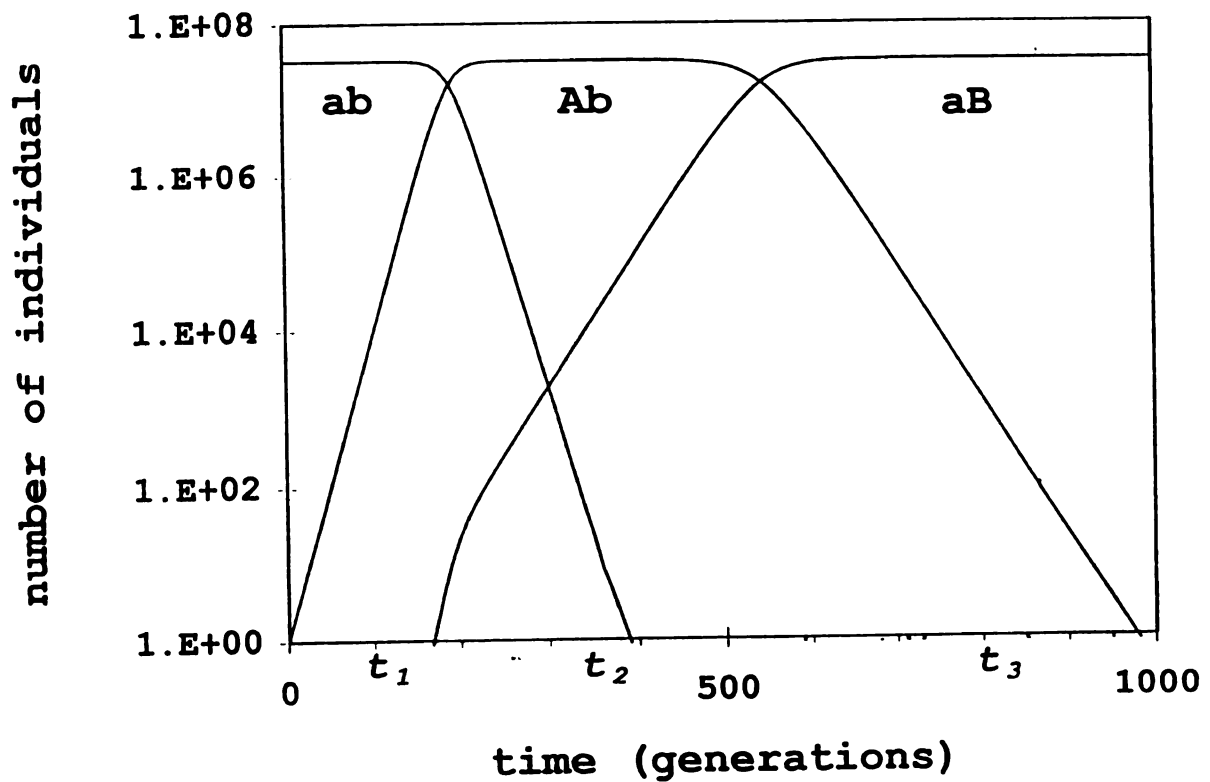


Figure 2.8. The leapfrog phenomenon illustrated dynamically. Genotype *ab* is displaced by mutant *Ab* which is later displaced by alternative mutant *aB*. Equations (25) and (26) with $s_y = 0.09$, $s_z = 0.13$. Note that genotypes sampled at time t_3 are more closely related to those sampled at t_1 than to those sampled at t_2 .

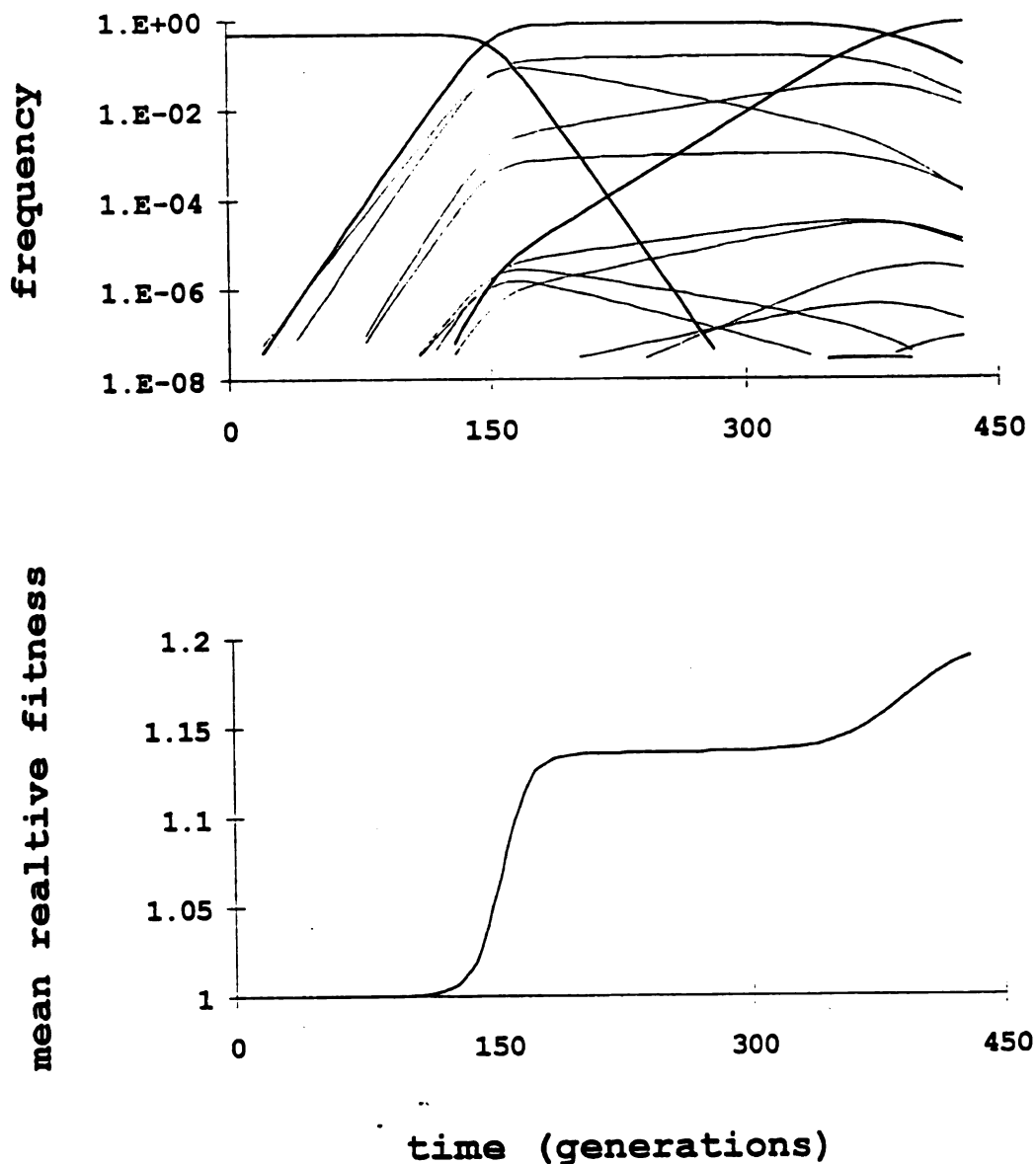


Figure 2.9. (a) A simulation of competition among numerous beneficial mutations. The heavy lines represent genotypes that achieve majority status. Note that a leapfrog event has occurred in this particular simulation. Selection coefficients were drawn at random from an exponential distribution. Parameters used are $\alpha = 35$, $\mu = 2.0 \times 10^{-9}$, $N = 3.3 \times 10^7$. (b) The mean fitness trajectory of the population simulated in panel (a). Note that the leapfrog phenomenon gives the appearance of two distinct periodic selection events.

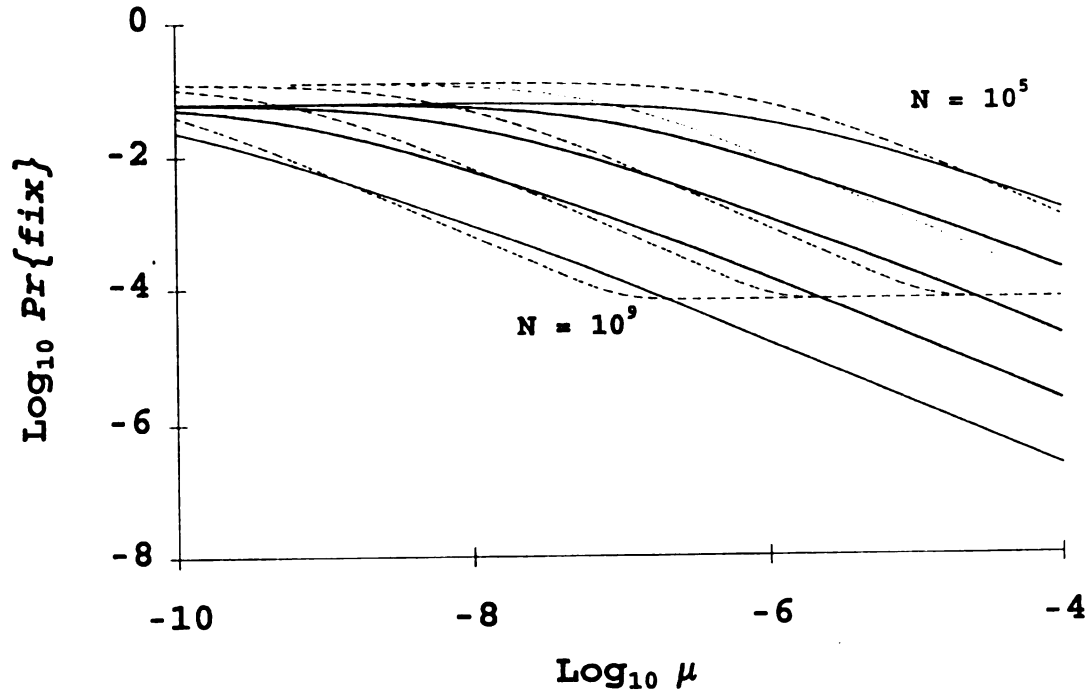


Figure 2.10. The probability of fixation of an arbitrarily chosen beneficial mutation is plotted against the beneficial mutation rate, μ , for various population sizes. The solid lines indicate probabilities assuming an exponential distribution of beneficial mutational effects (Eq. (4) with $\alpha = 35$, as estimated from the *E. coli* populations). The dashed lines indicate probabilities which assume a rectangular distribution of beneficial mutational effects (Eq. (33) with $s_{\max} = 0.12$, as estimated from the *E. coli* populations). The discrepancies resulting from the different distributional assumptions are small.

Chapter 3

THE ORDER OF FIXATION OF BENEFICIAL MUTATIONS IN ASEQUAL POPULATIONS¹

Abstract

In a constant environment, an organism's opportunities for further improvement are slowly depleted. To model adaptive evolution in such an environment, one should consider the number of available beneficial mutations to be finite. Given a finite set of beneficial mutations, the course of adaptive evolution is determined by both the timing and the ordering of fixations. In large, asexual populations, the ordering of fixations is affected by competition among beneficial mutations. Taking into account such competition, we derive the probability of any specified fixation ordering. The trajectories of fitness expectation and variance among replicate populations are then derived, given (i) a particular set of beneficial mutations, or given (ii) only a common distribution of mutational effects. Using fitness data from evolving E. coli populations, we solve the inverse problem to determine (i) the number of available beneficial mutations, (ii) the distribution of beneficial mutational effects, and (iii) the beneficial mutation rate.

¹

This chapter is written in the format of a paper. The "we" in this chapter refer to P.J. Gerrish, R.E. Lenski, and V. Mandrekar.

Introduction

In a continually changing environment, an organism's potential adaptations are never depleted. To model the adaptation of an organism in a changing environment, therefore, it is reasonable to assume the number of available beneficial mutations to be infinite (Gerrish & Lenski, 1998). When an organism's environment is constant, however, then the number of possible genetic improvements is slowly depleted as the organism adapts. To model the adaptation of an organism in a constant environment, therefore, the number of available beneficial mutations must be considered finite.

Given that a finite set of beneficial mutations is available to an organism, all available beneficial mutations are eventually fixed in the population. Whenever one of these beneficial mutations is fixed, the population fitness increases. Because each beneficial mutation is unique, each confers a different increase in population fitness. Thus, the adaptive dynamics of a population are determined as much

by

fix

inc

some

occ

ano

woul

woul

infe

Pani

woul

the

a and

the

Ab an

be ex

be ab

betwe

enough

are no

by the order in which the available beneficial mutations are fixed as by the timing of the fixations.

If there were only two different mutations that would increase the fitness of a particular asexual organism in some environment, and if one of these beneficial mutations occurred in one individual, while the other occurred in another individual, then the progeny of the two individuals would compete. The lineage carrying the superior mutation would competitively exclude both the wildtype and the inferior beneficial mutation (Muller, 1932; Muller, 1964; Pamilo et al., 1987). Eventually, the inferior mutation would appear on the background of the superior mutation and the double mutant would be fixed.

Let there be two beneficial mutations, A and B , and let a and b denote the corresponding wildtype alleles. Let A be the fitter of the two alleles, i.e., $w_A > w_B$. If genotypes Ab and aB are both present in the population, then aB will be excluded, and the order of fixation of the genotypes will be $ab \rightarrow Ab \rightarrow AB$. In small populations, the time interval between occurrence of beneficial mutations is typically long enough to ensure that both alternative beneficial mutations are not simultaneously present in the population. Thus, the

two

(ab

equi

depe

surv

time

shor

bene

popu

alte

Ab -

muta

alte

orde

muta

diff

fixa

each

whic

firs

is g

two possible fixation orderings of the beneficial mutations ($ab \rightarrow aB \rightarrow AB$, and $ab \rightarrow Ab \rightarrow AB$) are approximately equiprobable in small populations, their probabilities depending primarily on the relative rates of appearance and survival of A and B. In large populations, however, the time interval between occurrence of beneficial mutations is short, resulting in a high probability that alternative beneficial mutations are simultaneously present in the population. The simultaneous presence of the two alternative beneficial mutations promotes the ordering, $ab \rightarrow Ab \rightarrow AB$. In general, large populations (and/or high mutation rates) make probable the simultaneous presence of alternative beneficial mutations, which in turn promotes the ordering of fixations from largest to smallest effect mutations.

In what follows, we derive the probabilities of different fixation orderings. Because each ordering of fixations corresponds to a particular fitness trajectory, each trajectory is therefore assigned a probability, from which the aggregate expected trajectory may be computed. We first treat the case in which a set of beneficial mutations is given, such that the set of selection coefficients is

know

par

prol

dis

coe

prol

dif

mut

mut

the

evo

fit.

the

dis

ben.

Pro

Ord.

supe

known. Then we demonstrate how the number of unknown parameters may be reduced either by extracting the most probable set of selection coefficients from a parent distribution or by taking the limit as selection coefficients converge. Finally, we treat the fully probabilistic case in which, instead of assuming that different populations fix the same set of beneficial mutations, we assume only a distribution for beneficial mutational effects. In the discussion section, we employ these developments to estimate parameters of the model for evolving populations of *Escherichia coli* by regressing their fitness trajectories; parameters to be estimated are: (i) the number of available beneficial mutations, (ii) the distribution of beneficial mutational effects, and (iii) the beneficial mutation rate.

Probabilities of fixation orderings

Order of fixation of two beneficial mutations

Let subscripts S and I refer respectively to the superior and inferior of two available beneficial mutations.

Suppose the inferior mutation appears and survives drift before the superior mutation does so. Then the inferior mutation will be fixed before the superior mutation only if the superior mutation does not appear and survive drift before the inferior mutation is fixed. Let t_I denote the time of appearance of an inferior mutation which subsequently survives drift. Then the total time before fixation of the inferior mutation is $t_I + \frac{\ln N}{s_I}$ (see Appendix 3.1), where s_I is the selective advantage of the inferior mutation over the wildtype, and N is population number (assumed constant). Thus, the expected number of times the superior mutation occurs and survives drift before fixation of the inferior mutation is $r_s(t_I + \frac{\ln N}{s_I})$, where $r_s = \mu_s \pi(s_s) N$ is the rate at which the superior mutation appears and subsequently survives drift; μ_s is the rate of mutation to the superior mutation, s_s is the selective advantage of the superior mutation over the wildtype, and $\pi(u)$ is a linear function of u giving the probability that a beneficial mutation with selective advantage u survives the effects of drift (Haldane, 1927; Otto & Whitlock, 1997; Gerrish &

Lenski, 1998). Given t_I , the probability that inferior mutation I is fixed before superior mutation S appears and survives drift is the zero-class of the Poisson distribution, $\exp\{-r_s(t_I + \frac{\ln N}{s_I})\}$. Let $Pr\{I;S\}$ denote the

probability that I is fixed before S . (This is unconventional notation and should not be confused with a conventional notation in which S would denote a condition.) If the probability density function (p.d.f.) of t_I is denoted by $g(t_I)$, then the probability that I is fixed before S is given by:

$$Pr\{I;S\} = \int_0^{\infty} \exp\{-r_s(t_I + \frac{\ln N}{s_I})\} g(t_I) dt_I . \quad (14)$$

The rate at which the inferior mutation appears and subsequently survives drift is $r_I = \mu_I \pi(s_I) N$. This rate is time independent. Therefore, the waiting time p.d.f. for such an event is the homogeneous exponential:

$g(t_I) = r_I e^{-r_I t_I}$. Upon integration, equation (14) becomes

$$Pr\{I;S\} = \left(1 + \frac{\mu_s \pi(s_s)}{\mu_I \pi(s_I)}\right)^{-1} \exp\left\{-\mu_s \pi(s_s) N \frac{\ln N}{s_I}\right\}. \quad (15)$$

For later convenience, let μ denote overall beneficial mutation rate, such that $\mu = \mu_I + \mu_s$. In some of the subsequent figures, we assume individual mutation rates to be equal, implying in this case $\mu_I = \mu_s = \mu/2$.

We now derive the probability that S is fixed before I . The superior mutation will be fixed first only if its appearance and subsequent survival precede the fixation of the inferior mutation. Let t_s denote the time of occurrence of a superior mutation which subsequently survives drift. The inferior mutation cannot be fixed in the interval $(0, t_s)$ if the superior mutation is to precede. The time required for fixation of the inferior mutation is $\frac{\ln N}{s_I}$. Therefore, the superior mutation will be fixed first only if an inferior mutation does not appear and survive drift before time $(t_s - \frac{\ln N}{s_I})^+$, where the notation $(\cdot)^+$ is defined here as $\max(\cdot, 0)$. The expected number of inferior mutations appearing in this interval is $r_I(t_s - \frac{\ln N}{s_I})^+$. Therefore, the

unconditional probability that superior mutation S is fixed before inferior I is given by

$$Pr\{S;I\} = \int_0^{\infty} \exp\{-r_I(t_s - \frac{\ln N}{s_I})^+\} g(t_s) dt_s, \quad (16)$$

where $g(t_s) = r_s e^{-r_s t_s}$, and $r_s = \mu_s \pi(s_s) N$.

As expected, the probability that the superior mutation is fixed first is simply the complement of the probability that the inferior is fixed first: $Pr\{S;I\} = 1 - Pr\{I;S\}$. Note that $\lim_{N \rightarrow \infty} Pr\{S;I\} = 1$, confirming our claim that large population number promotes the ordering of fixation from largest to smallest effect mutations (Figure 3.1).

Moreover, $\lim_{N \rightarrow 1} Pr\{S;I\} = \frac{\mu_s \pi(s_s)}{\mu_I \pi(s_I) + \mu_s \pi(s_s)}$, confirming our

claim that the ordering of fixations in small populations is determined primarily by the relative rates of appearance and survival of beneficial mutations.

One limit of special interest to the empiricist is the limit in which the two selection coefficients converge. This limit is of interest because it is often the case that

little is known about the selection coefficients other than the relation $s_s > s_I$. And it is conveniently the case that selection coefficients cancel out in this limit, rendering the probability independent of either selection coefficient. The limit as selection coefficients converge yields a maximum value for $Pr\{I, S\}$. More precisely, define ϵ as the positive constant which satisfies $s_s = s_I + \epsilon$. Then, given only that $s_s > s_I$, the maximum probability, $Pr\{I, S\}$, is found by taking the limit of (15) as $\epsilon \rightarrow 0$.

A second limit of interest is that in which the mutation rate ratio, μ_I/μ_S , gets large. This limit is of interest because it again yields a maximum probability, $Pr\{I, S\}$, this time eliminating dependence on μ_I . In the cases where selection coefficients as well as the mutation-rate ratio are unknown (see HIV example in *Discussion*), it may be convenient to examine the double limit,

$$\begin{aligned} \sup\{Pr\{I; S\} | \mu_S, N\} &= \\ \lim_{\epsilon \rightarrow 0, \frac{\mu_I}{\mu_S} \rightarrow \infty} Pr\{I; S\} &= e^{-\mu_S K M n N}, \end{aligned} \tag{17}$$

where K is the linear constant in the drift survival function, and we assume the intercept of this function to equal zero, i.e., $\pi(s) = Ks$. (This assumption is reasonable when N is large, because $\pi(0) = 1/N \approx 0$; see Crow & Kimura, 1970.) Note that (17) does not depend on either selection coefficient and gives a maximum value for all cases in which $s_s > s_i$ and μ_i is unknown. A useful inequality may be derived by noting that the probability given by (17) is greater than 0.05 when

$$\mu_s K N \ln N < 3 . \quad (18)$$

The utility of this inequality is demonstrated in the discussion section.

Probability that a superior mutation is the first of many

In the case of many available beneficial mutations, the probability that the superior mutation will be fixed first is an extension of the previous developments. Let the set of M available beneficial mutations be ordered from largest to smallest effect such that $s_1 > s_2 > s_3 > . . . > s_M$. We

assume that these mutational effects are additive, so that the rank-order of selection coefficients does not depend on which mutations may have been fixed previously in the population. The probability that the superior mutation (mutation 1) is the first to be fixed is the probability that none of the $M-1$ inferior mutations are fixed first:

$$Pr\{1;2,3,\dots,M\} = \int_0^{\infty} \exp\left\{-\sum_{i=2}^M r_i \left(t_1 - \frac{\ln N}{s_i}\right)^+\right\} g(t_1) dt_1, \quad (19)$$

where $g(t_1) = r_1 e^{-r_1 t_1}$, and $r_i = \mu_i \pi(s_i) N$.

Probability that fixation ordering is from largest to smallest s

Fixations appear in order from largest to smallest s when the largest of the M available beneficial mutations is fixed first, and the largest of the $M-1$ remaining mutations is fixed next, etc. Given the ordered set of available beneficial mutations as described above and extending this logic, the probability that the mutations will be fixed in

order from largest to smallest selective advantage is, intuitively, $Pr\{1;2,\dots,M\} * Pr\{2;3,\dots,M\} * \dots * Pr\{M-1;M\}$, or:

$$\prod_{i=1}^{M-1} \int_0^{\infty} \exp\left\{-\sum_{j=i+1}^M r_j \left(t_i - \frac{\ln N}{s_j}\right)^+\right\} g(t_i) dt_i, \quad (20)$$

where $g(t_i) = r_i e^{-r_i t_i}$, and $r_i = \mu_i \pi(s_i) N$. Above a certain population number or beneficial mutation rate, this probability of rank-ordered fixations is equal to one (Figure 3.2).

General solution for any ordering of fixations

The previous developments may be generalized to derive an expression for the probability of any specified ordering of fixations. Let $\zeta\{\cdot\}$ denote a permuting function. A permuting function specifies an ordering. For example, $\zeta\{3\} = 1$ would imply that the third mutation to be fixed is the mutation of largest selective advantage. Put differently, the function $\zeta\{\cdot\}$ maps place in the fixation ordering onto rank by selective advantage. (Note, there are $M!$ such functions, corresponding to all possible fixation orderings.) Define I_j to be the set of all beneficial

mutations both inferior to and subsequent to the j^{th} mutation fixed, given the fixation ordering, ζ . Define \mathcal{S}_j to be the set of all beneficial mutations both superior to and subsequent to the j^{th} mutation fixed. Then, the probability of any specified fixation ordering, ζ , is given by:

$$\Phi(\zeta) = \prod_{i=1}^{M-1} \int_0^{\infty} \exp\left\{-\sum_{j \in \mathcal{I}_i} r_j \left(t_i - \frac{\ln N}{s_j}\right)^+ - \sum_{j \in \mathcal{S}_i} r_j \left(t_i + \frac{\ln N}{s_{\zeta(i)}}\right)\right\} g(t_i) dt_i. \quad (21)$$

where $g(t_i) = r_{\zeta(i)} e^{-r_{\zeta(i)} t_i}$. Note that

$$\lim_{N \rightarrow \infty} \Phi(\zeta) = \begin{cases} 1, & \mathcal{S}_j = \emptyset \\ 0, & \mathcal{S}_j \neq \emptyset \end{cases}, \text{ again confirming our claim that}$$

large populations promote rank-ordered fixations. Moreover,

$$\lim_{N \rightarrow \infty} \Phi(\zeta) = \prod_{i=1}^{M-1} \mu_i \pi(s_i) \left(\sum_{j=i}^M \mu_j \pi(s_j) \right)^{-1}, \text{ again confirming our}$$

claim that fixation ordering in small populations is determined primarily by the relative rates of appearance and survival of beneficial mutations. Indeed, if all such

rates, $\mu_i \pi(s_i)$, are equal, then this probability reduces to

$$\lim_{N \rightarrow 1, \mu_i \pi(s_i) = \mu_j \pi(s_j)} \Phi(\zeta) = \frac{1}{M!}, \text{ which agrees with intuition.}$$

The analytical integration of (21) is given in Appendix 3.2.

Fitness trajectories I: parallel populations share a common set of available beneficial mutations

How repeatable is evolution? Albeit on a miniature scale, evolutionary biologists are beginning to address this question experimentally by studying parallel microbial populations evolving in identical environments (e.g., Lenski & Travisano, 1994). We address this question theoretically for the case of adaptive evolution of replicate asexual populations. In the previous section, we derived the probability of any specified ordering of fixations. In this section, we derive the fitness trajectories that correspond to the various fixation orderings. From this, we compute trajectories of fitness expectation and variance among populations.

In identical environments, it is reasonable to assume that genetically identical clones will have the same set of

potential adaptive improvements. Hence, in this section, we assume that parallel populations share a common set of available beneficial mutations.

Expectation and variance

Given some fixation ordering, ζ , the time until fixation of the i^{th} beneficial mutation is $T_i = \sum_{j=1}^i \tau_j + \frac{\ln N}{s_{\{U\}}}$. The time τ_j is a waiting time between the $j-1^{\text{th}}$ fixation and the appearance and survival of the j^{th} mutation; this time plus the time required for fixation of the j^{th} mutation, $\frac{\ln N}{s_{\{U\}}}$, equals the total time between the $j-1^{\text{th}}$ and j^{th}

fixations. The τ_j depend on fixation ordering, ζ . If, for example, the $j+1^{\text{th}}$ mutation to be fixed is superior to the j^{th} mutation, given fixation ordering ζ , then the expected waiting time until occurrence and survival of the j^{th} mutation is shortened because of the condition that it must be fixed before the appearance and survival of the superior $j+1^{\text{th}}$ mutation. Mathematically, given that the j^{th} mutation appears and subsequently survives drift at time τ_j , the

probability that the $j+1^{\text{th}}$ mutation appears after time $\tau_j + \frac{\ln N}{s_{\{U\}}}$ is $\exp\{-r_{j+1}(\tau_j + \frac{\ln N}{s_{\{U\}}})\}$. The condition that a subsequent fixation, say the $j+2^{\text{th}}$ fixation, fixes an inferior mutation will shorten the expected waiting time slightly, because the waiting time for the j^{th} mutation must not be so long as to allow the prior fixation of the inferior $j+2^{\text{th}}$ mutation. Mathematically, the $j+2^{\text{th}}$ mutation appears and survives drift after time $(\tau_j - \frac{\ln N}{s_{\{U+2\}}})^+$ with probability $\exp\{-r_{j+2}(\tau_j - \frac{\ln N}{s_{\{U+2\}}})^+\}$. To generalize, let S_j

denote the set of all beneficial mutations both superior to and subsequent to the j^{th} mutation fixed, and let I_j denote the set of all beneficial mutations both inferior to and subsequent to the j^{th} mutation fixed. Then, given fixation ordering ζ , time τ_j has expectation

$$E(\tau_j | \zeta) = \int_0^{\infty} t g(t) \exp\left\{-\sum_{k \in S_j} r_k \left(t + \frac{\ln N}{s_{\{U\}}}\right) - \sum_{k \in I_j} r_k \left(t - \frac{\ln N}{s_k}\right)^+\right\} dt \quad (22)$$

where $g(t) = r_{\{U\}} e^{-r_{\{U\}} t}$. Thus, the accumulated expected time

until fixation of the i^{th} mutation is

$$E(T_i|\zeta) = \sum_{j=1}^l E(\tau_j|\zeta) + \frac{\ln N}{s_{\{U\}}} . \quad \text{Population fitness after the } i^{\text{th}}$$

fixation is defined as $w_i = 1 + \sum_{j=1}^l s_{\{U\}}$. Therefore, the

function describing the trajectory of population fitness over time is

$$w(t|\zeta) = w_i , \quad (23)$$

where $i = \{k : t \in [E(T_k|\zeta), E(T_{k+1}|\zeta))\}$.

Given the set of beneficial mutations available to an organism, one can determine the expected fitness of the population at time t by weighting each of the possible fitnesses at that time by its probability and summing over all possible trajectories. The set of possible trajectories has a one-to-one correspondence with the set of fixation orderings. To compute expected fitness at time t , one must multiply the population fitness (given a particular fixation ordering) by the probability of that fixation ordering and sum over all possible fixation orderings:

$$E(w|t) = \sum_{\zeta} w(t|\zeta) \Phi(\zeta) , \quad (24)$$

where $\Phi(\zeta)$ and $w(t|\zeta)$ are given by (21) and (23), respectively. Likewise, fitness variance is computed as

$$\text{Var}(w|t) = \left(\sum_{\zeta} w(t|\zeta)^2 \Phi(\zeta) \right) - E(w|t)^2 . \quad (25)$$

We emphasize that this equation gives the trajectory of fitness variance *among replicate populations*, and is therefore an indicator of the *repeatability* of the fitness trajectory (Lenski et al., 1991; Johnson et al., 1995). This quantity is quite distinct from the within-population, among-individual fitness variance, which is the basis of Fisher's Fundamental Theorem (Fisher, 1930).

Figure 3.3 displays the expected fitness trajectories for populations of various sizes. Plotted in the panel below each expected fitness trajectory is the corresponding trajectory for the standard deviation of fitness. This analysis shows the dramatic reduction in fitness variance as population size increases, which results from the increased predictability of the order of fixation of the available beneficial mutations.

Reducing the number of parameters

Selection coefficients of beneficial mutations are rarely known. Consequently, the utility of the previous developments may be enhanced by eliminating the need to specify a particular set of selection coefficients. This may be achieved in one of two ways: (i) the most probable set of selection coefficients may be extracted from some parent distribution, thereby requiring only the parameters of the distribution, or (ii) the limit may be examined in which all selection coefficients converge, which conveniently eliminates dependence on selection coefficients altogether.

Here, we assume a set of selection coefficients based on some parent distribution. To proceed analytically, we have chosen the set of expectations of the M order-statistics (Feller, 1968). Given an exponential parent distribution, as we have assumed for selection coefficients, the i^{th} order-statistic has density

$$g(s_{[i]}) = M \binom{M-1}{i-1} \alpha (1 - e^{-\alpha s_{[1]}})^{M-i} e^{-\alpha s_{[1]}^i} . \quad (26)$$

From this density, we have derived the expectation:

$$E(s_{[i]}) = \frac{1}{\alpha} \sum_{j=i}^M \binom{M}{j} \binom{j}{i} (-1)^{j-i} \frac{i}{j^2} . \quad (27)$$

This gives a set of M selection coefficients. Since the set is based solely on the exponential parameter, α , the number of parameters of our model is reduced. Thus, given only the population number, N , the number of available beneficial mutations, M , the beneficial mutation rate, μ , and an exponential parameter, α , it is now possible to compute the expected fitness trajectory as described in the previous section. Only N , however, is normally known. In the discussion section, we solve the inverse problem of estimating M , μ , and α , given fitness data of evolving populations.

A disadvantage of the above approach is that it requires an assumption about the shape of the distribution of mutational effects. Now, we eliminate the need for such an assumption by examining the limit as selection coefficients converge. This approach may serve as alternate means of reducing the number of parameters of the model.

The probability of any specified fixation ordering in the limit as selection coefficients converge is derived in Appendix 3.2. Conveniently, this limit is independent of any selection coefficient but depends simply on the ordering of fixations of beneficial mutations. This limit gives a maximum probability for all fixation orderings in which beneficial mutations are not fixed in order from largest to smallest effect; it gives a minimum probability for the case in which fixations do appear in this order. Also derived in Appendix 3.2 is the analytical integration of (22), which gives $E(\tau_j | \zeta)$. From this derivation, the limit as selection coefficients converge is determined. This limit is dependent on the average selection coefficient of beneficial mutations. Therefore, given this average value, the trajectory of fitness expectation and variance among replicate populations may be computed as previously described.

Fitness trajectories II: parallel populations share a common distribution of beneficial mutational effects

In previous sections, we assumed that different populations evolving in identical environments would fix the same set of beneficial mutations. In this section, we assume instead that every set of beneficial mutations fixed by such parallel populations is extracted from a common parent distribution. Such an assumption might be appropriate, for example, to describe the within-host evolution of a parasite. The parasite may invade different organs of the host, giving rise to several different subpopulations. Each subpopulation may have its own set of potential adaptations, yet the entire set of potential adaptations across all subpopulations within the host would form a common parent distribution of mutational effects.

Given such a distribution, we determine the probability densities for both population fitness and time of fixation. In terms of the evolving parasite example, the expected fitness trajectory derived from these p.d.f.s may be interpreted as the average fitness trajectory of the entire parasite population within the host. The variance may be

interpreted as the fitness variance among samples taken throughout the entire parasite population. The probabilistic derivation of the fitness trajectories consists of two parts. Firstly, we determine the p.d.f. for population fitness after each fixation. Secondly, we determine the p.d.f. for the time until fixation. In this way, a complete picture of the adaptive dynamics may be drawn. For brevity, we analyze only two regions of parameter space. In the first region, the *non-ordered region*, all fixation orderings are assumed equiprobable. In the second region, the *ordered region*, beneficial mutations are fixed in order from largest to smallest effect with probability approaching one.

The non-ordered region

When populations are small and/or mutation rates are low, all orderings of fixations are nearly equiprobable under the condition that all $\mu_i \pi(s_i)$ are nearly equal. Because $\pi(s_i)$ is assumed to be a linear function, this condition may be restated to require simply that all $\mu_i s_i$ be nearly equal. This condition is met when (i) all selective

coefficients for beneficial mutations are nearly equal and the rates of mutation to all beneficial mutations are nearly equal, or (ii) large effect mutations have lower mutation rates (i.e., are rarer) than small effect mutations. When populations are small (or mutation rates are low) and the above condition is met, we refer to that region as the non-ordered region of parameter space.

Here follows a probabilistic derivation for fitness trajectories of evolving populations whose parameters belong to the non-ordered region. The selective advantage of beneficial mutations, S , is assumed to be exponentially distributed with parameter α . (Note that this distribution requires the second scenario, number (ii) above, to satisfy the condition that all orderings of fixations are nearly equiprobable.) In the non-ordered case, population fitness

is given by $W_i = 1 + \sum_{j=1}^i S_j$, where the S_j are independent,

identically distributed (i.i.d.) random variables. Thus, $W_i - 1$ is a gamma-distributed random variable with p.d.f.

$$\phi_{W_i-1}(w) = \frac{\alpha^i w^{i-1}}{(i-1)!} e^{-\alpha w} . \quad (28)$$

The expectation and variance are $E(W_i) = 1 + \frac{i}{\alpha}$ and $\text{Var}(W_i) = \frac{i}{\alpha^2}$, respectively.

The p.d.f. for time until fixation has two components. After the fixation of one mutation, a certain time, T_M , will elapse before the appearance of the next beneficial mutation. After this mutation has appeared, a certain amount of time, T_F , is required for its fixation. Thus the time until fixation of the i^{th} mutation is given by

$$T_i = \sum_{j=1}^i (T_{Mj} + T_{Fj}). \quad \text{The } T_{Mj} \text{ are i.i.d. exponential random}$$

variables; thus, the sum, $T_{Mi} = \sum_{j=1}^i T_{Mj}$, is a gamma-

distributed random variable with density

$$g_{T_{Mi}}(t) = \frac{\bar{r}^i t^{i-1}}{(i-1)!} e^{-\bar{r}t}, \quad \text{where } \bar{r} = \frac{1}{M} \sum_{j=1}^M \mu_j n(s_j) N \quad (\text{recall the}$$

condition that all $\mu_j n(s_j)$ be approximately equal). The T_{Fj}

are given by $T_{Fj} = \frac{\ln N}{S_j}$; therefore, they are i.i.d. with

density $f_{T_F}(t) = \frac{\alpha \ln N}{t^2} \exp\left\{-\frac{\alpha \ln N}{t}\right\}$. Because the random

variables are independent, the p.d.f. for their sum, $\sum_{j=1}^i T_{Fj}$,

may be computed numerically (there is no analytical solution) as the i -fold convolution of $f_{T_F}(t)$. Thus the total time until fixation of the i^{th} beneficial mutation has p.d.f.,

$$h_{T_i}(t) = g_{T_M}(t) * f_{T_F}(t)^{*i}, \quad (29)$$

where $g_{T_M}(t)$ is the gamma density with parameter i , *

denotes convolution, and the exponent $*i$ denotes i -fold convolution. Numerical solution may be facilitated by noting that (29) is equivalent to

$$h_{T_i}(t) = \mathcal{F}^{-1} \left(\mathcal{F}(g_{T_M}(t|i)) [\mathcal{F}(f_{T_F}(t))]^i \right), \text{ where } \mathcal{F} \text{ denotes}$$

Fourier transform.

The ordered region

When populations are large or mutation rates are high, mutations are guaranteed to be fixed in order from largest

to smallest selective advantage. In this case, matters are complicated by the fact that all independence among random variables is lost. Suppose that M beneficial mutations are available to an organism. If mutations must be fixed in order from largest to smallest selective advantage, then the first mutation to appear will confer a selective advantage whose p.d.f. is given by the largest order-statistic,

$$p_{S_{[1]}}(s) = M [F(s)]^{M-1} f(s) , \quad (30)$$

where $F(s)$ and $f(s)$ are, respectively, the cumulative density function (c.d.f.) and p.d.f. for the selective advantage of beneficial mutations. If we assume, as we have in the previous derivations, that selective advantage S of beneficial mutations is an exponentially distributed random variable, then (30) becomes

$$p_{S_{[1]}}(s) = M (1 - e^{-\alpha s})^{M-1} \alpha e^{-\alpha s} . \quad (31)$$

This gives the p.d.f. for W_1 , since $W_1 = 1 + S_{[1]}$. Derivation

of the p.d.f. for $W_i = 1 + \sum_{j=1}^i S_{[j]}$ does not yield the gamma

distribution (as in the non-ordered region) because the $S_{[j]}$ are not independent. In this case, the p.d.f. is found by implementing a convenient feature of the exponential distribution, namely, that $S_{[i]}$ and the differences $S_{[j+1]} - S_{[j]}$, $j=1,2,\dots,i-1$, are independent random variables. Thus, for computing the density of population fitness upon fixation of the i^{th} beneficial mutation, the following theorem obtains.

Theorem. For $i = 1, 2, \dots, M$, let $X_{[i]}$ be the order-statistics drawn from exponential parent density, $\alpha e^{-\alpha x}$, such that $X_{[1]} > X_{[2]} > \dots > X_{[M]}$. Define $z_i = \sum_{j=1}^i X_{[j]}$. Then, the probability density function of Z_i is

$$\phi_{Z_i}(z) = \begin{cases} M (1 - e^{-\alpha z})^{M-1} \alpha e^{-\alpha z} & , \quad i = 1 \\ M \binom{M-1}{i-1} \frac{\alpha^i}{i} e^{-\alpha z} \left(\frac{z^i}{i!} + \sum_{j=1}^{M-i} \binom{M-i}{j} (-1)^j \gamma \right) & , \quad i > 1 \end{cases} \quad (32)$$

$$\text{where } \gamma = (-1)^i \left(\frac{i}{\alpha j} \right)^{i-1} (1 - e^{-\alpha \frac{i}{j}}) + \sum_{k=1}^{i-2} \frac{(-1)^k}{(i-k-1)!} \left(\frac{i}{\alpha j} \right)^k z^{i-k-1}.$$

This theorem is readily proved by induction. Thus, the density for fitness after the i^{th} fixation is $\phi_{w_i-1}(w)$, as given by the theorem.

As in the non-ordered case, the time to fixation of the i^{th} beneficial mutation is given by the sum,

$$T_i = \sum_{j=1}^i (T_{Mj} + T_{Fj}). \text{ And as before, the waiting time is}$$

decomposed into two parts, (i) $\sum_{j=1}^i T_{Mj}$ and (ii) $\sum_{j=1}^i T_{Fj}$. Part

(i) is the same as in the non-ordered case. In part (ii),

the same transformation of variables, $T_{Fj} = \frac{\ln N}{S_{[j]}}$, applies;

however, the transformed variables in the ordered case are not independent. Again, the lack of independence is remedied by the convenient feature of the exponential distribution described above. The p.d.f. for T_{Fj} is

$$f_{T_{Fj}}(y_1) = \int_{\frac{i-1}{i}y_1}^{y_1} \int_{\frac{i-2}{i-1}y_2}^{y_2} \dots \int_{\frac{2}{3}y_{i-2}}^{y_{i-2}} \int_{\frac{1}{2}y_{i-1}}^{y_{i-1}} \gamma(y_1, y_2, \dots, y_i) dy_i dy_{i-1} \dots dy_2, \quad (33)$$

where, $\gamma =$

$$\frac{M!}{(M-i)!} (\alpha \ln N)^i \left(1 - \exp \left\{ \frac{-\alpha \ln N}{y_i} \right\} \right)^{M-i} \prod_{j=1}^i (y_j - y_{j+1})^{-2} \exp \left\{ \frac{-\alpha \ln N}{y_j - y_{j+1}} \right\},$$

and $y_{i+1} = 0$. As in the non-ordered case, the p.d.f. for the

sum of the two parts, $T_i = \sum_{j=1}^i (T_{Mj} + T_{Fj})$, is given by their

convolution,

$$h_{T_1}(t) = g_{T_M}(t) * f_{T_F}(t) = \mathcal{F}^{-1}(\mathcal{F}\{g_{T_M}(t)\} \mathcal{F}\{f_{T_F}(t)\}) \quad . \quad (34)$$

In this section, I have presented work in progress. It is hoped that some features of the fitness p.d.f. derived for the ordered region will enable one to distinguish it from the fitness p.d.f. derived for the non-ordered region. If such distinctive features are discovered between sums of order-statistics (ordered region) vs. sums of i.i.d. random variables (non-ordered region), it may be possible to determine whether the evolution of a meta-population is selection limited (corresponding to the ordered region) or mutation limited (corresponding to the non-ordered region), simply based upon the fitness distribution.

Discussion

Previous work by Johnson et al. (1996) explored the case in which a single beneficial mutation is available to a population. They computed the variance in mean fitness among replicate populations and discovered the existence of two distinct domains: (i) When the product of beneficial mutation rate and population number, μN , is above a critical value, adaptive evolution is repeatable; they called this the "coincident-event" domain. (ii) When μN is small, adaptive evolution is not repeatable; they called this the "isolated-event" domain. Qualitatively, we have reached the same conclusion by our discovery of the distinct *ordered* and *non-ordered* regions of parameter space, the main difference being that we have treated the case of two or more available beneficial mutations.

Assumptions of the models

The expected fitness trajectory given by equation (24) was derived by assuming that fitness jumps suddenly at each expected time of fixation of a beneficial mutation. (These

sudden jumps in fitness explain the jagged trajectories in Figures 3.3.) While this assumption is justified in Appendix 3.1 and in Lenski et al. (1991) for the expected fitness trajectory, it is not justified for the trajectory of fitness variance. The fitness variance given by equation (25) is the variance among replicate populations in which fixations always occur at their expected time. Equation (25) does not account for the variance in timing of occurrence of beneficial mutations. While this timing variance may be large, the corresponding fitness variance is typically much smaller than the fitness variance due to the different possible fixation orderings.

The expected fitness trajectory given by equation (24) also assumes that selection coefficients are additive. That is, the analysis does not allow any epistatic interactions, whereby the selection coefficients of certain mutations might change in magnitude or even sign. This absence of epistatic effects in conjunction with the finite number of available beneficial mutations allows the roughly monotonic deceleration of the fitness trajectory to its final plateau (Figure 3.3, panels A, C, E). If there were epistasis, then the trajectory could suddenly re-accelerate after, for

example, the fixation of a beneficial mutation of small effect which resulted in certain other mutations becoming much more beneficial than they were against the former genetic background. The absence of epistasis ensures the eventual convergence of all populations to the same fitness plateau, such that the variance in fitness among populations eventually goes to zero (Figure 3.3, panels B, D, F). With strong epistasis in which the sign of certain mutational effects varies according to genetic background different populations can become stuck indefinitely on fitness peaks of different heights (Wright, 1982). That is, the initial fixation order of beneficial mutations will vary among finite populations, which may then open up (or close off) alternative adaptive routes. Therefore, sustained fitness variation among replicate populations, even while they individually approach fitness plateaus, implies a rugged adaptive landscape with strong epistasis (Lenski et al. 1991; Lenski & Travisano, 1994). In future work, we intend to examine formally the dynamical consequences of epistatic interactions among beneficial (and conditionally beneficial) mutations in finite populations.

The previous section, entitled *Fitness trajectories II*, relies heavily on the assumption that the parent distribution of beneficial mutational effects is the exponential. The notion that this distribution should be some monotonic decreasing function was proposed by Fisher (1930). His claim was based on certain assumptions about both the structure and the dimensionality of fitness landscapes. More recently, Gillespie (1991) elegantly defended the use of the exponential distribution for beneficial mutational effects, basing his argument solely on the statistical theory of extreme values. Despite the sound logic of these arguments, however, very little biological data are available to test this assumption.

Estimating parameters

For ten thousand generations of bacterial evolution, Lenski et al. (e.g. Lenski & Travisano, 1994) have recorded fitnesses of twelve *E. coli* populations relative to a common ancestor. We have performed a least-squares regression of their average fitness data to a simplified variation of equation (24) (Figure 3.4). A simplification was necessary

because the number of terms in the sum of equation (24) is $M!$; hence, exploration of parameter space for moderately large M is computationally prohibitive. Instead of summing over all possible fixation orderings, our simplified equation sums only two terms: (i) the fitness trajectory in which fixations are rank-ordered times the probability of rank-ordering, and (ii) the expected fitness trajectory in which all orderings are equiprobable times the complement of the probability of rank-ordering. The expected fitness trajectory in which all orderings are equiprobable is given by:

$$E[w(t|no\ order)] = w_i, \quad \text{where } i = \{k: t \in [t_i, t_{i+1})\}, \quad (35)$$

where $w_i = 1 + i/\alpha$ is the fitness after the i^{th} fixation, and

$$t_i = \sum_{j=1}^i \left(\frac{j}{\mu \pi (1/\alpha) N} + \alpha \ln N \right) = \frac{i(i+1)}{2\mu \pi (1/\alpha) N} + i\alpha \ln N \quad \text{is the timing}$$

of the i^{th} fixation; μ is the overall beneficial mutation rate at time zero, and α is the exponential parameter. This simplification is systematically biased toward extreme unpredictability, because it assumes that fixation orderings are either perfectly rank-ordered or completely

unpredictable. This assumption neglects to favor "well-ordered" fixations (e.g. 124356) over "poorly-ordered" fixations (e.g. 516324), but instead treats both as equiprobable. In spite of this bias, however, tests of our simplified approach, for the computationally feasible case of small M , yielded a close approximation to exact solution of equation (24) when there is a moderate to strong tendency toward rank-ordering of fixations. Previous estimates of μ and α for the *E. coli* populations (Gerrish & Lenski, 1998) gave us *a priori* reason to believe that there was a strong tendency toward rank-ordering of fixations, justifying the use of our simplified approach for these populations. Least-squares regression of the fitness data from the Ara-1 population (Lenski & Travisano, 1994) gave: exponential parameter, $\alpha = 20$, beneficial mutation rate, $\mu = 5 \times 10^{-9}$, and number of available mutations, $M = 8$. Two of these parameters correspond reasonably well with previous estimates, $\alpha = 35$ and $\mu = 2 \times 10^{-9}$, from a model in which an infinite number of available beneficial mutations is assumed (Gerrish & Lenski, 1998). That the infinite-mutations model estimates a larger value for α reflects the fact that the beneficial mutation rate does not change when there is no

limit to the number of available beneficial mutations. In the present model, the beneficial mutation rate effectively decreases over time because the finite number of available beneficial mutations are slowly expended. Thus, to achieve a similar fitness increase, the average selective advantage of each mutation must be greater for the present model in which there is a finite number of available beneficial mutations. This would explain why the present model estimates a lower α (average selective advantage = $1/\alpha$). The parameter, M , has not previously been estimated.

In future work, I will develop a modification of the estimation procedure described here that will take into account all "well-ordered" fixations. I envision a routine that computes the probabilities of all fixation orderings in which zero or only two fixations do not appear in rank-order. In addition, a bootstrap routine could then sample other permutations at random. Such a routine will be used not only to estimate parameters with greater accuracy, but also to compute the corresponding trajectory of fitness variance. Comparison of this theoretical variance with observed fitness variance among *E. coli* populations could be useful in addressing the question of whether these

populations are following the same evolutionary pathway (Lenski et al., 1991).

Ordered substitutions in nature, with an application to HIV

Some recent studies with bacteria and viruses have shown that certain pairs of mutations are fixed in evolving populations in a more or less predictable order (Hall, 1988; Mittler & Lenski, 1992; Cunningham et al., 1997). Three different hypotheses can, in principle, explain this predictability. First, mutation *A* may be much more advantageous than *B*. In that case, *A* deterministically out-competes *B* if they co-occur, as is likely in a sufficiently large population. Also, *A* has a correspondingly higher probability of escaping initial loss due to drift than does *B*. Second, mutation *A* may occur at a higher rate than *B* so that, in a small population, *A* is more likely to arise and be fixed first. Third, the mutations may interact epistatically, such that *B* is advantageous only in the presence of *A*. In that case, *B* will be fixed in the population only after *A* has been fixed. The theory that we have developed here provides a mathematical framework for

evaluating the first hypothesis. Moreover, it can be readily extended to also incorporate the second hypothesis, at least for the case of two mutations. In the following paragraphs, we apply this framework to the case of resistance mutations in HIV (human immunodeficiency virus).

The effective population number of virions in an HIV-infected person has been a matter of recent debate (Mascolini, 1997; Leigh Brown & Richman, 1997). The outcome of this debate is of central importance to the understanding of *in vivo* evolution of HIV. On one side of the debate, Coffin (1995) argues that, with respect to population dynamics, *in vivo* populations of HIV are effectively infinite and therefore behave deterministically. On the other side, Leigh Brown (1997) contends that *in vivo* HIV has the stochastic behavior of a small, certainly finite population. Based on molecular data and an algorithm developed by Kuhner et al. (1995), Leigh Brown (1997) calculated that the effective population number of virions in an HIV-infected patient was on the order of one thousand. This number is five to eight orders of magnitude smaller than the actual number of virions present in an HIV-infected person (Ho et al., 1989; Piatak et al., 1993). To support

this calculation, he pointed to the fact that mutations conferring resistance to anti-retroviral drugs do not always appear in order from largest to smallest effect. He reasoned that, if the effective population number were very large as modelers had previously assumed, one would expect resistance mutations to always appear in order from largest to smallest effect. Yet, despite the sound logic of this argument, his calculation was met with some criticism (Mascolini, 1997). Here, we evaluate Leigh Brown's claim by mathematically formalizing his argument concerning the ordering of resistance mutations.

Based on known parameters for two resistance mutations (Boucher et al., 1992), the ordering of their fixation *in vivo*, and an estimate of the *in vivo* point mutation rate (Mansky & Temin, 1995), we estimate the probability of the observed ordering as a function of effective population number. (Effective population number is simply the parameter N in our previous developments.) Then, employing the double limit given by equation (17), we compute the maximum possible probability of the observed ordering.

In 4 out of 18 patients receiving AZT treatment for HIV infection, Boucher et al. (1992) observed that an inferior

resistance mutation, K70R, appeared to be fixed before a superior resistance mutation, T215Y/F.¹ (In several cases, the appearance of K70R was transient; we did not count these transient appearances as fixations.) The 50 percent inhibitory coefficients (IC_{50}) for wildtype, mutant K70R, and mutant T215Y/F are 0.006, 0.01, and 0.15 μ M AZT, respectively. Employing a saturation function, we compute selection coefficients of mutants by

$$s = \frac{IC_{50}(M) [C + IC_{50}(W)]}{IC_{50}(W) [C + IC_{50}(M)]} - 1, \text{ where } C \text{ denotes AZT concentration}$$

in vivo, W and M denote wildtype and mutant, respectively.

Using equation (15), the solid line in Figure 3.5 plots the probability of the observed fixation ordering against effective population number. It suggests that such an ordering would never occur if effective population number were greater than ~ 3000 .

1

A. Leigh Brown (personal communication) has pointed out that a caveat to our interpretation of this data is that both mutations may be present in the population prior to drug treatment. But such pre-existence (of any consequence) is unlikely if the effective population number is small. So, in this sense, our argument appears circular. Nevertheless, the observed fixation ordering could be explained (without invoking unusual epistatic interactions or extraordinarily large differences in mutation selection balance) only if the T215Y/F mutation does not pre-exist in the population.

One source of error in the above calculations is due to uncertainty in the conversion of the IC_{50} to selection coefficients. The accuracy of the resulting selection coefficient relies on (i) the accuracy of the IC_{50} , (ii) the dubious assumption that these values determined *in vitro* are the same *in vivo*, and (iii) both the accuracy and constancy of the *in vivo* concentration of AZT, C . This latter concentration is highly variable throughout the patient's body and over time, thus introducing considerable uncertainty in estimates of the selection coefficients. To circumvent this difficulty, we appeal to equation (17) for the maximum probability, $Pr\{I, S\}$, given only that $s_s > s_I$.

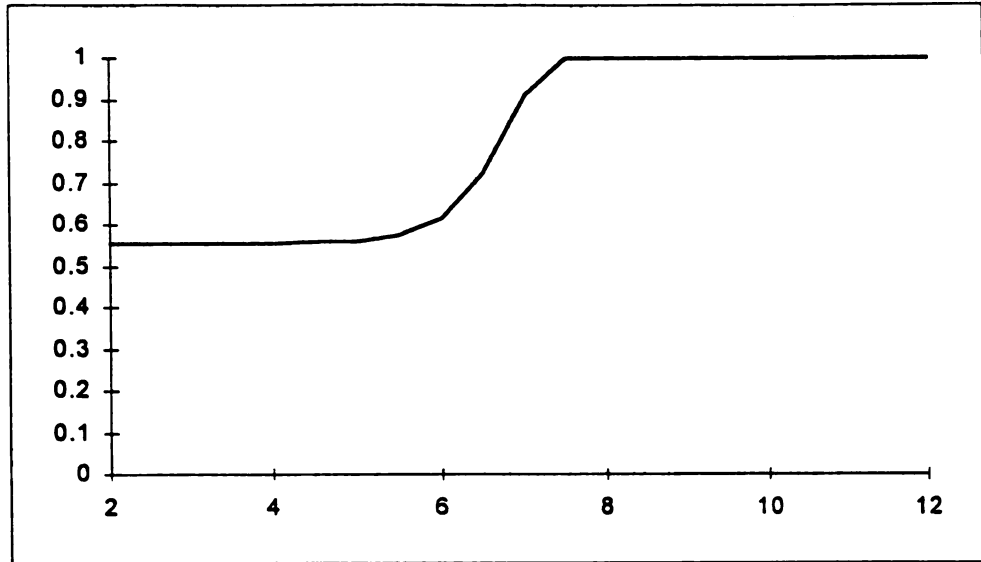
A second source of uncertainty in our calculations is introduced by the assumption of equal mutation rates. Indeed, that K70R appears first may be explained, at least in part, by invoking a higher mutation rate for K70R than for T215Y/F. This explanation may be ruled out, however, by calculating the maximum probability given by the limit as μ_{K70R} gets large; this limit is employed in equation (17).

The dotted line in Figure 3.5 employs equation (17) to plot the largest possible probability of observing K70R before T215Y/F against effective population number. By

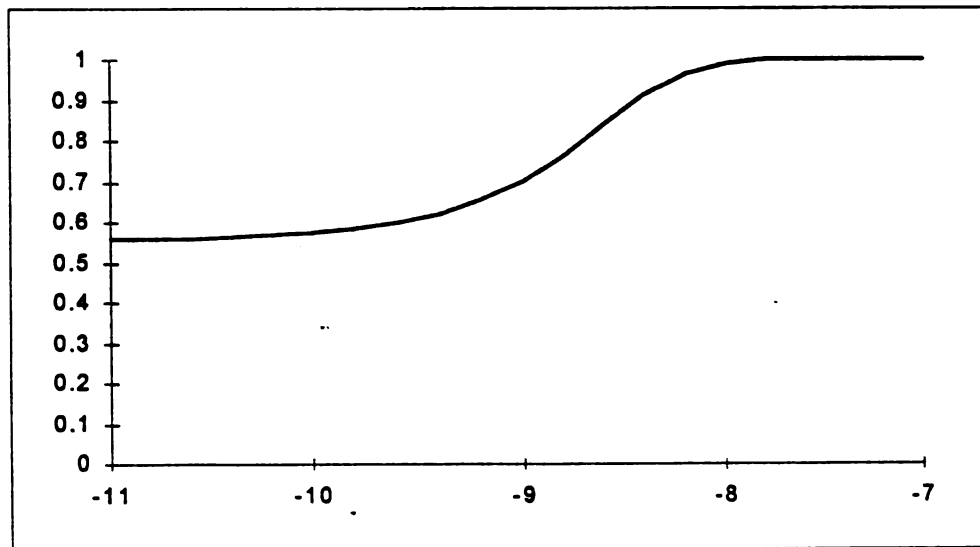
largest possible probability, we mean the maximum probability given simply that $s_{T215Y/F} > s_{K70R}$ and that μ_{K70R} is unknown. This figure demonstrates that these two simple conditions together with the observation of K70R before T215Y/F imply that the effective population number is less than about 5000. This observation is confirmed by numerically solving the inequality given by equation (18) for N .

Based on the Poisson distribution and the estimated probability (solid line of Figure 3.5), the observations of Boucher et al. statistically support the hypothesis that the effective population number is less than ~ 300 . Based on our maximum probability (dotted line), their observations support an effective population number of anything less than ~ 4000 . This number corresponds to an average production of one mutation per nucleotide per ~ 10 generations ($[4000 \text{ replications} \times (3)10^{-5} \text{ mutations per nucleotide per replication}]^{-1}$; see Mansky & Temin, 1995). Thus, our observations generally support the notion advocated by Leigh Brown that the effective population number of virions in an HIV-infected person is small enough that it must be considered finite.

probability of superior before inferior, $\Pr\{S,I\}$



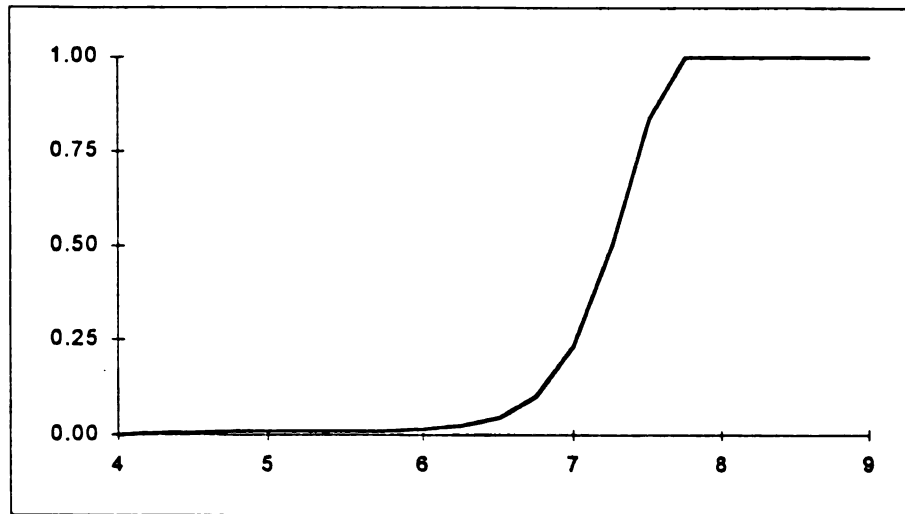
Log[population number, N]



Log[beneficial mutation rate, μ]

Figure 3.1. Probability that a superior mutation ($s_s = 0.1$) is fixed before an inferior mutation ($s_I = 0.08$) as a function of both population number, N (for which $\mu_s = \mu_I = 3 \times 10^{-9}$), and overall beneficial mutation rate, μ (for which $N = 3.3 \times 10^7$).

probability of rank-ordered fixations



Log[population number, N]

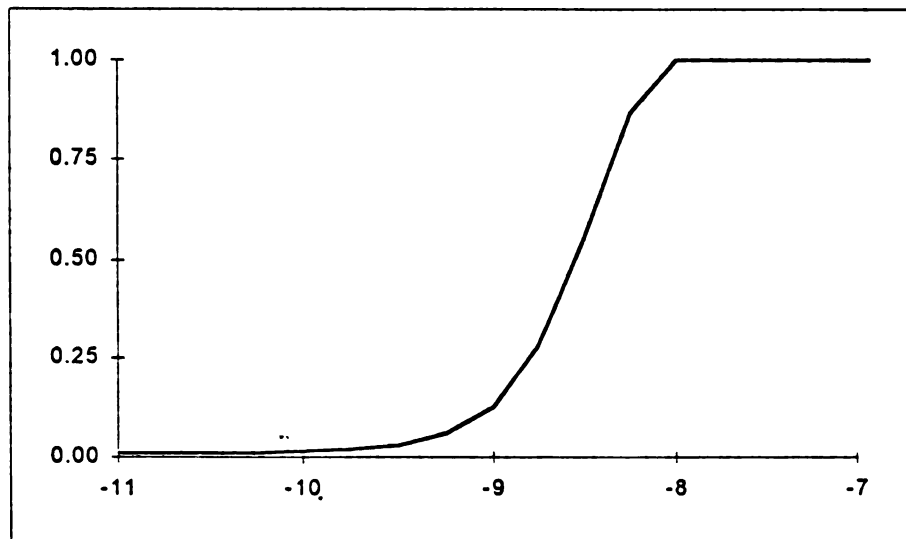

 Log[beneficial mutation rate, μ]

Figure 3.2. Probability of fixation ordering from largest to smallest selective advantage as a function of both population number, N (for which $\mu = 6 \times 10^{-9}$), and overall beneficial mutation rate, μ (for which $N = 3.3 \times 10^7$). The number of available beneficial mutations is $M = 7$, and the corresponding set of selection coefficients is $\{.08, .05, .04, .03, .02, .01, .005\}$. The individual beneficial mutation rates are $\mu_i = \mu/M$.

(A)

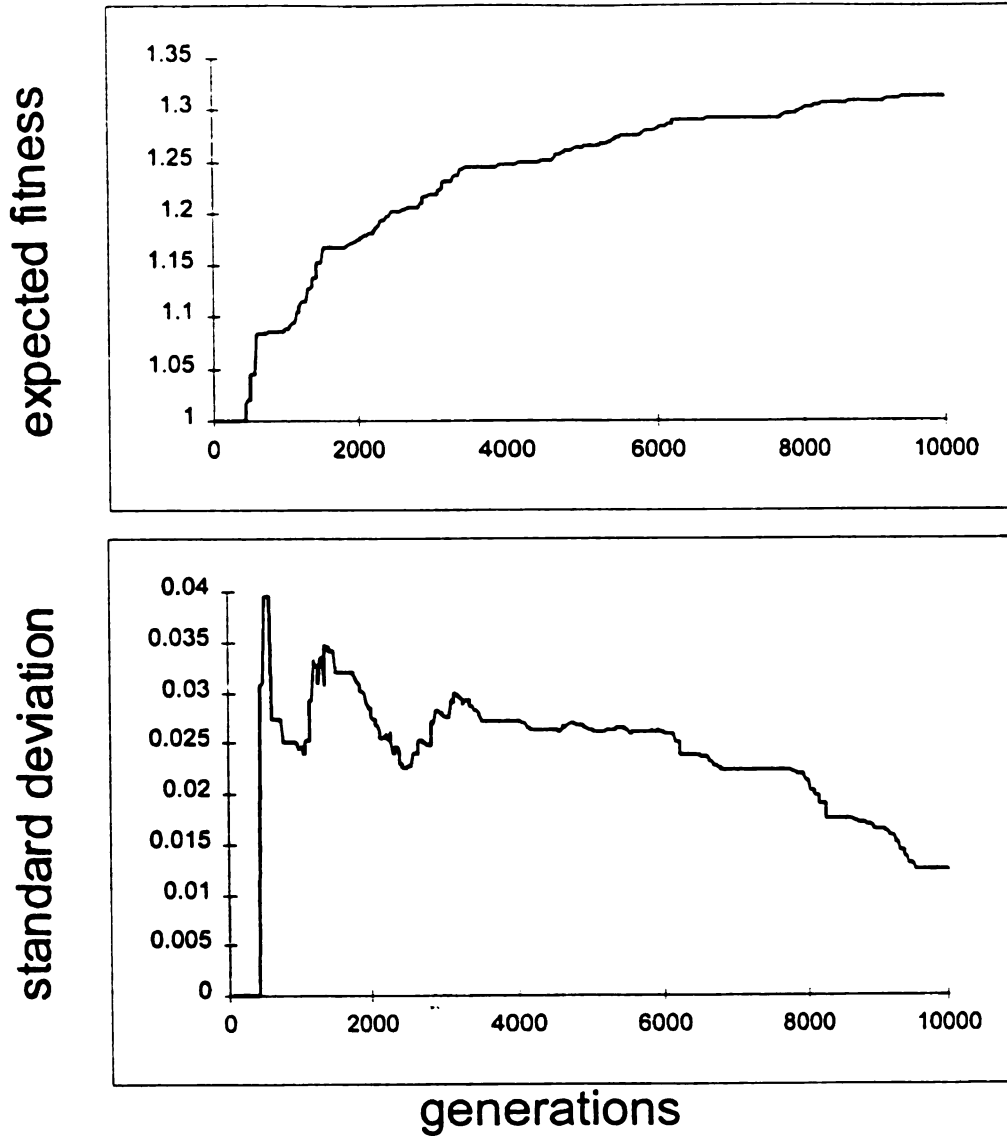
 $N = 1E6$ 

Figure 3.3. Evolutionary trajectories as a function of population size. The top and bottom panels in each pair show population fitness and its standard deviation. (A) $N = 10^6$, (B) $N = 10^7$, (C) $N = 10^8$. The number of available beneficial mutations is $M = 5$, and the corresponding set of selection coefficients is $\{.12, .08, .07, .05, .02\}$. The overall beneficial mutation rate is $\mu = 6 \times 10^{-9}$, and individual mutation rates are $\mu_i = \mu/M$.

standard deviation

expected fitness

(B)

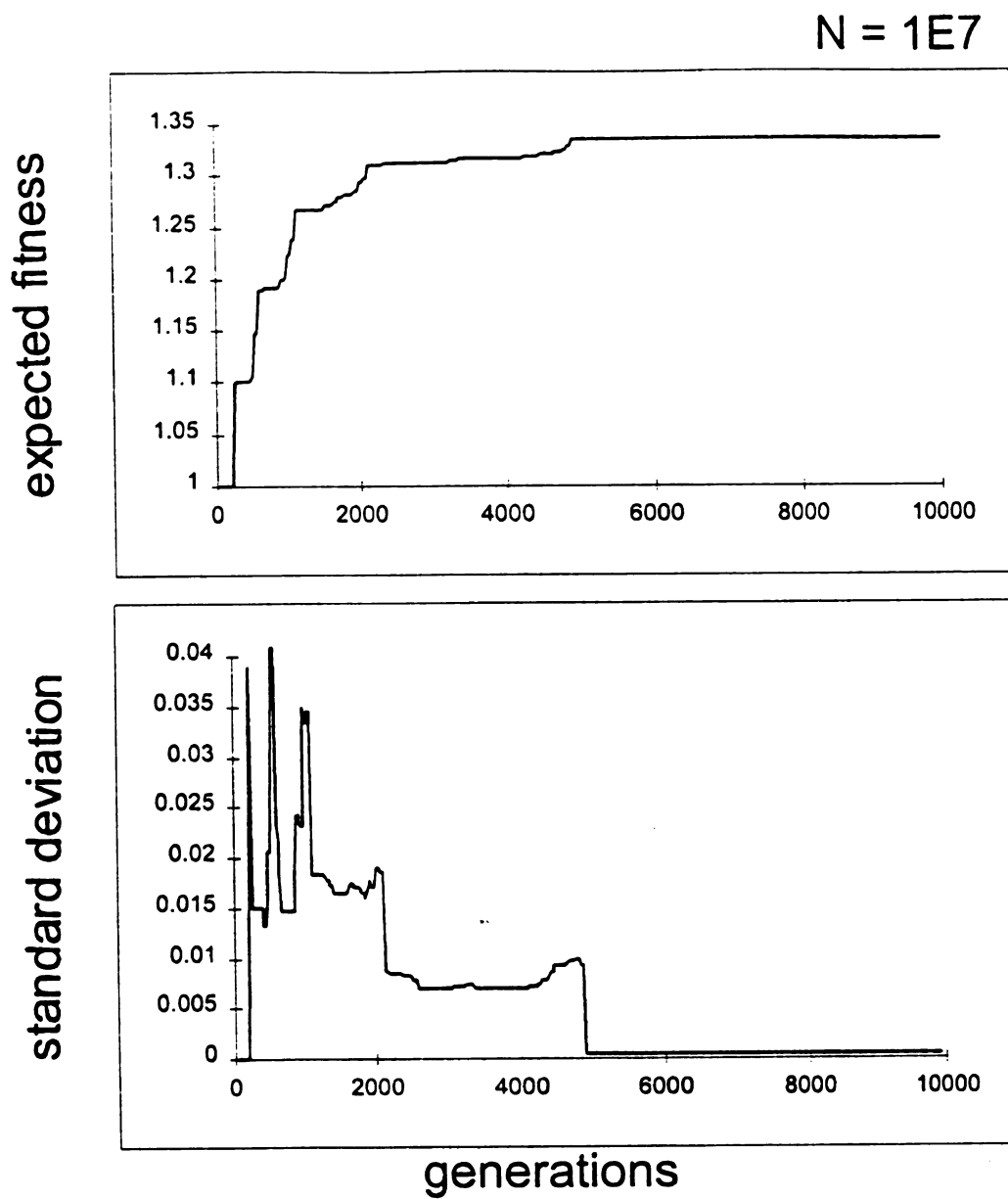
(B)

Figure 3.3 (cont.)

(c)

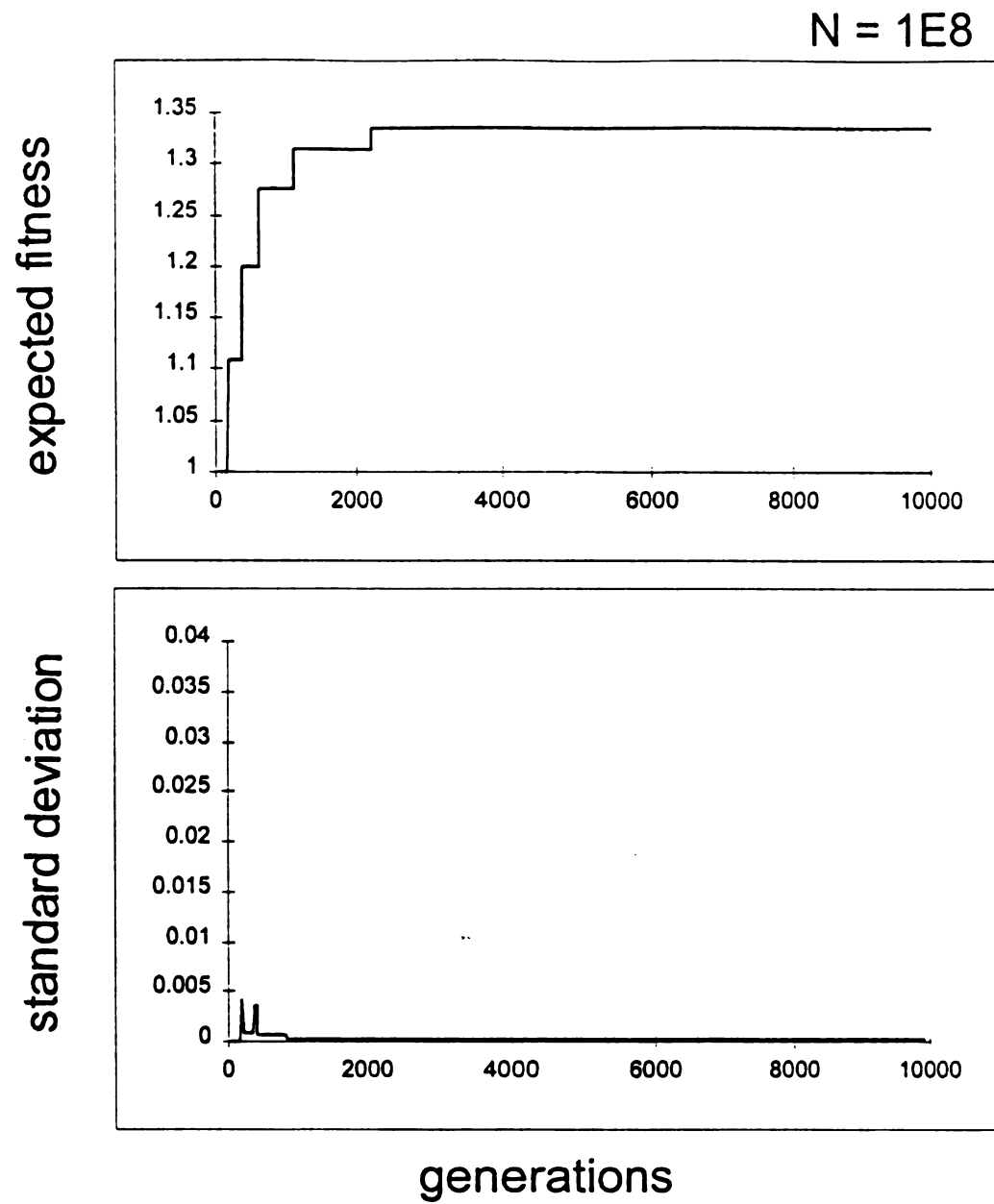


Figure 3.3 (cont.)

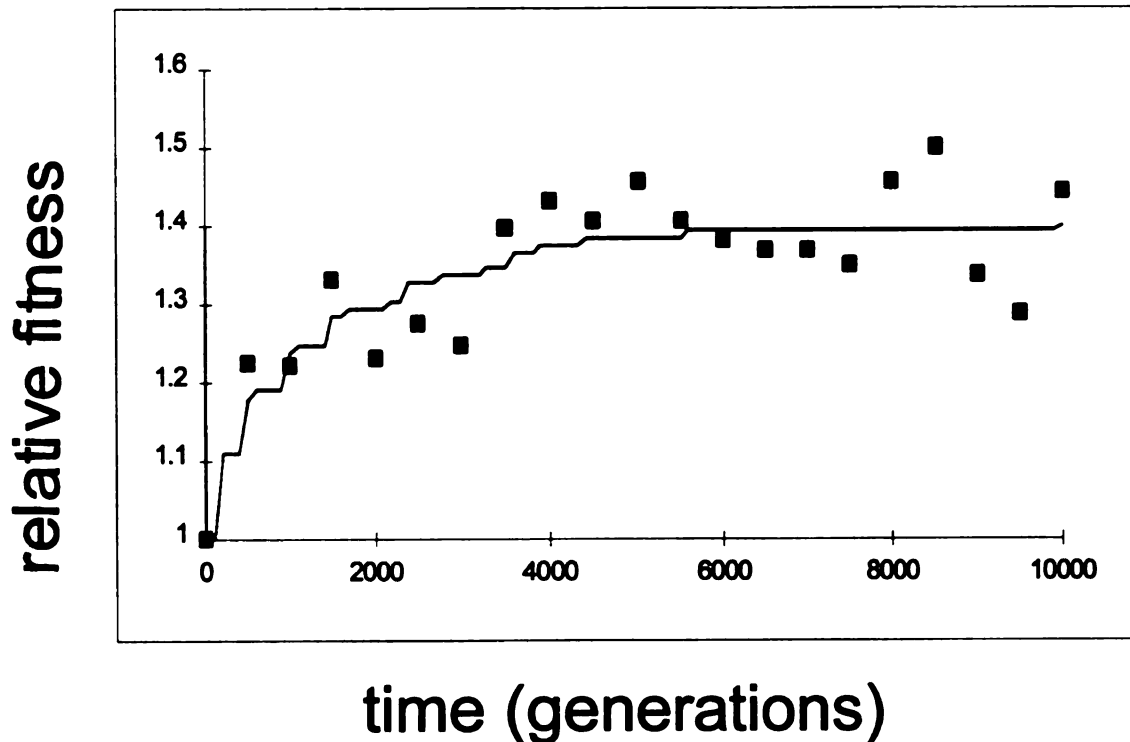


Figure 3.4. Least-squares fit of expected fitness trajectory to fitness data from evolving *E. coli* population, the Ara⁻ 1 line from Lenski & Travisano (1994). Effective population number is $N = 3.3 \times 10^7$. The generation number in the x-axis corresponds to a discrete-time model of population growth by binary fission, whereas the mathematical developments in this paper employ a continuous-time formulation. For purposes of parameter estimation, we have therefore multiplied the generation number by a factor of $\ln 2$ to adjust for this difference. Estimated parameters are $M = 8$, $\alpha = 20$, and $\mu = 5 \times 10^{-9}$.

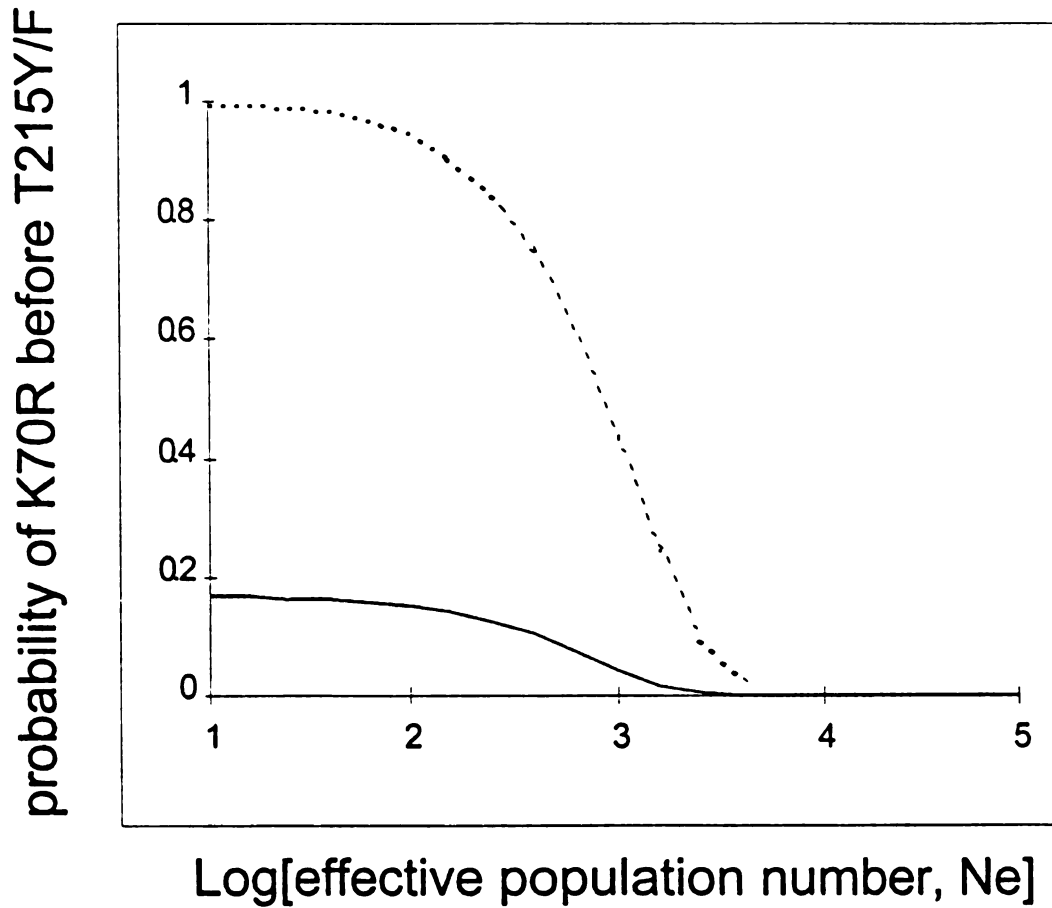


Figure 3.5. Estimated probability (solid line) and maximum probability (dotted line) that the inferior AZT-resistance mutation, K70R, is fixed before the superior mutation, T215Y/F, as a function of effective population number. We used a conservative value for the mutation rate, $\mu_{T215Y/F} = 10^{-5}$ (see Mansky & Temin, 1995). To compute estimated probabilities, we employed equation (15) and assumed (i) that the two mutations conferred *in vivo* selective advantages over the wildtype of $s_{K70R} = 0.3$ and $s_{T215Y/F} = 1.5$ (converted from IC_{50} values reported by Boucher et al., 1992), and (ii) that $\mu_{K70R} = \mu_{T215Y/F}$. To compute maximum probabilities, we employed equation (17) which relies only on the condition that $s_{T215Y/F} > s_{K70R}$ and an estimate for $\mu_{T215Y/F}$.

Chapter 4

HITCHHIKING OF DELETERIOUS MUTATIONS WITH SPECIAL ATTENTION TO MUTATOR ALLELES¹

Abstract

In asexual populations, the fate of a particular allele is determined as much by the selective values of alleles at other loci as by its own selective value. A deleterious allele may achieve high frequency or even fixation in a population as a result of linkage with a beneficial mutation. For a specified deleterious mutation, we derive the rate at which such "hitchhiking" takes place, taking into account competitive interactions among beneficial mutations. As a special case, we explore the hitchhiking of mutator alleles. We find that there exists a most probable mutator strength that is positive and typically of significant effect, suggesting a sort of "quantum" behavior in which a population's mutation rate either corresponds to the wildtype rate or is elevated significantly above that rate. This characteristic has been observed in both natural and laboratory populations. Our results indicate that hitchhiking is generally an important mechanism in the evolution of mutation rates in asexual populations.

¹

This chapter is written in the format of a paper. The "we" in this chapter refer to P.J. Gerrish, P.D. Sniegowski and R.E. Lenski.

Introduction

A neutral or deleterious mutation may rise to high frequency or even fixation in a population as a result of being linked to a beneficial mutation. This process is known as hitchhiking. Maynard Smith & Haigh (1974) and Kaplan et al. (1989) have explored the consequences of hitchhiking on heterozygosity for both selectively neutral alleles and selectively maintained polymorphisms. Among other things, they found that hitchhiking reduced average heterozygosity by an amount determined in part by the degree of linkage between the beneficial mutation and the locus in question. In asexual organisms, there is complete linkage among all loci. For the asexual case, Berg (1995) derives a stochastic model of hitchhiking in which beneficial mutations are implicitly assumed to be very rare events such that they do not interfere with each other's progression to fixation; he explores implications of this model for neutral and nearly neutral variation. In large populations, however, the fate of a given allele is affected not only by selection at linked loci but also by competition with other

fitness variants in the population, each with its own complex array of linked fitness alleles. Taking into account such competition, we explore the dynamics of hitchhiking, focusing our attention on the fate of a single specified deleterious allele. We derive the per-capita rate at which the specified deleterious allele is fixed by hitchhiking and the consequent *probability* that it is fixed within a given time interval.

In obligately asexual organisms, linkage simply means that alleles are present in the same genome. There are three ways in which a deleterious mutation and a beneficial mutation may appear on the same genome; the deleterious mutation may appear first, the beneficial mutation may occur first, or the two may occur simultaneously (in the same generation). If the beneficial mutation occurs first, then the deleterious mutation will not be fixed because it will compete with fitter organisms that carry the beneficial but not the deleterious mutation. The deleterious mutation may appear early in the growth of the beneficial mutation, in which case it will initially rise to high frequency with the beneficial mutation; however, it will subsequently decrease

in frequency and eventually stabilize at some frequency determined by the balance between mutation and selection.

In light of the above, a deleterious mutation can be *fixed* by selection only if it is in the background upon which a beneficial mutation appears. That is, the deleterious mutation must occur before, or simultaneously with, the beneficial mutation. Thus, the rate at which a specified deleterious mutation hitchhikes to fixation is reduced to the rate at which successful beneficial mutations are produced in the specified deleterious subpopulation.

Theoretical developments

The hitchhiking rate

For convenience, we refer to any beneficial mutation that is produced on a specified deleterious background (in that order) as a *double-mutation*. The frequency of a deleterious mutation in a population is increased by recurrent mutation and decreased by selection. The resulting mutation-selection balance determines the frequency of the deleterious mutation, which we denote by

ϕ . Let μ'_B denote the beneficial mutation rate in this deleterious subpopulation. Then, the per capita recruitment rate of double-mutations is $\phi \mu'_B$. We allow the beneficial mutation rate in the deleterious subpopulation to differ from the corresponding rate in the wildtype subpopulation (hence the prime), because one class of deleterious mutations that is of special interest causes a general increase in mutation rate.

Let $Pr\{fix\}$ denote the probability that any given beneficial mutation produced by the deleterious subpopulation achieves fixation in the population. Then, the per-capita rate of fixation of such double-mutations is

$$h = \phi \mu'_B Pr\{fix\} . \quad (36)$$

This is, in other words, the per-capita *hitchhiking* rate of the deleterious mutation in question. To assemble the pieces of equation (36), we first derive an effective mutation-selection balance, ϕ_e , and compare it with the commonly used equilibrium mutation-selection balance, $\hat{\phi}$. Then we derive the fixation probability, $Pr\{fix\}$, by considering that a beneficial mutation may be lost as a

result of either drift or competitive exclusion by a superior beneficial mutation. All derivations in this section are general for deleterious mutations.

An effective mutation-selection balance

The specified deleterious mutation appears at per-capita rate μ_D , and selection acts to remove this mutation at per-capita rate $s_D\phi$, where s_D is the coefficient of selection against the specified deleterious mutation. Therefore, the dynamic equation for mutation-selection balance is $\frac{d\phi}{dt} = (\mu_D - s_D\phi)(1-\phi)$, where $(1-\phi)$ imposes logistic frequency dependence and ensures that $\phi \leq 1$. Because mutation selection balance is typically a low frequency (such that $1-\phi \approx 1$), this equation is well approximated by $\frac{d\phi}{dt} = \mu_D - s_D\phi$. The equilibrium solution to this dynamic equation is $\hat{\phi} = \frac{\mu_D}{s_D}$, which is the conventional expression for mutation-selection balance.

Evolving populations, however, are not at equilibrium. Hitchhiking of deleterious alleles only occurs in adaptively evolving populations. In such populations, genetic variation is periodically purged by adaptive substitutions.

In Appendix 4.1, we derive an effective mutation-selection balance, ϕ_e , which employs the above dynamic equation and takes into account such periodic homogenizing of a population. In subsequent figures, we plot curves employing both equilibrium ($\phi = \hat{\phi}$) and effective ($\phi = \phi_e$) mutation-selection balance.

Figure 4.1 illustrates the importance of accounting for the mutation-selection dynamics. It uses mathematical developments and equations that are presented in Appendix 4.1, but the basic points can be readily understood graphically. Panel A shows, for two different values of s_D , the dynamics of a deleterious mutation as it approaches its equilibrium without any interruption. Notice that the more deleterious mutation not only has a lower equilibrium frequency, but also that it approaches that equilibrium much faster. However, the on-going substitution of beneficial mutations will perturb these dynamics away from their approach to the equilibrium, periodically re-setting the deleterious mutation frequency to near zero. Panels B and C show this effect for two different values of the substitution rate, σ , the inverse of which approximates the median time between fixation of successive beneficial

mutations. As can be seen, the discrepancy between the equilibrium (dashed line) and effective (solid line) mutation-selection balance becomes greater as the substitution rate increases. The discrepancy also becomes greater at lower values of s_d because, as shown in panel A, the equilibrium frequency is higher and the approach to equilibrium slower, leaving correspondingly more time for a given rate of adaptive substitutions to depress the frequency of the deleterious mutation substantially below its equilibrium. Thus, the discrepancy between the equilibrium and effective mutation-selection balance is a decreasing function of s_d and an increasing function of σ .

Probability of fixation of double-mutations, $\Pr\{\text{fix}\}$

If a beneficial mutation on a deleterious background is to achieve fixation, it must (i) confer an advantage that outweighs the disadvantage of the deleterious background, (ii) survive the effects of drift in the first few generations of growth, and (iii) outcompete any alternative beneficial mutations that may arise on the wildtype or deleterious backgrounds. Condition (iii) may be rephrased

to state that a beneficial mutation will be fixed only if it does not encounter an *interfering mutation*, defined as follows. If a superior mutation appears and itself survives drift, then it is called an *interfering mutation* since it interferes with, indeed prevents, the fixation of the original beneficial mutation. Because loss by drift occurs in the first few generations of growth whereas loss by *clonal interference* (competition among alternative beneficial mutations) occurs in later generations of growth, we can make the assumption that these two processes are independent. Thus, we compute the probability of fixation of a beneficial mutation on a deleterious background as the product of the probabilities that (i) this double-mutation survives drift and (ii) no interfering mutation is encountered.

Suppose a population consists of two homogeneous subpopulations, wildtype and deleterious, until the appearance of a beneficial mutation on the deleterious background. Before the beneficial mutation appears, the size of the deleterious subpopulation is determined by the dynamic balance between mutation and selection. Let $x(t)$ denote the number of wildtype individuals at time t , let

$y(t)$ denote the number of individuals carrying only the deleterious mutation, and let $z(t)$ denote the number of individuals carrying both beneficial and deleterious mutations. Let s_B and s_D denote, respectively, the selection coefficients for the beneficial mutation and against the deleterious mutation, such that, relative to the wildtype, the fitness of an individual carrying both mutations is $1 + s_B - s_D$ (i.e., both s_B and s_D assume positive values). Define t_f as the time to virtual fixation of the double mutation. Start time with the appearance of a beneficial mutation on the deleterious background, i.e., $z(0) = 1$. Then the total number of alternative beneficial mutations is the number produced on either the wildtype or deleterious background during the time interval $(0, t_f)$. On the wildtype background, this number is

$$\mu_B \int_0^{t_f} x(t) dt = \frac{\mu_B}{s_B - s_D} (1 - \phi) N \ln N , \quad (37)$$

where μ_B denotes the rate of beneficial mutation in the wildtype subpopulation, and the dynamics of $x(t)$ are assumed to be logistic, which carries the assumption of constant population size (see Crow & Kimura, 1970). Likewise, the

number of alternative beneficial mutations produced on the deleterious background between the times of appearance and fixation of the original beneficial mutation is

$$\mu'_B \int_0^t y(t) dt = \frac{\mu'_B}{s_B - s_D} \phi N \ln N . \quad (38)$$

The total expected number of alternative beneficial mutations is the sum of equations (37) and (38).

We now determine what fraction of these alternative mutations are interfering mutations. To proceed requires an assumption about how selection coefficients of beneficial mutations are distributed. For reasons discussed in Gillespie (1991, p. 262), we have chosen the exponential distribution. Let α denote the parameter of this distribution. Then, given some function, $\pi(s)$, describing the probability of surviving drift (see Appendix 4.1), a beneficial mutation chosen at random from the *wildtype* subpopulation (i) is superior to the double-mutant, and (ii) survives drift, with probability $\int_{s_B - s_D}^{\infty} \pi(s) \alpha e^{-\alpha s} ds$. When $\pi(s)$ is a linear function of s (which we assume in the following developments), this integral reduces to

$\pi(s_B - s_D + 1/\alpha) e^{-\alpha(s_B - s_D)}$. The second factor in this product is

the probability that an arbitrarily chosen beneficial mutation is superior to the double mutant in question, given the exponential distribution of mutational effects. The first factor is the expected probability that an arbitrarily chosen *superior* mutation survives drift. The probability that a beneficial mutation chosen at random from the *deleterious* subpopulation is (i) superior to the original mutation and (ii) survives drift is $\int_{s_B}^{\infty} \pi(s-s_D) \alpha e^{-\alpha s} ds$, or $\pi(s_B-s_D+1/\alpha) e^{-\alpha s_B}$ when $\pi(s)$ is linear. Note the change in the lower integration limit from s_B-s_D for the wildtype subpopulation to s_B for the deleterious subpopulation. This is because a mutation occurring on the deleterious background must have a selection coefficient greater than s_B if it is to be superior to the double-mutation, whereas a mutation occurring on the wildtype background needs a selection coefficient only greater than $s_B - s_D$.

Interfering mutations are those which (i) occur in the interval $(0, t_f)$, (ii) survive the effects of drift, and (iii) are superior to the double-mutation. Therefore, the number of interfering mutations produced by the wildtype subpopulation is

$$\lambda = \frac{\mu_B}{s_B - s_D} (1 - \phi) N \ln(N) \pi(s_B - s_D + 1/\alpha) e^{-\alpha(s_B - s_D)} . \quad (39)$$

Likewise, the number of interfering mutations produced by the deleterious subpopulation is

$$\lambda' = \frac{\mu'_B}{s_B - s_D} \phi N \ln(N) \pi(s_B - s_D + 1/\alpha) e^{-\alpha s_B} . \quad (40)$$

The total number of interfering mutations in the population, therefore, is simply the sum $\lambda + \lambda'$. The number of interfering mutations is Poisson distributed, so the conditional probability of fixation of a beneficial mutation with net selective advantage, $s_B - s_D > 0$, is equal to the probability that (i) it survives drift and (ii) no interfering mutation appears:

$$Pr\{fix|s_B\} = \pi(s_B - s_D) e^{-(\lambda + \lambda')} . \quad (41)$$

Given our assumption that s_B is exponentially distributed with parameter α , the expected probability that a beneficial mutation on the deleterious background achieves fixation is

$$Pr\{fix\} = \alpha \int_{s_D}^{\infty} \pi(s_B - s_D) e^{-(\lambda + \lambda') - \alpha s_B} ds_B . \quad (42)$$

The lower limit of integration reflects the obvious condition that a beneficial mutation is defined as one that has a net selective advantage, i.e., $s_B - s_D > 0$.

Substituting equation (42) into (36) gives the final expression for the per capita hitchhiking rate of a deleterious mutation:

$$h = \phi \mu'_B \alpha \int_{s_D}^{\infty} \pi(s_B - s_D) e^{-(\lambda + \lambda') - \alpha s_B} ds_B . \quad (43)$$

Converting per-capita rate to populational probability

To make sense of a per-capita hitchhiking rate, it is helpful to convert it into the corresponding probability that a hitchhiking event occurs in a population within a given time interval. A hitchhiking event is said to have occurred in a population if the specified deleterious mutation (i) has produced a successful beneficial mutation, and (ii) has achieved fixation as a result of linkage with this beneficial mutation. Given per-capita hitchhiking rate, h , the populational probability that the specified deleterious mutation produces a successful beneficial

mutation in the interval $(0, T)$ is $1 - e^{-hNT}$. After the successful beneficial mutation is produced, a certain time, T_f , elapses until the resulting double mutation is fixed. Only then has the hitchhiking event taken place. Thus, the populational probability that the specified deleterious mutation hitchhikes to fixation in the interval $(0, T)$ is

$$Pr\{hitchhike|T\} = \begin{cases} 1 - e^{-hN(T-T_f)} & , T \geq T_f \\ 0 & , T < T_f \end{cases} . \quad (44)$$

This equation may be understood as the complement of the probability that zero hitchhiking events take place by time T . In other words, it is the probability that one or more hitchhiking events take place in the interval $(0, T)$, thus accounting for the redundancy of multiple hitchhiking events. (We assume that once the mutation is fixed, it remains fixed.)

The time required for fixation of a double mutation depends on the selective advantage of the successful beneficial mutation. While this value is different for each successful beneficial mutation, its expected value may be determined following the logic of Gerrish & Lenski (1998). They reason that "a beneficial mutation whose selective

advantage is small is not likely to become fixed because it must compete with many superior mutations. On the other hand, a beneficial mutation whose advantage is large is less likely to be produced. Hence, there must be some intermediate selection coefficient that balances the fixation advantage of large s with the more frequent occurrence of small s . This balance corresponds to the expected selection coefficient of successful mutations."

Let $p(s_B) = K\pi(s_B - s_D) e^{-(\lambda + \lambda') - \alpha s_B}$, where K is a constant

such that $\int_{s_D}^{\infty} p(s_B) ds_B = 1$. Given that a beneficial mutation (i) is in linkage with the specified deleterious mutation and (ii) achieves fixation in the population, then its selective advantage, s_B , has probability density $p(s_B)$. Thus, the expected value for the selection coefficient of successful double mutations is

$$E(s_B - s_D) = \int_{s_D}^{\infty} s_B p(s_B) ds_B - s_D . \quad (45)$$

This may also be interpreted as the expected increase in population mean fitness during a hitchhiking event. Hence, to a first approximation, the time required for the fixation

of a successful double mutation is $T_f \approx \frac{\ln N}{E(s_B - s_D)}$, thereby completing equation (44).

Some of the subsequent figures plot both per-capita hitchhiking rates as well as populational probabilities of hitchhiking. The benefit of plotting rates is to avoid dependence on the time interval, and the benefit of plotting per-capita rates is to avoid dependence on population size. Elimination of these two dependencies facilitates interpretation of the observed trends in terms of the processes involved, yet it somewhat obfuscates the consequences of these processes at the population level. Thus, in order to elucidate these more intuitively accessible consequences, subsequent figures also plot populational probabilities.

An application

Hitchhiking rate of mutator mutations

A class of deleterious mutations that is of particular interest, especially in light of Chapter 1, is the class known as mutator mutations. These are mutations that

disrupt DNA synthesis and repair, and thereby elevate the mutation rate of the organism. Both experimental and theoretical evidence (Sniegowski et al., 1997; Taddei et al., 1997) suggests that the evolution of mutation rates in asexual populations may be determined more by chance hitchhiking of mutators than by the evolutionary "fine-tuning" of mutation rates proposed in previous work (Leigh, 1970; Ishii et al., 1989).

Before proceeding, a clarification of terminology is essential. Two subpopulations are of interest, wildtype and mutator; what we previously called the deleterious subpopulation, we now call the mutator subpopulation. Whereas previous reference to deleterious mutation implied a specific mutation, the following developments make reference to an *overall* deleterious mutation rate. A genotype's overall deleterious mutation rate is the total rate at which all deleterious mutations are produced. Define μ_o to be the overall deleterious mutation rate in the wildtype subpopulation; this is to be contrasted with μ_d which is now defined as the rate of mutation to the mutator genotype.

A mutator mutation is deleterious most of the time. The overwhelming majority of mutations are either neutral or

deleterious. We assume that a mutator allele elevates the mutation rate of all mutations by the same factor such that $\mu'_B = m\mu_B$ and $\mu'_\delta = m\mu_\delta$, where m is called the mutator strength. Because the wildtype deleterious mutation rate, μ_δ , is typically several orders of magnitude greater than the wildtype beneficial mutation rate, μ_B , a general increase in the mutation rate increases the total number of deleterious mutations produced much more than the total number of beneficial mutations produced. Therefore, an increased mutation rate confers reduced fitness. We now calculate how much fitness is expected to decline for a given increase in the mutation rate based on the increased production of deleterious mutations.

The number of mutations occurring per generation is Poisson distributed, so that the fraction of the subpopulation in which zero deleterious mutations occur in a single generation is $e^{-\mu_\delta}$. This is equivalent to the absolute fitness of the wildtype subpopulation, because each deleterious mutation is ultimately a genetic death in a large asexual population. That is, in a large asexual population, each deleterious mutation is eliminated sooner or later by selection (except for the occasional deleterious

mutation that hitchhikes to fixation), irrespective of the form of the fitness function.¹ Likewise, the absolute fitness of the mutator subpopulation is $e^{-m\mu_\delta}$, where $m\mu_\delta$ is the overall deleterious mutation rate of the mutator. Therefore, the selection coefficient against the mutator is

$$s_D = 1 - \frac{e^{-m\mu_\delta}}{e^{-\mu_\delta}} \approx \mu_\delta(m-1). \quad \text{This definition of } s_D \text{ may now be}$$

inserted into equation (43). The resulting expression gives the hitchhiking rate of a mutator.

Estimating rate of mutation to mutator for evolving E. coli populations

As described in Sniegowski et al. (1997) and in Chapter 1, three of twelve evolving *E. coli* populations fixed a mutator mutation within 10,000 generations. The times of

¹This is most easily understood by first computing the average fitness of an individual in which all deleterious mutations confer the same selective disadvantage. If deleterious mutation rate is μ_1 and the selection coefficient against this class of mutations is s_1 , then expected fitness of an individual in the population is

$E(w_1) = e^{-\mu_1/s_1} + (1-s_1)(\mu_1/s_1)e^{-\mu_1/s_1} + (1-s_1)^2(\mu_1/2s_1)e^{-\mu_1/s_1} + \dots = e^{-\mu_1}$. If there are M s -classes of deleterious mutations, then the expected fitness of an individual in the population is $e^{-\mu_\delta}$, where $\mu_\delta = \sum_{i=1}^M \mu_i$.

these three fixations are known; they are approximately $t_1 = 2000$, $t_2 = 3000$, and $t_3 = 8000$ generations. The censored likelihood function for the mutator hitchhiking rate is $(e^{-h(10,000)})^9 h^3 \exp\{-h \sum_{i=1}^3 t_i\}$. Note that mutator hitchhiking rate here refers to the rate at which the *population* fixes a mutator by hitchhiking (i.e., it is not a per-capita rate).

From this, the maximum likelihood estimate of the mutator hitchhiking rate for the *E. coli* populations is

$h_{obs} = 3 (90,000 + \sum_{i=1}^3 t_i)^{-1} = 2.9 \times 10^{-5}$ hitchhiking events per generation. Variance was determined from the likelihood

function to be $Var(h) = \frac{h_{obs}^2}{3} = 2.8 \times 10^{-10}$. Assuming a normal

density truncated at zero and normalized, the approximate 95% confidence interval for the hitchhiking rate is $(3.8 \times 10^{-6}, 6.2 \times 10^{-5})$.

All parameters of our model have previously been estimated for *E. coli* populations, except the rate of mutation to mutator, μ_D . These parameters are $\alpha = 35$ and $\mu_B = 2.0 \times 10^{-9}$ (Gerrish & Lenski, 1998), $\mu_\delta = .0003$ (Kibota & Lynch, 1996), $m = 100$ (Sniegowski et al., 1997), and $N = N_e = 3.3 \times 10^7$ (Lenski et al., 1991). The rate of mutation to

mutator may now be determined, given the estimate of the hitchhiking rate. This estimation procedure is facilitated by noting that $\lambda + \lambda'$ from (42) is well approximated by

$$\frac{\mu_B N \ln N}{s_B - s_D} \pi(s_B - s_D + 1/\alpha) e^{-\alpha(s_B - s_D)}, \text{ which is independent of } \mu_D.$$

(This approximation is due to the fact that mutation-selection balance, ϕ , is typically quite low.) Thus, μ_D may be estimated explicitly by inserting $\phi = \phi_e$ (given by (84) in Appendix 4.1) into (36) and rearranging:

$$\mu_D \approx \frac{h_{obs} s_D}{N \mu'_B Pr\{fix\}} \left(1 - \frac{1 - e^{-s_D/\sigma}}{s_D/\sigma} \right)^{-1}, \text{ where } \sigma \text{ is the rate of}$$

beneficial substitutions (equation (88) in Appendix 4.1), and $Pr\{fix\}$ is given by (42). The rate of mutation to the mutator type is then estimated to be $\mu_{D_{obs}} = 2.0 \times 10^{-6}$ mutator mutations produced per replication. Based solely on the variance in the hitchhiking rate, the actual rate of mutation to mutator may be assumed to be in the interval $(2.6 \times 10^{-7}, 4.3 \times 10^{-6})$ with 95% confidence. This estimate is quite reasonable, in view of the following considerations. There are four loci (*mutS*, *mutL*, *mutH*, and

uvrD) that together encode the methyl-directed mismatch repair pathway. Mutations causing defects in any of these genes will yield a mutator with strength $m \approx 100$, and all three of the mutators found in the experiments by Sniegowski et al. (1997) mapped to this pathway. Assuming ~1000 base-pairs for each locus as well as a mutation rate for *E. coli* of $\sim 5 \times 10^{-10}$ per base-pair per replication (Drake, 1991), and estimating that perhaps 25% of all mutations cause a major loss of protein function, one would predict an overall mutation rate for this repair pathway of about $4 \times 1000 \times 5 \times 10^{-10} \times 0.25 = 5 \times 10^{-7}$. This calculation lies within the confidence limits of our estimate for μ_D based on the observed dynamics of the hitchhiking process.

Discussion

Assumptions

An assumption of the above developments is that the number of beneficial mutations is unlimited, such that their rate (μ_B) and distribution of effects (α) are constant. This simplification renders our developments best suited to

populations in slowly changing environments in which the potential for adaptive improvement is renewed at a rate roughly equal to the rate of adaptation. In what follows, we compare our results to the numerical simulations of Taddei et al. (1997) and to the experimental results of Sniegowski et al. (1997). The assumption of unlimited mutations, however, is not appropriate for either of these comparisons. To evaluate sensitivity to this assumption, we are presently working on the analytical derivation for the case when the number of possible beneficial mutations is finite.

We have also assumed that a beneficial mutation appears either on the wildtype background or on the background carrying the specified deleterious mutation. In reality, a population is composed of several deleterious mutations; thus, a beneficial mutation may appear on the background of any number of deleterious mutations, thereby effectively reducing the fitness of what we have called the "wildtype" background. Nevertheless, our assumption may be justified by considering that (i) mutation-selection balance is quite low for strongly deleterious mutations, meaning that the beneficial mutation recruitment rate in that subpopulation

is very small, and (ii) mutation-selection balance may be high for weakly deleterious mutations, but these will hardly affect the outcome of a linked beneficial mutation.

As previously mentioned, the assumption that selection coefficients of beneficial mutations are exponentially distributed is nicely defended both by Fisher (1930) and by Gillespie (1991). Their arguments, of course, do not absolve this assumption of criticism, but they offer logical support to the choice of the exponential when the shape of the distribution is unknown. Lastly, we have assumed that there are no gene interactions between the deleterious mutation in question and any beneficial mutation. The validity of this assumption is a purely biological question and appears to be highly contingent on the particulars of the mutations in question.

Comparison with simulation

Taddei et al. (1997) simulated populations that grew exponentially from 10^8 to 10^{10} individuals, of which 10^8 were then sampled, with the process repeated to mimic a serial

transfer regime. For comparison with their simulations, we employ the fixation effective population size, which we calculate as $N_e \approx \frac{1}{2}N_0\tau$ (Appendix 4.2), where N_0 is the number transferred (10^8), and τ is the number of generations between transfers ($\log_2 100$). We have determined that the set of beneficial mutations they used approximates an exponential distribution with parameter $\alpha = 98$ (maximum likelihood estimate). Also, for comparison, we employ the equilibrium approximation for mutation-selection balance, $\phi \approx \hat{\phi}$; for comparison, we do not employ our *effective mutation-selection balance* because their simulations do not appear to take into account the periodic purging of genetic variation caused by substitutions.² Inserting their simulation parameters ($\mu_s = 10^{-4}$, $\mu_B = 10^{-8}$, and $\mu_D = 5 \times 10^{-7}$) into our analytical model, we computed the probability that such a population would fix a mutator by hitchhiking within a time span of 20,000 generations. Out of 100 simulations, Taddei et al. found that 19 fixed a mutator in 20,000

²This is evidenced by their Figure 3 in which mutator frequencies never drop much below mutation-selection balance. Given parameters used by Taddei et al., their Figure 3 should show occasional downward spikes of between one and two orders of magnitude if the variation-purging effect of substitutions in the wildtype subpopulation were allowed by their simulations.

generations when $m = 10$, and 7 fixed a mutator when $m = 100$. Employing their assumption of mutation-selection equilibrium ($\phi = \hat{\phi}$), our model gives probabilities of 0.223 and 0.073 for $m = 10$ and $m = 100$, respectively, which shows good agreement with their findings. However, when we correctly account for the variation-purging effect of substitutions by employing the *effective mutation-selection balance* ($\phi = \phi_e$), probabilities are 0.040 and 0.057 for $m = 10$ and $m = 100$, respectively. We restate that the fundamental differences between our analytical model and the simulations are: (i) the analytical model assumes an unlimited number of available beneficial mutations whereas the simulations assume that number to be finite, and (ii) the simulations assume mutation-selection equilibrium whereas our analytical model takes into account the dynamics of mutation-selection balance.

Effect of population size on hitchhiking rate

Panel A of Figure 4.2 plots per-capita hitchhiking rate, h , against population number, N , for a mutator; it shows that h remains constant below about $N = 10^6$ and decreases steadily above that value. This figure elucidates the effect of increased clonal interference on per-capita hitchhiking rate. In small populations, clonal interference is highly improbable. Thus, the principal determinant of h in small populations is the per-capita probability of occurrence of a beneficial mutation on the specified deleterious background. Because it is per-capita, this probability is independent of N , explaining why h is essentially independent of N below a certain population size. Above this population size, the decline in h reflects the fact that clonal interference is probable and intensifies with increasing N , thus reducing the probability that any given beneficial mutation (including those occurring on the specified deleterious background) will achieve fixation.

Panel B of Figure 4.2 plots the corresponding probability that a *population* fixes the specified

deleterious mutation by hitchhiking within a time interval of 10,000 generations against population size, N .

Hitchhiking probability increases with population number but this increase decelerates at large N because of intensifying clonal interference.

Effect of selective disadvantage on hitchhiking rate

Panel A of Figure 4.3 shows the per capita hitchhiking rate as a function of the strength of selection against the specified deleterious mutation, s_d . Interestingly, this rate is essentially independent of s_d when it is small. This is most easily understood by considering that the initial rise in the deleterious mutation's frequency after a substitution is determined much more by recurrent mutation than by selection against the mutation. It is not until the frequency of the mutation approaches its equilibrium value that selection begins to significantly affect its trajectory. Because very slightly deleterious mutations have relatively high equilibrium frequencies, there is a good chance that a substitution in the population will occur before selection becomes important to the trajectories of

such mutations. In other words, in the time interval between substitutions, slightly deleterious mutations may behave as neutral mutations, in which case their probability of hitchhiking is essentially that of a neutral mutation. (In the population genetic language, substitutions reduce the effective population size, N_e , and any mutation whose selective disadvantage is less than N_e^{-1} , in a haploid population, is effectively neutral.)

Panel B of Figure 4.3 plots the corresponding probability that a population fixes the specified deleterious mutation by hitchhiking within a time interval of 10,000 generations as a function of s_d . A trend similar to that observed in panel A is observed.

Effect of mutator strength on hitchhiking rate

Panel A of Figure 4.4 shows how the strength of a mutator affects its own per-capita hitchhiking rate. Recall that a mutator allele elevates the general mutation rate, which includes the beneficial mutation rate, by a factor, m (i.e., $\mu'_B = m \mu_B$). Thus, equation (36) becomes $h = m \mu_B \phi \Pr\{\text{fix}\}$, where μ_B denotes wildtype beneficial

mutation rate. This equation is the product of (i) the per-capita recruitment rate, $m \mu_b \phi$, of double-mutations, and (ii) the probability of fixation, $Pr\{fix\}$, of a double-mutation. Factor (i) is plotted in Figure 4.5, and factor (ii) is plotted in Figure 4.6. Thus, Figure 4.4A, which plots hitchhiking rate against mutator strength, may be understood simply as the product of Figures 4.5 and 4.6. Our aim in the next several paragraphs is therefore to explain the trends in Figures 4.5 and 4.6 in order to understand Figure 4.4.

We begin by explaining the dashed line in Figure 4.5. This shows that the invalid equilibrium assumption ($\phi = \hat{\phi}$) renders the double-mutation recruitment rate essentially independent of mutator strength over the region of biological interest. This may be understood as follows. The rate at which a mutator subpopulation produces beneficial mutations is directly proportional to the number of individuals carrying the mutator allele and hence to the mutation-selection balance of that allele. As derived in the subsection entitled *Hitchhiking rate of mutators*, the coefficient of selection against a mutator is $1 - e^{-\mu_b(m-1)}$. Thus, if deleterious mutations were always maintained at

their equilibrium frequency, $\phi = \hat{\phi} = \frac{\mu_D}{s_D}$, then the population would produce beneficial mutations on the mutator background at a per-capita rate of $m \mu_B \mu_D \left(1 - e^{-\mu_D(1-m)}\right)^{-1} \approx \frac{\mu_B \mu_D}{\mu_\delta}$. In words, a weaker mutator allows for higher equilibrium mutation-selection balance but has a lower per capita mutation rate. These two factors influencing the double-mutation recruitment rate, $m \mu_B \hat{\phi}$, directly cancel each other out, rendering this rate essentially independent of mutator strength. Note that this rate is also independent of population size, N .

When the dynamics of mutation-selection balance are properly accounted for by employing $\phi = \phi_e$, then the double-mutation recruitment rate is not independent of mutator strength. In fact, the solid line in Figure 4.5 shows this rate to be a monotonically increasing function of mutator strength. This is explained as follows. After an adaptive substitution takes place, the frequency of mutators in the population is very low. It takes some time for a mutator allele to approach its equilibrium frequency. And the higher that equilibrium frequency, the longer it takes to approach (Figure 4.1A). This can be seen in the dynamic solution to the mutation-selection balance equation,

$\frac{\mu_D}{s_D} (1 - e^{-s_D t})$. Weak mutators have small s_D and therefore require long times to approach their high equilibrium frequency. If adaptive substitutions occur in the population at a given rate, then mutator frequency drops with a certain periodicity determined by this rate. Since stronger mutators recuperate their lower equilibrium frequency more rapidly, their effective mutation-selection balance is less affected by the periodic purging of genetic variation caused by adaptive substitutions. Thus, their effective recruitment rate of double-mutations is higher. Hence the trend of monotonic increasing recruitment rate with increasing mutator strength, as shown by the solid line in Figure 4.5.

Figure 4.6 shows that probability of fixation, $Pr\{fix\}$, decreases monotonically with mutator strength. The stronger the mutator, the more deleterious it is. Therefore, the net fitness of a beneficial mutation that is linked to a mutator is lower for stronger mutators. Hence, the probability of fixation of such double-mutations decreases monotonically with increasing mutator strength.

Figure 4.4 shows that, when the dynamics of mutation-selection balance are properly accounted for, there exists

an intermediate mutator strength that maximizes the hitchhiking rate. (Panel B of Figure 4.4 reveals that the same trend is observed at the population level.) This observation reflects a balance between the increased double-mutation recruitment rate of stronger mutators (Figure 4.5) which *increases* hitchhiking rate, with the decreased probability of fixation of stronger mutators (Figure 4.6) which *decreases* hitchhiking rate.

Observations of bacterial populations in nature and in laboratories seem to suggest that mutation rates either correspond to a wildtype rate or they are elevated from this rate by one, two or occasionally three orders of magnitude (Sniegowski et al., 1997; LeClerc et al., 1996; Mao et al., 1997). Of course, this observation is complicated by the fact that strong mutators are more likely to be noticed and their mutation rate is more likely to be statistically distinguishable from the wildtype. Yet, it seems that the strength of the evidence outweighs this complication. Fragility of the genetic mechanisms involved in DNA synthesis, proofreading and repair may offer one explanation for this observation (Cox, 1976; Miller, 1996). Our work suggests another explanation which is based solely on the

adaptive dynamics of asexual populations. If a mutator is to be observed, it must achieve high frequency in the population. Hitchhiking is a plausible mechanism by which a mutator may achieve high frequency (Sniegowski et al., 1997). Our work shows that a mutator's hitchhiking rate is maximized when its strength is at some intermediate value. Therefore, given that a mutator is observed, it is most likely to be a mutator of intermediate strength.

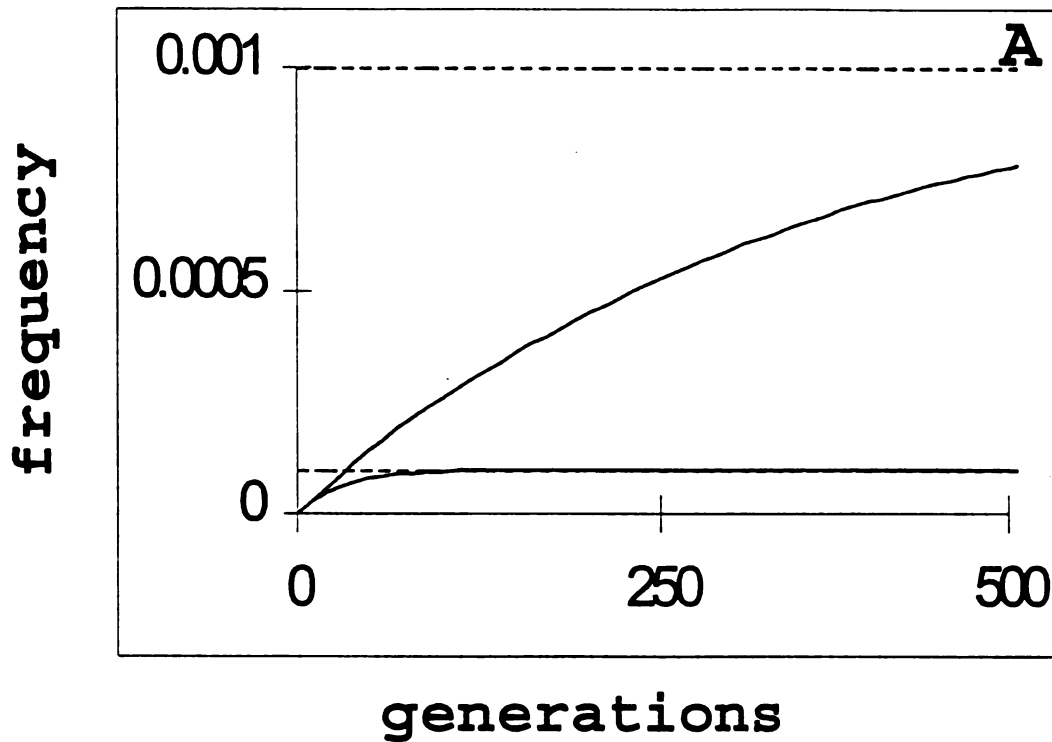


Figure 4.1. The dynamics of mutation-selection balance. The two solid lines in panel A show the frequency trajectories of two selectively distinct deleterious mutations, starting with a frequency of zero, as after an adaptive substitution, and asymptotically approaching their equilibrium frequency. (These lines plot the dynamic equation, $\phi(t) = \frac{\mu_D}{s_D} (1 - e^{-s_D t})$). Dashed lines show the corresponding equilibrium frequencies ($\hat{\phi} = \frac{\mu_D}{s_D}$). Parameters used are $\mu_D = 3 \times 10^{-6}$, $s_D = 0.03$ and 0.003 .

frequency

Fig
tra
ada
0.0
det
 α =
lin
cal

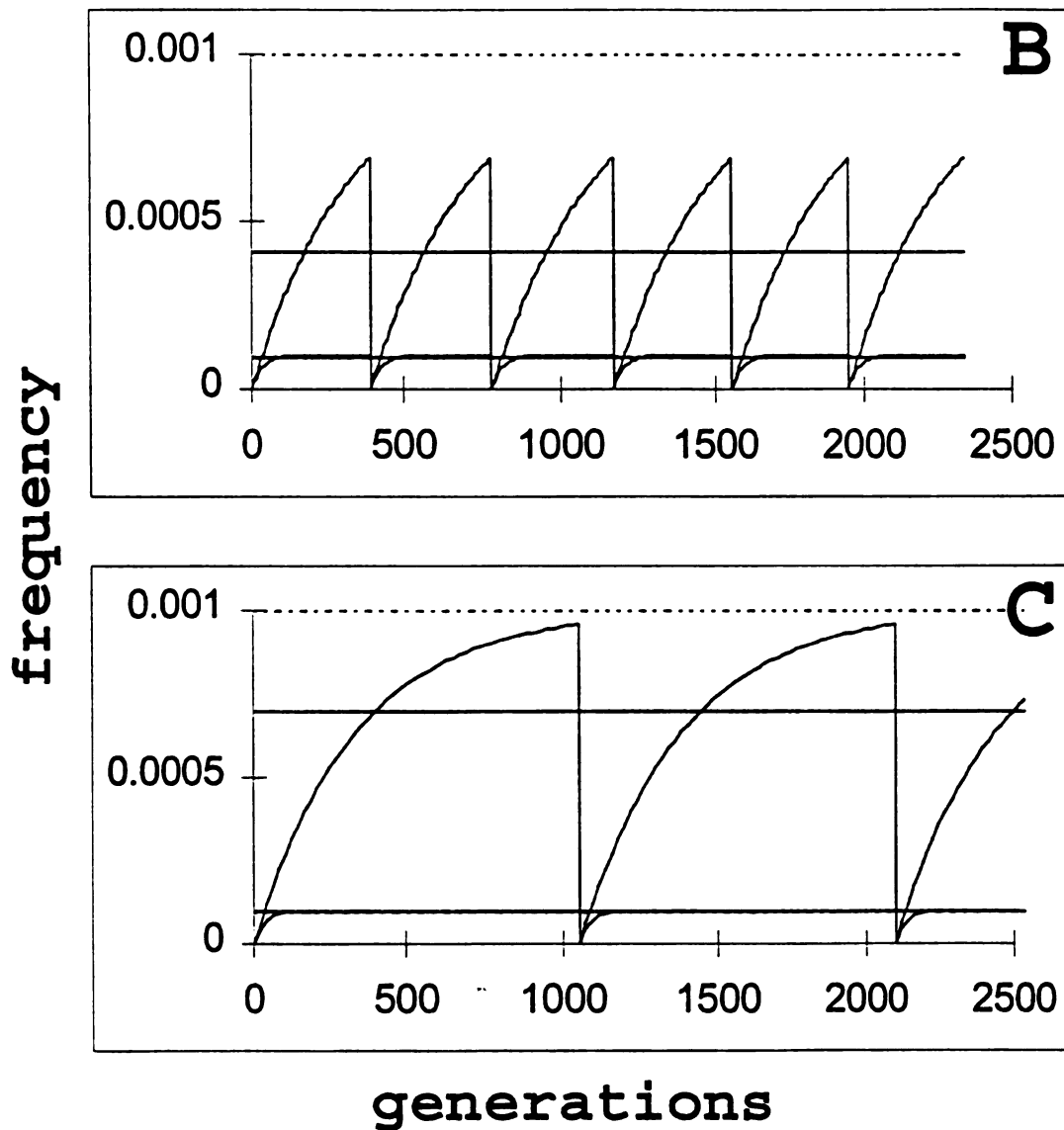


Figure 4.1 (continued). Panels B and C plot the frequency trajectories of the same deleterious mutations, but here adaptive substitutions occur at rates $\sigma = 0.0009$ and $\sigma = 0.003$, respectively. These substitution rates are determined from equation (88) using parameters $N = 3 \times 10^7$, $\alpha = 35$, $\mu_B = 4 \times 10^{-10}$ and 2×10^{-9} . The solid horizontal lines represent effective mutation-selection balance, as calculated by equation (87).

Fig
of
pop
pro
by
The
mut
 $\alpha =$

 $\alpha =$

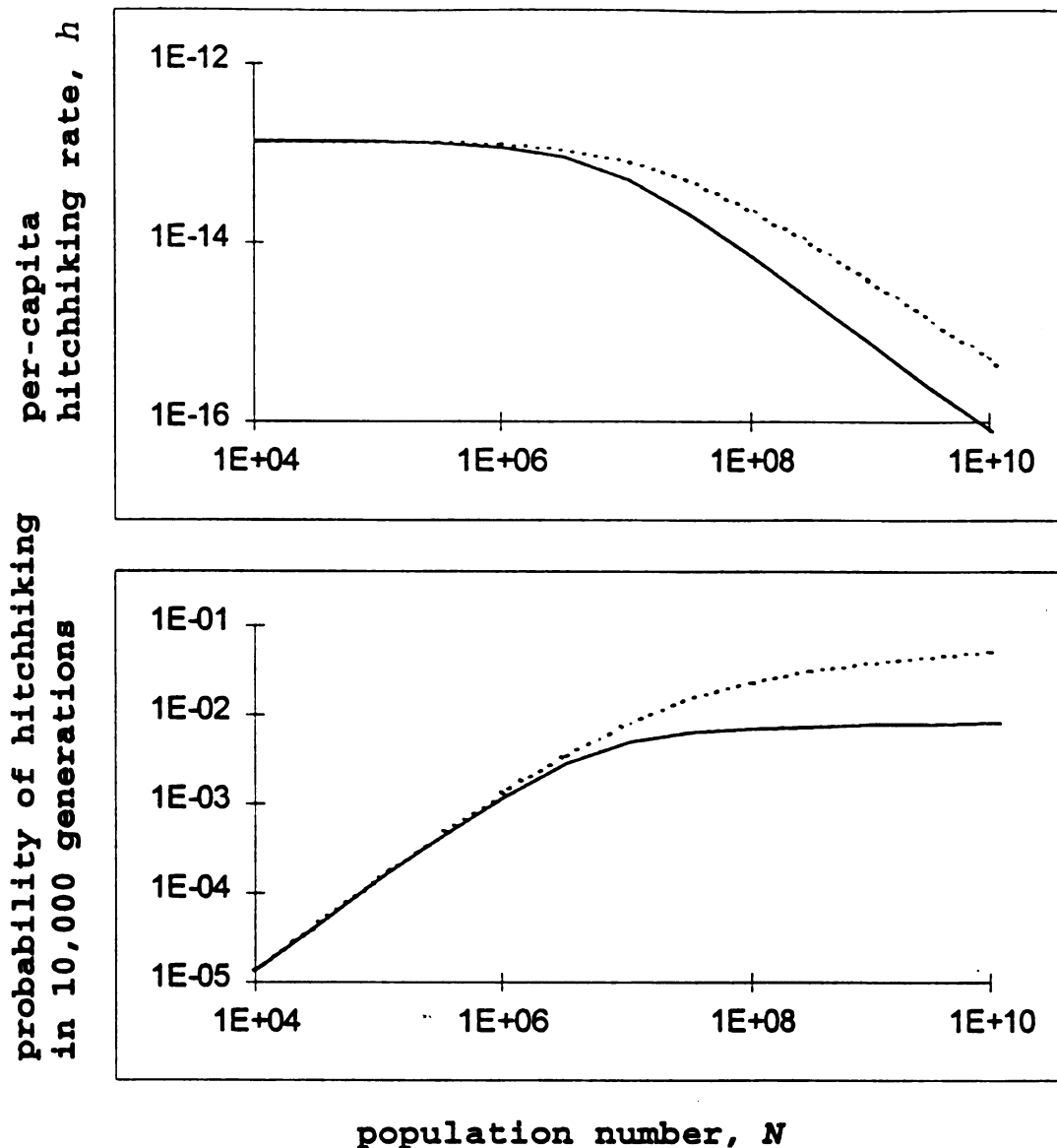


Figure 4.2. Panel A shows the per-capita hitchhiking rate of a specified deleterious mutation as a function of population number, N . Panel B shows the corresponding probability that a population fixes the deleterious mutation by hitchhiking within a time interval of 10,000 generations. The dashed lines show the result of incorrectly assuming mutation-selection equilibrium ($\phi = \hat{\phi}$). Parameters used are $\alpha = 35$, $\mu_b = 2 \times 10^{-9}$, $\mu_D = 2 \times 10^{-6}$, $s_D = 0.03$.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Fi
of
se
pr
by
Th
mu
α

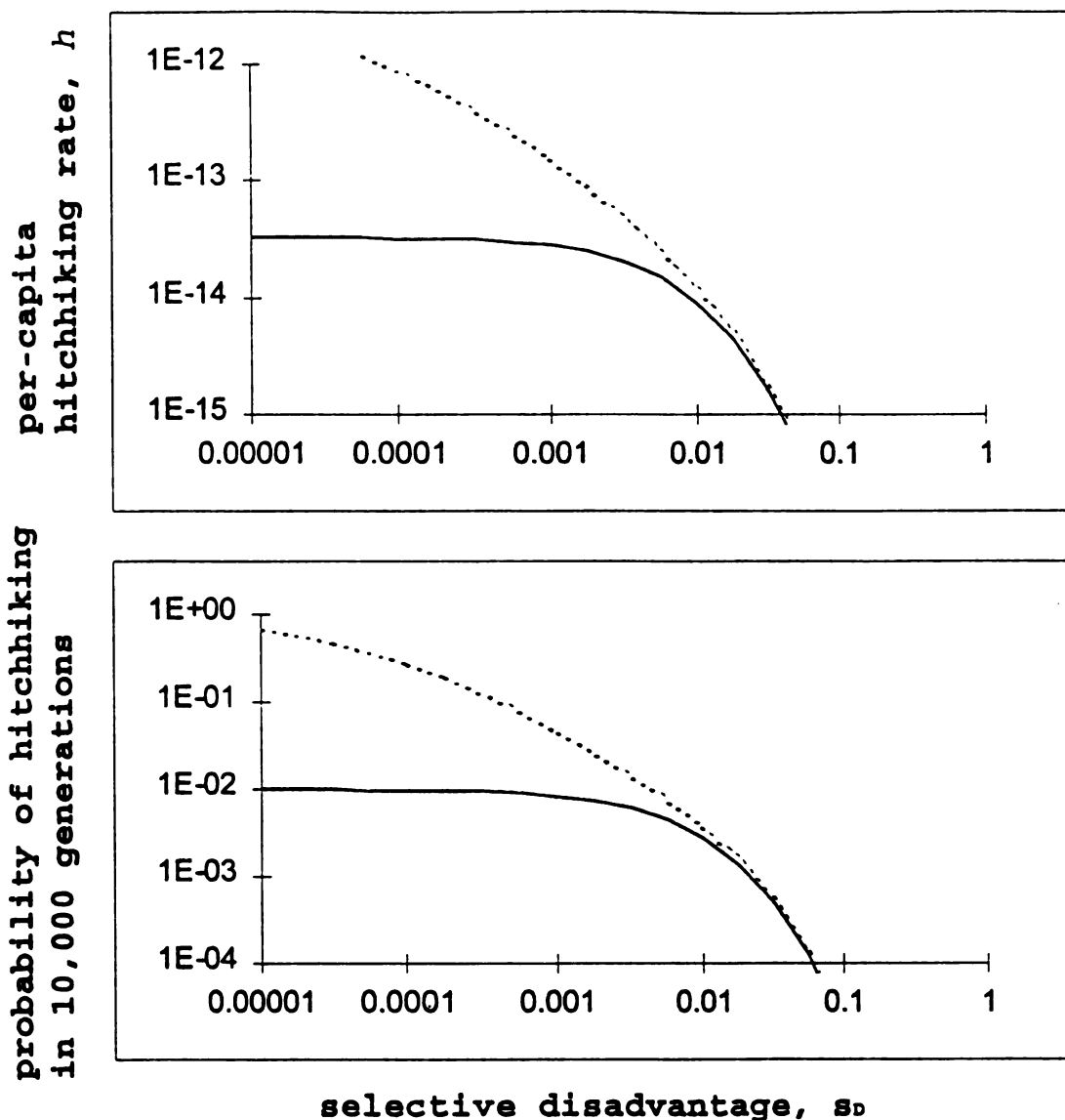


Figure 4.3. Panel A shows the per-capita hitchhiking rate of a specified deleterious mutation as a function of its selective disadvantage, s_d . Panel B shows the corresponding probability that a population fixes the deleterious mutation by hitchhiking within a time interval of 10,000 generations. The dashed lines show the result of incorrectly assuming mutation-selection equilibrium ($\phi = \hat{\phi}$). Parameters used are $\alpha = 35$, $\mu_b = 2.0 \times 10^{-9}$, $\mu_d = 2 \times 10^{-5}$, and $N = 3 \times 10^7$.

ner-canti ta

but the quality of the

Fi
of
B
fi
in
re
eq
 μ_E

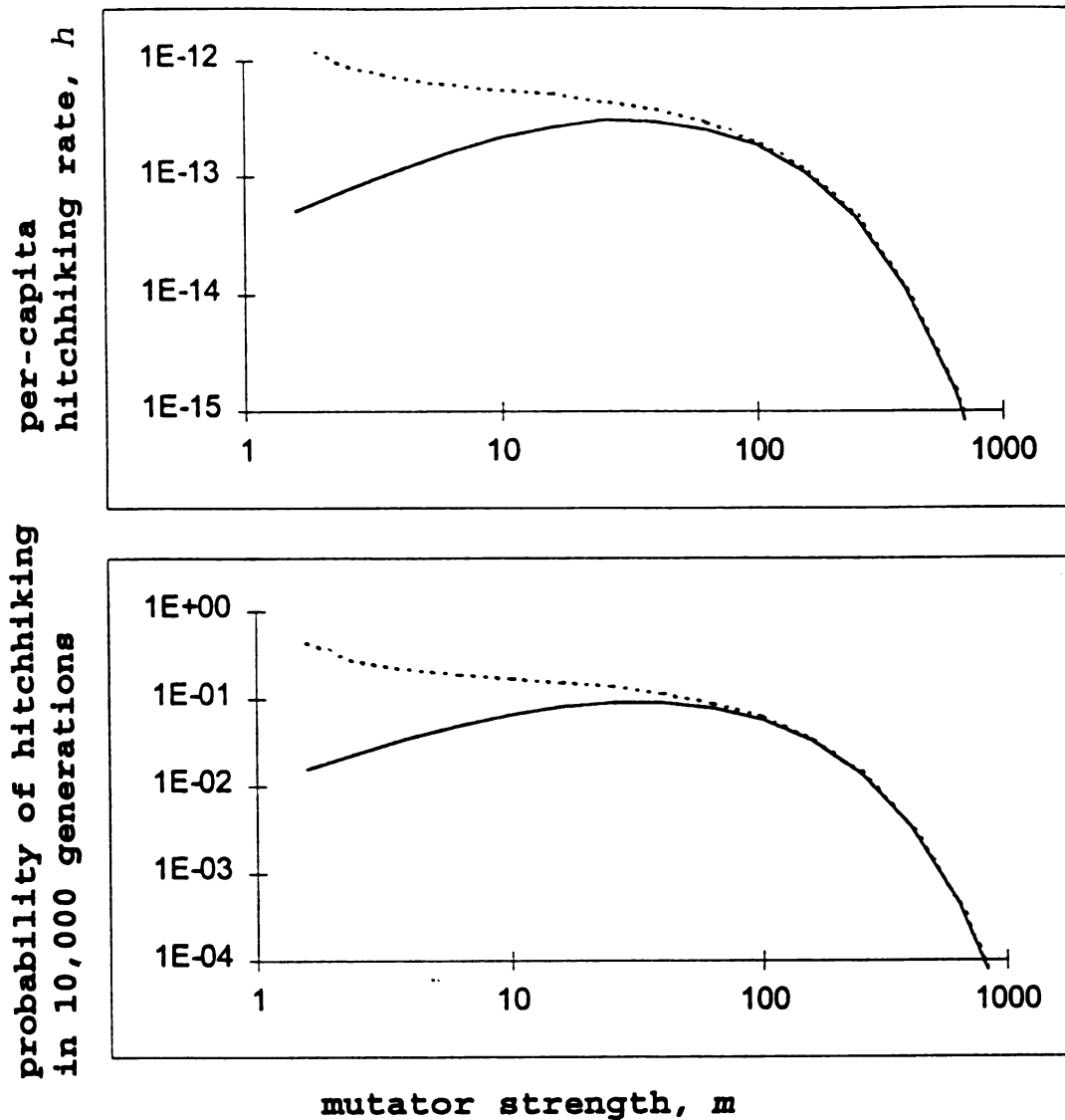


Figure 4.4. Panel A shows the per-capita hitchhiking rate of a mutator allele as a function of its strength, m . Panel B shows the corresponding probability that a population fixes the mutator allele by hitchhiking within a time interval of 10,000 generations. The dashed line shows the result of incorrectly assuming mutation-selection equilibrium ($\phi = \hat{\phi}$). Parameters used are $\alpha = 35$, $\mu_b = 2.0 \times 10^{-9}$, $\mu_D = 2 \times 10^{-6}$, $\mu_g = 0.0003$, and $N = 3 \times 10^7$.

per-capita double-mutation

Fig
mut
fun
res
equ
 $\alpha =$

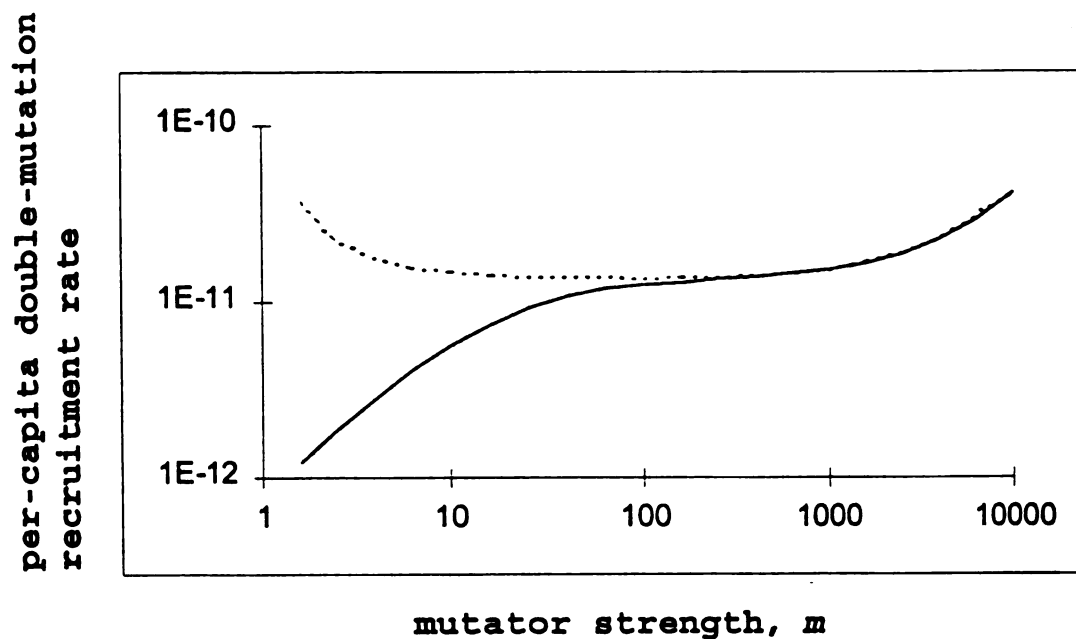


Figure 4.5. The per-capita recruitment rate of double-mutations (beneficial mutations on mutator background) as a function of mutator strength, m . The dashed line shows the result of incorrectly assuming mutation-selection equilibrium ($\phi = \hat{\phi}$). Parameters used are $N = 3 \times 10^7$, $\alpha = 35$, $\mu_B = 2.0 \times 10^{-9}$, $\mu_D = 2 \times 10^{-6}$, and $\mu_S = 0.0003$.

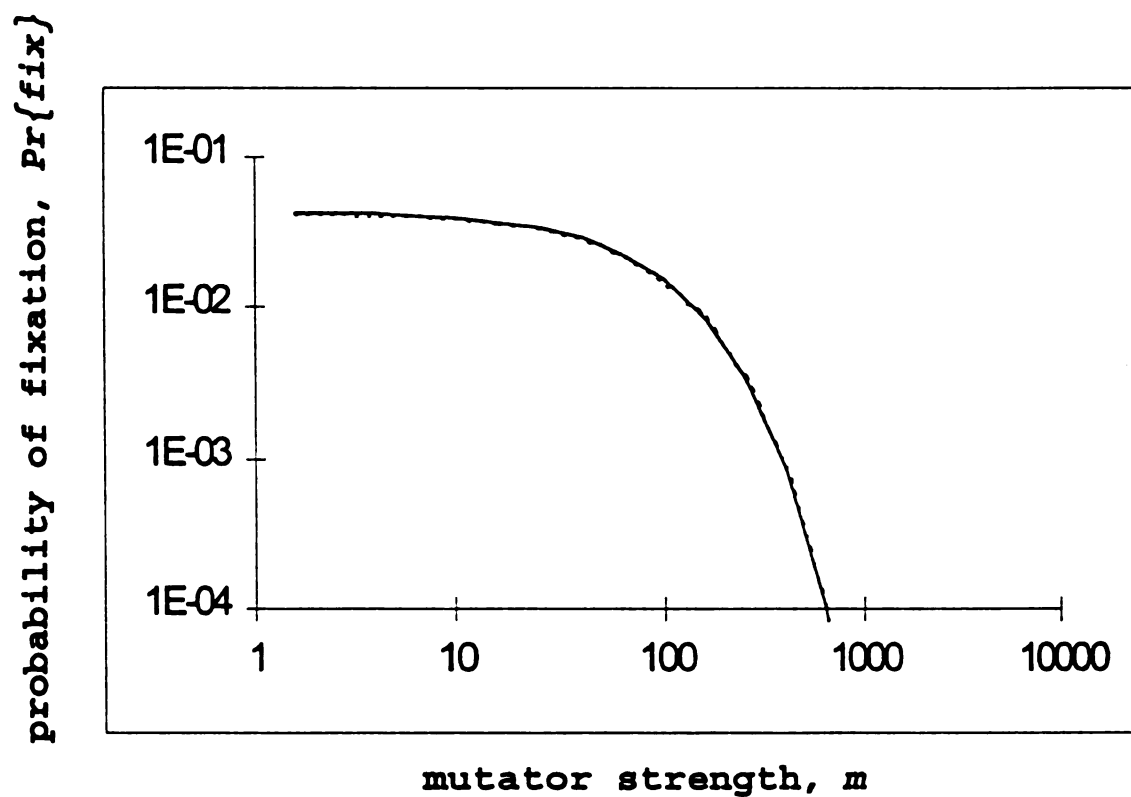


Figure 4.6. Probability of fixation of a double-mutation (beneficial mutation on mutator background) as a function of mutator strength, m . Parameters used are $N = 3 \times 10^7$, $\alpha = 35$, $\mu_b = 2.0 \times 10^{-9}$, $\mu_D = 2 \times 10^{-6}$, and $\mu_s = 0.0003$.

frequency

Fig
sub
bac
ach
Exa
sol
 $N =$

APPENDICES

APPENDIX 1.1

MUTATION RATE ESTIMATION PROGRAM

Theory

The computer program FT.EXE was written to estimate mutation rates from fluctuation test data (Luria & Delbruck, 1943). Briefly, fluctuation tests report the numbers of mutants that have accumulated by spontaneous mutation and subsequent replication during exponential growth of a population (Chapter 1). Computation of the expected distribution is notoriously difficult (Stewart et al., 1990). The probability generating function for this distribution was derived by Lea & Coulson (1949). From this generating function, an algorithm for generating the corresponding distribution was derived independently by Gurland (1958) and Ma et al. (1992). (See also Gurland, 1963; Sarkar et al., 1992; Jaegger & Sarkar, 1995.) The program FT.EXE employs this algorithm for computing the distribution of numbers of mutants in a fluctuation test as follows.

Let μ denote mutation rate of the wildtype to the selected mutation, let N denote the final population number, and let i denote the number of mutants in the final population. Then the probability, p_i , that i mutants are present in the final population is given by

$$p_0 = e^{-m}$$

$$p_i = \frac{m}{i} \sum_{j=0}^{i-1} \frac{p_j}{i-j+1} , \quad (46)$$

where $m = \mu N$ is the expected number of *mutations* (not to be confused with *mutants*). To estimate m , the program FT.EXE employs the Newton series (derived below) for the recurrence relation given by (46).

The program also implements the empirical formulas for 95% confidence limits derived by Stewart (1994). For small sample sizes, however, Stewart's formulas have no solution. To make estimates from small sample sizes, I derive (below) the analytical variance based on the maximum likelihood surface. As an alternative to Stewart's confidence limits, the program also gives 95% confidence limits based on computation of this analytical variance. Given this variance, one can base statistical inference on either the

assumption of normality or the implementation of Chebyshev's inequality (Feller, 1968).

Let S denote the sample set, let $m = \mu N$ denote expected number of mutations, and let $u = \ln m$. Given u , $p_j(u)$ denotes the probability of j mutants in the final population. Then, the log-likelihood function is

$$L(u) = \sum_{j \in S} \ln p_j(u) \quad . \quad (47)$$

From this, the estimate of u is computed by solution of

$$L'(u) = \sum_{j \in S} \frac{p'_j(u)}{p_j(u)} = 0 \quad , \quad (48)$$

where the prime indicates derivative. The variance of u is given by

$$\text{Var}(u) = -[L''(u)]^{-1} \quad (49)$$

(see Hogg & Craig, 1995). To compute the variance, we have

$$L''(u) = \sum_{j \in S} \frac{p''_j(u)}{p_j(u)} - \left(\frac{p'_j(u)}{p_j(u)} \right)^2 \quad . \quad (50)$$

The derivatives necessary for computing (48) and (50) are given by

$$\mathbf{P}_i(u) = \frac{e^u}{i} \sum_{j=0}^{i-1} \frac{\mathbf{A}\mathbf{P}_j(u)}{(i-j+1)}, \quad (51)$$

$$\text{where } \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad \mathbf{P}_k(u) = \begin{bmatrix} p_k(u) \\ p'_k(u) \\ p''_k(u) \end{bmatrix}, \quad \text{and } \mathbf{P}_0(u) = \begin{bmatrix} e^{-e^u} \\ -e^{u-e^u} \\ (e^u - 1)e^{u-e^u} \end{bmatrix}.$$

To estimate u , the program solves (48) numerically by Newton's method. This is achieved by iterating on r :

$$u_{r+1} = u_r - \left[\frac{L'(u)}{L''(u)} \right]_{u=u_r}. \quad (52)$$

Once an estimate for u is obtained, this estimate is used to compute the variance by (49). From here, the program assumes that u is normally distributed to compute confidence limits. (Stewart (1994) suggests that u is more nearly normal than m .) However, the program may be easily modified

to implement Chebyshev's inequality in order to avoid assumptions about how u is distributed. Lastly, the estimate of the number of mutations, m , as well as confidence limits on this estimate are obtained by back-transformation, $m = e^u$. From here, the mutation rate estimate and confidence limits are obtained from $\mu = m/N$.

Obtaining and using the program

The compiled program, FT.EXE, as well as the source code, FT.BAS (or FT.TXT for ASCII format), may be obtained at the following ftp site:

`ftp://ftps.cdc.gov/pub/MuRates`

The username and password needed to log on to this site are "ncidftp" and "12emerG" (case sensitive), respectively.

Also at that site are: (i) FTDOC.TXT, documentation on how to use the program, (ii) FORMAT.TXT, an outline of the input format, and (iii) SAMPLE.DAT, an example data file.

APPENDIX 2.1

PROBABILITY OF SURVIVING DRIFT

In the first few generations of growth, a beneficial mutation may be lost by random sampling events, or drift. Haldane (1927) derived the probability of surviving drift for a single beneficial mutation. His derivation made use of a result from the theory of branching processes, which states that probability of extinction (i.e., *not* surviving drift) is obtained by solving the equation $f(\theta) = \theta$, where $f(\theta)$ is the probability generating function for number of offspring (see Ewens, 1969, p. 79). A simple assumption for multicellular, sexual organisms is that this function generates a Poisson distribution, in which case the probability of survival of a beneficial mutation approximates $2s$. Our analyses, however, are based on the fundamental assumption of no recombination. We may further restrict our analysis to a particular kind of asexual organism, namely asexual bacteria. Bacteria reproduce by binary fission, and so we derive the generating function as

follows. Our assumption of a constant population size (see below) implies a sampling event every generation. Thus, a bacterium that divides before sampling will leave zero, one, or two offspring after sampling. In the case of bacteria, therefore, the probability generating function for number of offspring is

$$f(\theta) = (1-c/2)^2 + c(1-c/2)\theta + (c/2)^2\theta^2, \quad (53)$$

where c is the expected number of offspring after division and sampling. Thus, the probabilities of passing zero, one, and two offspring to the next generation are, respectively, $(1-c/2)^2$, $c(1-c/2)$, and $(c/2)^2$. The selective advantage of the mutant is $s = \ln c$ by definition, or approximately $s \approx c - 1$ when s is small. Let $\pi(s)$ denote the probability that a beneficial mutant survives drift. Then, by substituting $1 + s$ for c in (53) and solving the equation

$$f(1-\pi(s)) = 1-\pi(s), \text{ we obtain } \pi(s) = \frac{4s}{(1+s)^2}, \text{ which is}$$

approximately $4s$ for small s . All derivations in this dissertation employ the general notation, $\pi(s)$, whereas all computations implement the approximation, $\pi(s) \approx 4s$.

APPENDIX 2.2

n-GENOTYPE LOGISTIC SYSTEM WITH MUTATION

General solution

Logistic dynamics of an *n*-genotype system are modeled by assuming that (i) total population size is constant, i.e.,

$$\sum_{i=1}^n x_i = N, \text{ where } x_i \text{ is number of individuals of genotype } i,$$

and (ii) the differences in Malthusian parameters are constant:

$$m_i - m_1 = s_i, \quad i = 2, \dots, n, \quad (54)$$

$$\text{where } m_i = \frac{1}{x_i} \frac{dx_i}{dt}, \text{ and } m_1 = \frac{1}{x_1} \frac{dx_1}{dt} = \left(\sum_{i=2}^n x_i - N \right)^{-1} \left(\sum_{i=2}^n \frac{dx_i}{dt} \right).$$

Equation (54) may, therefore, be rewritten as:

$$\frac{1}{x_i} \frac{dx_i}{dt} + \left(N - \sum_{j=2}^n x_j \right)^{-1} \left(\sum_{j=2}^n \frac{dx_j}{dt} \right) = s_i, \quad i = 2, 3, \dots, n. \quad (55)$$

This system of $n-1$ equations can be rearranged as follows:

$$\frac{dx_i}{dt} = x_i \left(s_i - \frac{1}{N} \sum_{j=2}^n s_j x_j \right) , \quad (56)$$

where $i = 2, 3, \dots, n$. While this system of equations is non-linear, its symmetry makes an analytical solution possible.

The key to its solution is the transformation,

$X_i = \ln x_i - s_i t$. The system of equations now becomes:

$$\frac{dX_i}{dt} = -\frac{1}{N} \sum_{j=2}^n s_j e^{X_j + s_j t} , \quad (57)$$

where $i = 2, 3, \dots, n$. Thus, the time derivatives of all

transformed variables are equal:

$$\frac{dX_i}{dt} - \frac{dX_j}{dt} = 0 , \quad (58)$$

where $i, j = 2, 3, \dots, n$. Integration of (58) yields

$X_i - X_j = k_{ij}$, and k_{ij} is a constant of integration that is

determined from initial conditions:

$$k_{ij} = X_i(0) - X_j(0) = \ln x_i(0) - \ln x_j(0) , \quad (59)$$

where $i, j = 2, 3, \dots, n$. Thus, the system of equations is uncoupled by substituting X_j from (57) with $X_i - k_{ij}$, which yields:

$$\frac{dX_i}{dt} = -\frac{1}{N} \sum_{j=2}^n s_j e^{X_i - k_{ij} + s_j t}, \quad (60)$$

where $i = 2, 3, \dots, n$. From solution and subsequent back-transformation of equation (60), the analytical solution of an n -genotype logistic system is obtained:

$$x_i(t) = x_i(0) e^{s_i t} \left(1 + \frac{1}{N} \sum_{j=2}^n x_j(0) (e^{s_j t} - 1) \right)^{-1}, \quad (61)$$

where $i = 2, 3, \dots, n$, and

$$x_1(t) = N - \sum_{j=2}^n x_j(t). \quad (62)$$

Application of boundary conditions due to mutation

If genotype i appears by mutation at time τ_i , then boundary conditions are $x_i(\tau_i) = 1$. From these, the initial conditions are determined; they are

$$\mathbf{X}_0 = \mathbf{R}^{-1} \mathbf{N} , \quad (63)$$

where \mathbf{X}_0 is a vector whose elements are $x_i(0)$, $i =$

$2, 3, \dots, n$, \mathbf{R} is an $n-1 \times n-1$ matrix whose elements are

$$r_{ij} = \begin{cases} N e^{s_i \tau_i} + 1 , & i=j \\ 1 - e^{s_j \tau_i} , & i \neq j \end{cases}, \quad i, j = 2, 3, \dots, n, \text{ and } \mathbf{N} \text{ is a vector}$$

whose $n-1$ elements are the constant N .

Notation for the 3-genotype case

The developments in this appendix use a more general notation than is used in the rest of the paper, where x_1 is simply denoted by x , x_2 is denoted by y , and x_3 is denoted by z . This 3-genotype case has the particular solution,

$$\begin{aligned} y(t) &= e^{s_y t} \left[\frac{1}{y(0)} + \frac{1}{N} \left(e^{s_y t} - 1 + \frac{z(0)}{y(0)} \left(e^{s_z t} - 1 \right) \right) \right]^{-1} \\ z(t) &= e^{s_z t} \left[\frac{1}{z(0)} + \frac{1}{N} \left(e^{s_z t} - 1 + \frac{y(0)}{z(0)} \left(e^{s_y t} - 1 \right) \right) \right]^{-1} . \end{aligned} \quad (64)$$

The initial conditions are determined from the boundary conditions, $y(0) = 1$ and $z(t_z) = 1$; they are

$$\begin{aligned} y(0) &= 1 \\ z(0) &= \frac{e^{s_y t_z} + N}{N e^{s_z t_z} + 1} . \end{aligned} \tag{65}$$

APPENDIX 2.3

EXPECTED NUMBER OF CANDIDATE REPLICATIONS

Here we derive the expected number of replications that may generate superior mutations that prevent a given beneficial mutation from attaining some frequency, f . We have called these *candidate replications*, denoted by R , in the subsection, *Clonal interference - a general model*. The crucial step in the derivation of R is finding an expression for the time, t_z , at which a superior mutation must appear if the original mutation is to attain a maximum frequency of exactly f .

The time, t_{\max} , at which y reaches a maximum number is determined from $\left. \frac{dy}{dt} \right|_{t_{\max}(t_z)} = 0$; it is

$$t_{\max}(t_z) = \frac{1}{s_z} \ln \left(\frac{s_y}{s_z - s_y} \frac{N^2}{e^{(s_y - s_z)t_z} + N e^{-s_z t_z}} \right) . \quad (66)$$

When $e^{(s_y - s_z) t_z} < N e^{-s_z t_z}$ (i.e. when $e^{s_y t_z} < N$), equation (66)

is well approximated by omitting the term $e^{(s_y - s_z) t_z}$ from the denominator, resulting in the approximation

$$t_{\max}(t_z) \approx \frac{1}{s_z} \ln \left(\frac{N s_y}{s_z - s_y} \right) + t_z. \quad (67)$$

The constraints under which this approximation works well are discussed later.

We now calculate the time t_z at which superior mutation z must appear if y is to achieve a maximum of exactly fN .

The solution to $y(t_{\max}(t_z)) = fN$ is

$$t_z = \frac{1}{s_y} \ln \left(\frac{N \left(1 + \frac{s_y}{s_z - s_y} \right)}{\left(\frac{1}{f} - 1 \right) \left(\frac{N s_y}{s_z - s_y} \right)^{s_y/s_z} - \frac{s_y}{s_z - s_y}} \right). \quad (68)$$

Next, we use the fact that the expected number of candidate replications, R , given s_y and $s_z > s_y$, is well approximated by evaluating R at $s_z = \langle s_z | s_z > s_y \rangle = s_y + \frac{1}{\alpha}$, to derive the

expected number of candidate replications:

$$R \approx \int_0^{\hat{t}_z} x(t) dt = \frac{N}{s} \ln N - \frac{N}{s} \ln \left[1 + \left(\frac{1}{1+\alpha s} \right) \left(\left(\frac{1}{f} - 1 \right) (\alpha s N)^{\frac{\alpha s}{\alpha s + 1}} - \alpha s \right) \right], \quad (69)$$

where \hat{t}_z is simply t_z evaluated at $s_z = s_y + \frac{1}{\alpha}$.

The approximation made in equation (67) is, for our purposes, essentially an equality when $e^{s_y t_z} < N$. If we combine this condition with equation (68), then the approximation works well only when the frequency f meets the following condition:

$$f < \left[\frac{2}{N} \left(\frac{N s_y}{s_z - s_y} \right)^{1 - \frac{s_y}{s_z}} + 1 \right]^{-1}. \quad (70)$$

If we let $s_z = \langle s_z | s_z > s_y \rangle = s_y + \frac{1}{\alpha}$, and if we simplify the

notation so that $s = s_y$, then the above condition becomes

$$f < \left[\frac{2}{N} (\alpha s N)^{\frac{1}{1+\alpha s}} + 1 \right]^{-1} . \quad (71)$$

This upper bound on f reaches a minimum value when $\frac{\partial f}{\partial s} = 0$,

so that an overall bound below which the approximation works well is obtained by solving for the value of s that

satisfies $\ln(\alpha s N) = 1 + \frac{1}{\alpha s}$ and using that value in

equation (71). In general, the approximation is valid when

$f < 0.95$ provided that N is greater than 10^4 . For the

purposes of this paper, the approximation is essentially an equality because we are concerned only with the cases

$f = 0.01$ and $f = 0.5$, for which the approximation works

extremely well. We compute fixation probabilities, i.e.,

the boundary case $f > \frac{N-1}{N}$, using the simpler derivations in

Clonal interference and fixation.

APPENDIX 2.4

FUNCTIONS EMPLOYING THE RECTANGULAR DISTRIBUTION

We present here the results only of the derivations in which a rectangular distribution replaces the exponential distribution of beneficial mutational effects. The probability of fixation of an arbitrarily chosen beneficial mutation is:

$$Pr\{fix|s_{\max}, \mu, N\} = \frac{1}{s_{\max}} \int_0^{s_{\max}} \pi(s) e^{-\lambda_R(s, s_{\max}, \mu, N)} ds, \quad (72)$$

$$\text{where } \lambda_R(s, s_{\max}, \mu, N) = \frac{\mu}{s} N \ln N \pi\left(\frac{s_{\max}^2 - s^2}{2s_{\max}}\right) \text{ (assuming that}$$

$\pi(u)$ is approximately linear). The expected rate of substitution of beneficial mutations is:

$$\langle \sigma_R(s_{\max}, \mu, N) \rangle = \mu N \Pr\{\text{fix} | s_{\max}, \mu, N\} . \quad (73)$$

The expected selection coefficient of successful mutations is:

$$\langle s_R(s_{\max}, \mu, N) \rangle = \frac{\int_0^{s_{\max}} s \Pi(s) e^{-\lambda_R(s, s_{\max}, \mu, N)} ds}{\int_0^{s_{\max}} \Pi(s) e^{-\lambda_R(s, s_{\max}, \mu, N)} ds} , \quad (74)$$

where $\lambda_R(s, s_{\max}, \mu, N)$ is as defined above for equation (72).

The expected number of superior mutations in the interval $(0, \hat{t}_z)$ is:

$$\psi_R(s, s_{\max}, \mu, N, f) = \frac{\mu}{s} N \ln \left(\frac{N}{\chi_R} \right) \Pi \left(\frac{s_{\max}^2 - s^2}{2s_{\max}} \right) . \quad (75)$$

The expected number of superior mutations in the interval (\hat{t}_z, t) is:

$$\gamma_R(s, s_{\max}, \mu, N, f) = \frac{\mu}{s} N \ln(\chi_R) \Pi \left(\frac{s_{\max}^2 - s^2}{2s_{\max}} \right) , \quad (76)$$

where $\chi_R = 1 + \frac{1}{s_{\max} + s} \left[(s_{\max} - s) \left(\frac{1}{f} - 1 \right) \left(\frac{2sN}{s_{\max} - s} \right)^{\frac{2s}{s_{\max} + s}} - 2s \right]$. The

probability that an arbitrarily chosen beneficial mutation transiently achieves polymorphic frequency ($f > 0.01$) is:

$$\begin{aligned} Pr\{poly|s_{\max}, \mu, N\} = \\ \frac{1}{s_{\max}} \int_0^{s_{\max}} \pi(s) e^{-\psi_R(s, s_{\max}, \mu, N, 0.01)} (1 - e^{-\gamma_R(s, s_{\max}, \mu, N, 0.01)}) ds. \end{aligned} \quad (77)$$

Finally, the probability that an arbitrarily chosen beneficial mutation transiently achieves majority status is obtained by replacing 0.01 in equation (77) with 0.5.

APPENDIX 3.1

ADAPTIVE SUBSTITUTIONS IN ASEXUAL POPULATIONS ARE ACCURATELY MODELED AS INSTANTANEOUS REPLACEMENTS

Conveniently, the continuous process of adaptive substitution in asexual populations is well approximated by a discrete process of instantaneous replacement. Lenski et al. (1991) suggested the use of such an approximation. They reasoned that in a large population the frequency of a substituting variant would remain very low for a considerable time and then rise sharply, thus approximating a step function.

Define time of adaptive substitution, t^* , as the time at which the frequency of a beneficial mutation achieves 0.5. Then, the process of adaptive substitution may be approximated by a simplified scenario in which the substituting variant remains at a frequency of zero until time t^* and assumes a frequency of one thereafter. That we have defined t^* appropriately is evidenced by the symmetry of logistic growth. More generally, let t_i^* denote the time of the i^{th} adaptive substitution in an evolving population. Then populational processes are well approximated by

assuming the population to be numerically dominated by a single "wildtype" variant throughout the interval (t_{i-1}^*, t_i^*) , for every i .

Suppose a beneficial mutation appears in a population at time $t=0$ and subsequently spreads to fixation. Let p denote the frequency of the beneficial allele. Then, if the unit of time is generations, the dynamic equation for the growth of the beneficial mutation is $dp/dt = sp(1-p)$, where s is the selection coefficient, and $p(0)=1/N$. Define time until fixation, t_f , as that which satisfies $p(t_f)=(N-1)/N$. The time of substitution, t^* , as defined above, is that which satisfies $p(t^*)=0.5$; it is $t^* = \ln(N)/s$. Thus, according to the continuous substitution process, the total number of wildtype replications after time $t=0$ is given by

$$\int_0^{t_f} [1-p(t)] dt = \frac{N \ln(N)}{s} = Nt^*.$$

The discrete approximation to the above substitution process is given by the equation, $p(t) = \begin{cases} 0, & t \leq t^* \\ 1, & t > t^* \end{cases}$, where

t^* is as defined above. According to this approximation,

the number of wildtype replications after time $t=0$ is simply Nt^* , which is exactly equal to the number obtained for the continuous process.

We conclude that (i) the time of substitution, t^* , is appropriately defined and (ii) the continuous process of substitution is closely approximated by a discrete process of instantaneous replacement at time t^* .

APPENDIX 3.2

ANALYTICAL INTEGRATION OF EQUATIONS (21) AND (22).

To analytically integrate equation (21), a more precise notation is essential. Given fixation ordering, ζ , the i^{th} mutation fixed will be followed by a number of fixations of mutations that are superior to the i^{th} mutation. We have denoted the set of such superior mutations as \mathbf{S}_i . Let the subscript S indicate membership in set \mathbf{S}_i , such that $s_S(k)$ denotes the k^{th} member of this set. Now, order this set of selection coefficients such that $s_S(1) > s_S(2) > \dots > s_S(\#\mathbf{S}_i)$, where $\#\mathbf{S}_i$ denotes the cardinality of set \mathbf{S}_i . The i^{th} mutation fixed will also be followed by a number of fixations of mutations that are inferior to the i^{th} mutation. We have denoted the set of such inferior mutations as \mathbf{I}_i . Let the subscript I denote membership in set \mathbf{I}_i , such that $s_I(k)$ denotes the k^{th} member of this set. Now, order this set of selection coefficients such that $s_I(1) > s_I(2) > \dots > s_I(\#\mathbf{I}_i)$, where $\#\mathbf{I}_i$ denotes the

cardinality of set I_i . With this new notation, the integral in equation (21) may be rewritten as

$$\begin{aligned}
& \int_0^{\ln N/s_I(1)} \exp\left\{-\left(t_i + \frac{\ln N}{s_i}\right) \sum_{k=1}^{\#S_i} r_S(k)\right\} g(t_i) dt_i \\
& + \int_{\ln N/s_I(1)}^{\ln N/s_I(2)} \exp\left\{-\left(t_i + \frac{\ln N}{s_i}\right) \sum_{k=1}^{\#S_i} r_S(k) - \left(t_i - \frac{\ln N}{s_I(1)}\right) r_I(1)\right\} g(t_i) dt_i \\
& + \int_{\ln N/s_I(2)}^{\ln N/s_I(3)} \exp\left\{-\left(t_i + \frac{\ln N}{s_i}\right) \sum_{k=1}^{\#S_i} r_S(k) - \sum_{k=1}^2 \left(t_i - \frac{\ln N}{s_I(k)}\right) r_I(k)\right\} g(t_i) dt_i \\
& \vdots \\
& + \int_{\ln N/s_I(\#I_i-1)}^{\ln N/s_I(\#I_i)} \exp\left\{-\left(t_i + \frac{\ln N}{s_i}\right) \sum_{k=1}^{\#S_i} r_S(k) - \sum_{k=1}^{\#I_i-1} \left(t_i - \frac{\ln N}{s_I(k)}\right) r_I(k)\right\} g(t_i) dt_i \\
& + \int_{\ln N/s_I(\#I_i)}^{\infty} \exp\left\{-\left(t_i + \frac{\ln N}{s_i}\right) \sum_{k=1}^{\#S_i} r_S(k) - \sum_{k=1}^{\#I_i} \left(t_i - \frac{\ln N}{s_I(k)}\right) r_I(k)\right\} g(t_i) dt_i
\end{aligned} \tag{78}$$

Thus, when analytically integrated, (21) becomes

$$\begin{aligned}
\Phi(\zeta) = & \prod_{i=1}^M \sum_{j=0}^{\#I_i} \frac{r_i}{R} \exp\left\{-\frac{\ln N}{s_i} \sum_{k=1}^{\#S_i} r_S(k) + \sum_{k=1}^j \frac{\ln N}{s_I(k)} r_I(k)\right\} \\
& \times \left(\exp\left\{-R \frac{\ln N}{s_I(j)}\right\} - \exp\left\{-R \frac{\ln N}{s_I(j+1)}\right\} \right),
\end{aligned} \tag{79}$$

where $R = r_i + \sum_{k=1}^{\#S_i} r_S(k) + \sum_{k=1}^j r_I(k)$, $s_I(0) = \infty$ and $s_I(\#I_i+1) = 0$.

We now derive the limit as selection coefficients converge. Let the ordered set of selection coefficients be

evenly spaced such that $s_j = s_{j-1} + \epsilon$. Then, the limit as selection coefficients converge of the probability of a given fixation ordering is

$$\lim_{\epsilon \rightarrow 0} \Phi(\zeta) = \prod_{i=1}^{M-1} \left(\frac{1}{\#s_i + 1} \exp\{-\#s_i \mu_i K N \ln N\} \left[1 - \exp\{-(\#s_i + 1) \mu_i K N \ln N\} \right] + \frac{1}{\#s_i + \#I_i + 1} \exp\{-(2\#s_i + 1) \mu_i K N \ln N\} \right), \quad (80)$$

where K denotes the linear constant in the drift survival function, and the intercept of that function is assumed to equal zero, i.e., $\pi(s) = Ks$. The mutation rate to each beneficial mutation is assumed to be equal such that $\mu_i = \mu/M \forall i$ where μ is the overall beneficial mutation rate. When population number is very large, such that $\mu N > 1$, this limit is closely approximated by

$$\Phi(\zeta) \approx \begin{cases} \prod_{i=1}^{M-1} \frac{1}{\#s_i + 1} \exp\{-\#s_i \mu_i K N \ln N\} & , \text{ not rank-ordered} \\ \prod_{i=1}^{M-1} \left[1 - \left(1 - \frac{1}{M-i+1} \right) \exp\{-\mu_i K N \ln N\} \right] & , \text{ rank-ordered} \end{cases} \quad (81)$$

Equation (22) may be integrated analytically by decomposition of the integral into intervals as before. The appropriate decomposition is achieved by simply multiplying each integrand of (78) by t_i . Then, analytical integration yields

$$E(\tau_i | \zeta) = \sum_{j=0}^{\#I_i} \frac{r_i}{R} \exp \left\{ -\frac{\ln N}{s_i} \sum_{k=1}^{\#S_i} r_S(k) + \sum_{k=1}^I \frac{\ln N}{s_I(k)} r_I(k) \right\} \times \left[e^{-R \frac{\ln N}{s_{I(j)}} \left(\frac{\ln N}{s_I(j)} + \frac{1}{R} \right)} - e^{-R \frac{\ln N}{s_{I(j+1)}} \left(\frac{\ln N}{s_I(j+1)} + \frac{1}{R} \right)} \right], \quad (82)$$

where $R = r_i + \sum_{k=1}^{\#S_i} r_S(k) + \sum_{k=1}^I r_I(k)$, $s_i(0) = \infty$ and $s_i(\#I_i + 1) = 0$.

In the limit as selection coefficients converge, this expected time is reduced to

$$\lim_{\epsilon \rightarrow 0} E(\tau_i | \zeta) = \frac{\exp\{-\#S_i \mu KN \ln N\}}{\#S_i + 1} \left(\frac{1}{R} - \left(\frac{\ln N}{s} + \frac{1}{R} \right) \exp\{-(\#S_i + 1) \mu KN \ln N\} \right) + \frac{\exp\{-(\#S_i + \#I_i) \mu KN \ln N\}}{\#S_i + \#I_i + 1} \left(\frac{\ln N}{s} + \frac{1}{R} \right) \exp\{-(\#S_i + \#I_i + 1) \mu KN \ln N\} \quad (83)$$

where s is the average selection coefficient (to which all s_j converge).

APPENDIX 4.1

EFFECTIVE MUTATION-SELECTION BALANCE

Let $\phi(t)$ denote the frequency of a given deleterious mutation, where $t=0$ at the time of the most recent adaptive substitution (see below for precise definition). Given that the time until the next adaptive substitution is t_s , the average mutation selection balance during the time interval between the most recent and the next substitution is

$\frac{1}{t_s} \int_0^{t_s} \phi(t) dt$. Let σ denote the expected rate of adaptive

substitutions. Then, to a first approximation, we define *effective mutation selection balance* as

$$\phi_e = \sigma \int_0^{1/\sigma} \phi(t) dt \quad . \quad (84)$$

Suppose a beneficial mutation appears on the wildtype background at time $t = 0$. As the number of individuals carrying this beneficial mutation grows, they produce at

rate μ_D a deleterious mutation, which has selective disadvantage s_D . Assuming constant population size, N , the frequency, ϕ , of this deleterious mutation on the beneficial background is given by the dynamic equation,

$$\frac{d\phi}{dt} = ([(N-1) e^{-s_B t} + 1]^{-1} \mu_D - s_D \phi) (1-\phi) , \quad (85)$$

with initial condition $\phi(0) = 0$, reflecting the fact that the beneficial mutation occurs on the wildtype background. It is commonly the case that $\phi \ll 1$, for which the above equation is well approximated by

$$\frac{d\phi}{dt} = \begin{cases} \mu_D - s_D \phi & , \quad t \geq t_{1/2} \\ 0 & , \quad t < t_{1/2} \end{cases} \quad (86)$$

where $t_{1/2} = \frac{\ln N}{s_B}$ is the time necessary for the beneficial

mutation to attain a frequency of 0.5, and again this equation has initial condition $\phi(0) = 0$. For convenience, we now shift our time axis such that the time at which the beneficial mutation achieves frequency 0.5 is zero, i.e.,

$t_{1/2} \doteq 0$. We define this time as the time of substitution.

That equation (85) is well approximated by (86) is good evidence that the time of substitution, as we have defined it, marks a "resetting" of genetic variability in the population. Figure 4.7 plots the frequency of a deleterious mutation in a population during a substitutional event. The vertical dotted line marks the time of substitution as defined here (the $t=0$ axis in the shifted coordinate system). Note the close agreement between exact and approximate solutions.

In our shifted coordinate system, equation (86) becomes

$$\frac{d\phi}{dt} = \mu_D - s_D \phi. \text{ When the initial condition, } \phi(0) = 0, \text{ is}$$

applied, the solution is $\phi = \frac{\mu_D}{s_D} (1 - e^{-s_D t})$. This solution

is plotted in Figure 4.1A for two different values of s_D ;

note that as $t \rightarrow \infty$, $\phi \rightarrow \frac{\mu_D}{s_D} = \hat{\phi}$. Given that the time between

two substitutions is t_s , the *effective mutation-selection balance* during that time is

$$\sigma \int_0^{1/\sigma} \phi(t) dt = \frac{\mu_D}{s_D} \left(1 - \frac{\sigma}{s_D} (1 - e^{-s_D/\sigma}) \right), \quad (87)$$

where σ is the rate of adaptive substitutions as determined by Gerrish & Lenski (1997):

$$= \alpha \mu_B N \int_0^\infty \pi(s) \exp \left\{ -\frac{\mu_B}{s} N \ln(N) e^{-\alpha s} \pi(s + \frac{1}{\alpha}) - \alpha s \right\} ds \quad (88)$$

This expression makes the assumption that selection coefficients for beneficial mutations are exponentially distributed with parameter α . The function $\pi(s)$ describes the probability that a beneficial mutation of selective advantage, s , is not lost by drift. Assuming constant total population size and Poisson-distributed offspring, this function is approximately $\pi(s) \approx 2s$ (Haldane, 1927). For bacteria, which reproduce by binary fission, $\pi(s) \approx 4s$ (Gerrish & Lenski, 1998).

APPENDIX 4.2

EFFECTIVE POPULATION NUMBER UNDER A SERIAL TRANSFER REGIME

I derive the effective population number with respect to non-neutral mutations. First, I find the probability of fixation¹ of a beneficial mutation in a population of constant number. Then, I find this probability for a population that is subject to periodic bottlenecks and grows exponentially between these bottlenecks, as in a serial transfer regime. Equating these two probabilities gives an expression for effective population number. Conveniently, the same equations apply for deleterious mutations as well, implying that this effective population number is general for non-neutral mutations.

Employing a diffusion approximation, the Kolmogorov backward equation is solved to find the ultimate probability

1

By *probability of fixation*, we mean the probability that the mutant gene in question is not lost by drift. In an asexual system, the mutant gene may be lost as a result of competition with alternative beneficial mutant genes (Haigh, 1978; Gerrish & Lenski, 1997); such competition, however, becomes important only when frequencies become relatively high, at which point stochastic effects are negligible. Because effective population size is a stochastic equivalent, calculations here do not take such competition between beneficial mutations into account.

of fixation $u(p)$ given starting frequency p . (The Kolmogorov backward equation is derived in Chapter 8 of Crow & Kimura (1970).) Let $u(p, t)$ denote the probability that a mutant gene becomes fixed by the t^{th} generation, given that its starting frequency is p . Then, the diffusion approximation for this probability is given by

$$\frac{\partial u(p, t)}{\partial t} = \mu \frac{\partial u(p, t)}{\partial p} + \frac{1}{2} \sigma^2 \frac{\partial^2 u(p, t)}{\partial p^2} . \quad (89)$$

We are interested in the ultimate probability of fixation (probability that the mutant gene ever achieves fixation), given by $u(p) = \lim_{t \rightarrow \infty} u(p, t)$, for which $\partial u / \partial t = 0$ and which therefore satisfies

$$\mu \frac{du(p)}{dp} + \sigma^2 \frac{d^2 u(p)}{dp^2} = 0 , \quad (90)$$

with boundary conditions, $u(0) = 0$ and $u(1) = 1$. The solution which satisfies these boundary conditions is

$$u(p) = \frac{\int_0^p G(x) dx}{\int_0^1 G(x) dx} , \quad (91)$$

where $G(x) = \exp\left\{-\int \frac{2\mu}{\sigma^2} dx\right\}$. Let x_t denote frequency of the

mutant gene at time t , and let δx_t denote the change in x_t

between generations t and $t+1$, i.e. $x_{t+1} = x_t + \delta x_t$. Then μ

and σ^2 may be understood as expectations $E(\delta x_t)$ and

$E[(\delta x_t)^2]$, respectively. Given that the system is asexual

and that the mutant gene in question has selective advantage

s , the change in mean frequency is simply $\mu = N_e x s / N_e = s x$.

To calculate σ^2 , consideration must be given to both the mode of reproduction as well as the predominant mode of selection. At this point, I restrict the derivation to the case of reproduction by binary fission; conveniently,

however, the result approximates that for other modes of reproduction, implying generality. If selection occurs

mainly by differential growth, then the change in variance

is $\sigma^2 = 2N_e x(1+s) \frac{1}{2} (1 - \frac{1}{2}) / N_e^2 = \frac{x(1+s)}{2N_e}$, the binomial variance

in which $2N_e x(1+s)$ offspring are sampled with probability

$\frac{1}{2}$. If selection occurs mainly by differential death (or differential sampling), then the change in variance is

$$\sigma^2 = 2N_e x \left(\frac{1+s}{2} \right) \left(\frac{1-s}{2} \right) / N_e^2 = \frac{x(1-s^2)}{2N_e}, \text{ the binomial variance in}$$

which $2N_e x$ offspring are sampled with probability $(1+s)/2$.

Thus, $\frac{2\mu}{\sigma^2} = 4N_e f(s)$, where $f(s) = \frac{s}{1+s}$ for differential growth

or $f(s) = \frac{s}{1-s^2}$ for differential death, and (91) becomes

$$u(p) = \frac{1 - \exp\{-4N_e f(s)p\}}{1 - \exp\{-4N_e f(s)\}}. \quad (92)$$

Given ploidy number, ϕ , the starting frequency of an individual mutant gene is $p = 1/(\phi N)$, where N is the actual number of individuals present in the population at the time of mutation. When this starting frequency is inserted into (92), the probability of fixation of an individual mutant gene is closely approximated by

$$u \approx 4s \left(\frac{N_e}{\phi N} \right), \quad (93)$$

regardless of the choice of $f(s)$. (The sexual case, assuming Poisson-distributed offspring, yields the same approximation; see Crow & Kimura, 1970.) Rearrangement of this equation gives an expression for effective population number, $N_e = \frac{u\phi N}{4s}$.

If the actual population number, N , fluctuates over time, as in a serial transfer regime, then the effective population number will also fluctuate. To leave the effective population number as a function of time, however, would defeat the purpose of having an effective population number, which is to simplify the math. Thus, we employ the geometric mean effective population number,

$$\ln(\bar{N}_e) = \frac{1}{\tau} \int_0^{\tau} \ln \left(\frac{u\phi N_t}{4s} \right) dt, \text{ where } \tau \text{ is the period of time}$$

between serial transfers.

If the population grows exponentially between transfers, then $N_t = N_0 e^{rt}$, where r is the exponential

growth parameter. For continuous growth of bacteria, $r = 1$; for discrete growth of bacteria, $r = \ln 2$.

It remains to derive the probability of fixation, u . Let v denote the probability that the beneficial mutation in question does not survive the effects of random sampling, such that $u = 1 - v$. The number of offspring of the beneficial mutant that make it through the first population bottleneck (due to the first transfer dilution) is a Poisson random variable. The Poisson parameter of this random variable is the product of (i) the number of offspring of the beneficial mutant just before dilution, and (ii) the dilution factor, D . Given that the beneficial mutant appears at time t , factor (i) is $e^{r(1+s)(\tau-t)}$. Thus, the Poisson parameter for the number of offspring of the beneficial mutant that are sampled at the first dilution is $e^{r(1+s)(\tau-t)}D$. For convenience, I break this parameter into two factors, $\lambda = e^{r(1+s)\tau}D$ and $\gamma_t = e^{-r(1+s)t}$.

The total probability of loss of the beneficial mutation is then given by the following logic. Either zero offspring are transferred at the first dilution or one

offspring is transferred and lost in subsequent dilutions or two offspring are transferred and both lineages are lost in subsequent dilutions or . . . etc. The probability that zero offspring are transferred at the first dilution is $e^{-\lambda\gamma_t}$. The probability that one offspring is transferred at the first dilution is $\lambda\gamma_t e^{-\lambda\gamma_t}$. Let x denote the probability that one lineage is lost in subsequent dilutions. Then the probability that one offspring is transferred at the first dilution and its lineage is lost in subsequent dilutions is $\lambda\gamma_t e^{-\lambda\gamma_t} x$. Likewise, the probability that two offspring are transferred at the first dilution and both lineages are lost in subsequent dilutions is $\frac{1}{2}(\lambda\gamma_t)^2 e^{-\lambda\gamma_t} x^2$. The same logic applies for any number of offspring transferred at the first dilution. Thus the total probability of loss is given by the sum,

$$v = \sum_{j=0}^{\infty} \frac{(\lambda\gamma_t)^j}{j!} e^{-\lambda\gamma_t} x^j = e^{\lambda\gamma_t(x-1)} . \quad (94)$$

The remaining unknown is x , the probability that a single lineage is lost in subsequent dilutions.

If one beneficial mutant is transferred at the first dilution, then the number of its offspring present just before the next dilution is $e^{r(1+s)\tau}$, such that the Poisson parameter for the number of its offspring that make it through that dilution is $e^{r(1+s)\tau}D$, or simply λ . Each of those offspring that make it through that dilution will have the same Poisson distribution of offspring after the next dilution, etc. This is called a branching process (see p. 58 in Bailey, 1964). A result of branching process theory is that the probability of extinction is given by the smallest positive root of the equation $g(x) = x$, where $g(x)$ is the probability generating function for the number of offspring produced by an individual in one "generation". If we define one "generation" as starting just after one dilution and ending just after the next, then the probability generating function for the number of offspring left by one individual after one "generation" is

$$g(x) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} x^j = e^{\lambda(x-1)}. \quad \text{Thus, the lineage started by}$$

one beneficial mutant transferred at the first dilution is lost in a subsequent dilution with probability x , the smallest positive root of the equation, $e^{\lambda(x-1)} = x$. When this probability is determined, then the total probability of survival may be calculated:

$u = 1 - v = 1 - e^{\lambda\gamma_t(x-1)} = 1 - x^{\gamma_t}$. From here, the effective population number may be calculated:

$$\bar{N}_e = \exp \left\{ \frac{1}{\tau} \int_0^{\tau} \ln \left[\frac{\phi N_0}{4s} e^{r' (1-x^{\gamma_t})} \right] dt \right\} . \quad (95)$$

When $|\lambda\gamma_t(x-1)|$ is small, then the following approximations are satisfactory for intermediate values of s . The probability of survival may be approximated by $u = 1 - e^{\lambda\gamma_t(x-1)} \approx \lambda\gamma_t(1-x)$ (from a first order expansion), where x is approximately $x \approx 1 - 2rs\tau$ (from a second order expansion of the equation $e^{\lambda(x-1)} = x$). Insertion of these two approximations into (95) yields the following simplified expression:

$$\bar{N}_e \approx \frac{1}{2} \phi N_0 r \tau \quad . \quad (96)$$

If τ is given in generations, this expression is further simplified:

$$\bar{N}_e \approx \frac{1}{2} \phi N_0 \tau \quad . \quad (97)$$

LIST OF REFERENCES

LIST OF REFERENCES

- Bailey, N.T.J., 1964. The Elements of Stochastic Processes. John Wiley & Sons, Inc., New York.
- Barton, N.H., 1993. The probability of fixation of a favoured allele in a subdivided population. Genet. Res. 62: 149-157.
- Barton, N.H., 1994. The reduction in fixation probability caused by substitutions at linked loci. Genet. Res. 64: 199-208.
- Barton, N.H., 1995. Linkage and the limits to natural selection. Genetics 140: 821-841.
- Berg, O.G., 1995. Periodic selection and hitchhiking in a bacterial population. J. Theor. Biol. 173: 307-320.
- Bhatnagar, S.K. & M.J. Bessman, 1988. Studies on the mutator gene, *mutT*, of *Escherichia coli*. Molecular cloning of the gene, purification of the gene product, and identification of a novel nucleoside triphosphatase. J. Biol. Chem. 263: 8953-8957.
- Boucher, C.A.B., E. O'Sullivan, J.W. Mulder, C. Ramautarsing, P. Kellam, G. Darby, J.M.A. Lange, J. Goudsmit & B.A. Larder, 1992. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. Journal of Infectious Diseases 165: 105-110.
- Carlton, B.C. & B.J. Brown, 1981. Manual of Methods for General Bacteriology, ed. P. Gerhardt. American Society for Microbiology, Washington, D.C., p. 222-242.

- Chao, L. & E.C. Cox, 1983. Competition between high and low mutating strains of *Escherichia coli*. *Evolution* 37: 125-134.
- Coffin, J.M., 1995. HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis, and therapy. *Science* 267: 483-489.
- Cox, E.C., 1976. Bacterial mutator genes and the control of spontaneous mutation. *Annu. Rev. Genet.* 10: 135-156.
- Cox E. C. & T.C. Gibson, 1974. Selection for high mutation rates in chemostats. *Genetics* 77: 169-84.
- Crow, J.F. & M. Kimura, 1965. Evolution in sexual and asexual populations. *Am. Nat.* 99: 439-450.
- Crow, J.F. & M. Kimura, 1970. *An Introduction to Population Genetics Theory*. New York: Harper & Row.
- Cunningham, C.W., K. Jeng, J. Husti, M. Badgett, I.J. Molineux, D.M. Hillis & J.J. Bull, 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Molecular Biology & Evolution* 14: 113-116.
- Drake, J.W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* 88: 7160-7164.
- Drake, J.W., 1991. Spontaneous mutation. *Annu. Rev. Genet.* 25: 125-46.
- Elena, S.F., L. Ekunwe, N. Hajela, S.A. Oden & R.E. Lenski, 1998. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica*, in press.
- Elena, S.F., V.S. Cooper & R.E. Lenski, 1996. Punctuated evolution caused by selection of rare beneficial mutations. *Science* 272: 1802-1804.
- Ewens, W.J., 1979. *Mathematical Population Genetics*. New York: Springer-Verlag.

- Ewens, W.J., 1969. Population Genetics. London: Methuen Press.
- Feller, W., 1968. An Introduction to Probability Theory and Its Application. New York: John Wiley & Sons.
- Felsenstein, J., 1988. Sex and the evolution of recombination, pp. 74-86 in *The Evolution of Sex*, edited by R.E. Michod and B.R. Levin. Sunderland, Mass.: Sinauer Associates.
- Felsenstein, J., 1974. The evolutionary advantage of recombination. *Genetics* 78: 737-756.
- Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Oxford: Oxford Univ. Press.
- Gerrish, P.J. & R.E. Lenski, 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* (in press).
- Gillespie, J.H., 1991. The Causes of Molecular Evolution. Oxford: Oxford Univ. Press.
- Gillespie, J.H., 1981. Mutation rate modification in a random environment. *Evolution* 35: 468-476.
- Gurland, J., 1958. *Biometrics* 14: 229-249.
- Gurland, J., 1963. A method of estimation for some generalized Poisson distributions. International Symposium on Classical and Contagious Discrete Distributions, McGill, Montreal.
- Haigh, J., 1978. The accumulation of deleterious genes in a population -- Muller's ratchet. *Theor. Pop. Biol.* 14: 251-267.
- Haldane, J.B.S., 1927. The mathematical theory of natural and artificial selection. *Proc. Camb. Phil. Soc.* 23: 838-844.

- Hall, B.G., 1988. Adaptive evolution that requires multiple spontaneous mutations. I. Mutations involving an insertion sequence. *Genetics* 120: 887-897.
- Ho, D.D., T. Moudgil & M. Alam, 1989. Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *New England Journal of Medicine* 321: 1621-1625.
- Hogg, R.V. & A.T. Craig, 1995. Introduction to mathematical statistics. New Jersey: Prentice Hall.
- Holmes, E.C., L.Q. Zhang, P. Simmonds, C.A. Ludlam & A.J.L. Brown, 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc. Natl. Acad. Sci. USA* 89: 4835-4839.
- Horiuchi, T., H. Maki, M. Maruyama, & M. Sekiguchi, 1981. Identification of the *dnaQ* gene product and location of the structural gene for RNase H of *Escherichia coli* by cloning of the genes. *Proc. Natl. Acad. Sci. USA* 78: 3770-3774.
- Ishii, K., H. Matsuda, Y. Iwasa & A. Sasaki, 1989. Evolutionary stable mutation rate in a periodically changing environment. *Genetics* 121: 163-174.
- Jaeger, G. & S. Sarkar, 1995. On the distribution of bacterial mutants: the effects of differential fitness of mutants and non-mutants. *Genetica* 96: 217-223.
- Johnson, P., R.E. Lenski & F. Hoppensteadt, 1995. Theoretical analysis of divergence in mean fitness between initially identical populations. *Proceedings of the Royal Society, London B* 259: 125-130.
- Kaplan, N.L., R.R. Hudson & C.H. Langley, 1989. The "hitchhiking" effect revisited. *Genetics* 123: 887-889.
- Keightley, P.D., 1991. Genetic variance and fixation probabilities at quantitative trait loci in mutation-selection balance. *Genet. Res.* 58: 139-144.

- Kimura, M., 1960. Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. *J. Genet.* 57: 21-34.
- Kimura, M., 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* 9: 23-34.
- Kimura, M., 1979. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc. Natl. Acad. Sci. USA* 76: 3440-3444.
- Kuhner, M.K., J. Yamato & J. Felsenstein, 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140: 1421-1430.
- Lea, D.E. & C.A. Coulson, 1949. The distribution of the numbers of mutants in bacterial populations. *J. Genetics* 49: 264-285.
- LeClerc, J.E., B. Li, W.L. Payne, & T. Cebula, 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274: 1208-1211.
- Lederberg, S., 1966. Genetics of host-controlled restriction and modification of deoxyribonucleic acid in *Escherichia coli*. *J. Bacteriol.* 91: 1029-1036.
- Leigh, E.G., 1970. Natural selection and mutability. *Am. Nat.* 104: 301-305.
- Leigh, E.G., 1973. The evolution of mutation rates. *Genetics* 73 (suppl.): s1-s18.
- Leigh Brown, A.J. & D.D. Richman, 1997. HIV-1: gambling on the evolution of drug resistance? *Nature Medicine* 3: 268-271.
- Leigh Brown, A.J., 1997. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. USA* 94: 1862-1865.

- Lenski, R.E. & M. Travisano, 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA* 91: 6808-6814.
- Lenski, R.E., M.R. Rose, S.C. Simpson & S.C. Tadler, 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *Am. Nat.* 138: 1315-1341.
- Luria, S. E. & M. Delbrück, 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28: 491-511.
- Ma, W.T., G.H. Sandri, & S. Sarkar, 1992. Analysis of the Luria-Delbrück distribution using discrete convolution powers. *J. Appl. Prob.* 29: 255-267.
- Manning, J.T. & D.J. Thompson, 1984. Muller's ratchet accumulation of favourable mutations. *Acta Biotheor.* 33: 219-225.
- Mansky, L.M. & H.M. Temin, 1995. Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* 69: 5087-5094.
- Mao, E.F., L. Lane, J. Lee & J.H. Miller, 1997. Proliferation of mutators in A cell population. *J. Bacteriol.* 179: 417-422.
- Mascolini, M., 1997. Of clades and quasispecies: making sense of the HIV population census. *Journal of the International Association of Physicians in AIDS Care* 3: 13-25.
- Maynard Smith, J. & J. Haigh, 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23: 23-35.
- Maynard Smith, J., 1968. Evolution in sexual and asexual populations. *Am. Nat.* 102: 469-473.
- Miller, J.H., 1992. *A Short Course in Bacterial Genetics*. Cold Spring Harbor Lab. Press, Plainview, NY.

- Miller, J.H., 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu. Rev. Microbiol.* 50: 625-643.
- Mittler J.E. & R.E. Lenski, 1992. Experimental evidence for an alternative to directed mutation in the *bgl* operon. *Nature* 356: 446-448.
- Modrich, P., 1995. Mismatch repair, genetic stability and tumour avoidance. *Phil. Trans. Roy. Soc. Lond. Ser. B.* 347: 89-95.
- Modrich, P., 1991. Mechanisms and biological effects of mismatch repair. *Ann. Rev. Genet.* 25: 229-253.
- Moxon E.R., P.B. Rainey, M.A. Nowak & R.E. Lenski, 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current Biology* 4: 24-33.
- Muller, H.J., 1932. Some genetic aspects of sex. *Am. Nat.* 8: 118-138.
- Muller, H.J., 1964. The relation of recombination to mutational advance. *Mutat. Res.* 1: 2-9.
- Nowell, P.C., 1974. The clonal evolution of tumor cell populations. *Science* 194: 23-28.
- Otto, S.P. & M.C. Whitlock, 1997. The probability of fixation in populations of changing size. *Genetics* 146: 723-733.
- Pamilo, P., M. Nei & W. Li, 1987. Accumulation of mutations in sexual and asexual populations. *Genet. Res.* 49: 135-146.
- Pang, P.P., A.S. Lundberg & G.C. Walker, 1985. Identification and characterization of the *mutL* and *mutS* gene products of *Salmonella typhimurium* LT2. *J. Bacteriol.* 163: 1007-1015.
- Peck, J.R., 1994. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* 137: 597-606.

- Peck, J.R., G. Barreau & S.C. Heath, 1997. Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* 145: 1171-1199.
- Piatak, M., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hahn, G.M. Shaw & J.D. Lifson, 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 259: 1749-1755.
- Sambrook, E.F., T. Fritsch & J. Maniatis, 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Plainview, NY.
- Sarkar, S., 1991. Haldane's solution of the Luria-Delbruck distribution. *Genetics* 127: 257-261.
- Sarkar, S., W.T. Ma & G.v.H. Sandri, 1992. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* 85: 173-179.
- Schaaper, R.M. & R. Cornacchio, 1992. An *Escherichia coli* *dnaE* mutation with suppressor activity toward mutator *mutD5*. *J. Bacteriol.* 174: 1974-1982.
- Sniegowski, P.D., P.J. Gerrish & R.E. Lenski, 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387: 703-705.
- Stewart, F.M., D.M. Gordon & B.R. Levin, 1990. Fluctuation analysis: the probability distribution of the number of mutants under different conditions. *Genetics* 124: 175-185.
- Stewart, F.M., 1994. Fluctuation tests: how reliable are the estimates of mutation rates? *Genetics* 137: 1139-1146.
- Taddei, F., M. Radman, J. Maynard Smith, B. Toupance, P.H. Gouyon & B. Godelle, 1997. Role of mutator alleles in adaptive evolution. *Nature* 387: 700-702.

- Taucher-Scholz, G. & H. Hoffman-Berling, 1983.
Identification of the gene for DNA helicase II of
Escherichia coli. Eur. J. Biochem. 137: 573-580.
- Travisano, M. & R.E. Lenski, 1996. Long-term experimental
evolution in *Escherichia coli*. IV. Targets of selection
and the specificity of adaptation. Genetics 143: 15-
26.
- Tröbner, W. & R. Piechocki, 1984. Competition between
isogenic *mutS* and *mut⁺* populations of *Escherichia coli*
K12 in continuously growing cultures. Mol. Gen. Genet.
198: 175-176.
- Vasi, F., M. Travisano & R.E. Lenski, 1994. Long-term
experimental evolution in *Escherichia coli*. II. Changes
in life-history traits during adaptation to a seasonal
environment. Am. Nat. 144: 432-456.
- Wright, S., 1982. Character change, speciation, and the
higher taxa. Evolution 36: 427-443.

MICHIGAN STATE UNIV. LIBRARIES



31293017879895