

Genetic diversity in changing environments

A Thesis

Submitted for the Degree of
DOCTOR OF PHILOSOPHY

by

SACHIN KAUSHIK



THEORETICAL SCIENCES UNIT
JAWAHARLAL NEHRU CENTRE FOR ADVANCED SCIENTIFIC RESEARCH
(A Deemed University)
Bangalore – 560 064

January 2023

To my family

DECLARATION

I hereby declare that the matter embodied in the thesis entitled “**Genetic diversity in changing environments**” is the result of investigations carried out by me at the Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under the supervision of **Prof. Kavita Jain**, and that it has not been submitted elsewhere for the award of any degree or diploma.


In keeping with the general practice in reporting scientific observations, due acknowledgment has been made whenever the work described is based on the findings of other investigators.



Sachin Kaushik

CERTIFICATE

I hereby certify that the matter embodied in this thesis entitled “**Genetic diversity in changing environments**” has been carried out by **Mr. Sachin Kaushik** at the Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under my supervision and that it has not been submitted elsewhere for the award of any degree or diploma.



Prof. Kavita Jain
(Research Supervisor)

Acknowledgments

I am grateful to all of the wonderful people who have guided, supported, and accompanied me on this journey. This Ph.D. thesis is the outcome of the efforts of everyone who has helped make this research journey a success in so many ways. First and foremost, I would like to thank my research supervisor, **Prof. Kavita Jain**, for providing me with a home away from home. Throughout my Ph.D., her excellent guidance, constant motivation, and invaluable suggestions helped me grow every day. I am extremely fortunate to have her as my supervisor, who has provided me with unwavering support throughout my academic career and beyond, who cared so much about my work, who responded to my question and doubts so promptly, who has given me sufficient freedom to work on the topics of my choice, and who has patiently monitored my progress. I will be eternally grateful to her for making my research journey an enjoyable learning experience for me.

I acknowledge the financial support and excellent research facilities of JNCASR. I would like to express my sincere gratitude to Prof. Kavita Jain, Prof. Srikanth Sastry, Dr. Meher K. Prakash, Prof. Santosh Anshumali, Dr. Diwakar and Prof. Subir K. Das for the wonderful courses they provided. I am thankful to all the faculty members of TSU for their support in various ways. I would also like to thank my GSAC members Prof. Vidhyadhiraja N S and Dr. Deepa Agashe for their support and valuable discussions.

I would like to thank all of my high school and college teachers, as well as everyone else who has helped me grow in knowledge.

I thank all my past and present labmates Priyanka, Sayani, Archana, Himanshu, Lakshita and Sayantan for their help and support. I express my special thanks to Sayani and Archana for being supportive and caring. I am grateful to my good friend and former roommate Himanshu for always being there to help me out. I am deeply thankful to Harshita, and Vikram for their support in both good and bad times. I extend my special thanks to Vanshika, Saurabh, Keshav, Vinita, Ankit, Abhishek, Eklavya, Swaraj, Preeti, Narendra, Divya, Aditya, Akshay, Surishi, Rohit, Shalu, and Anjaney for their support. I truly appreciate the assistance I received from Nalina and Archana in resolving technical challenges that arose during the thesis compilation process, and they were always available and willing to assist.

I can not thank my parents and sister Nisha enough, who have always led me in the right direction. They have been a constant source of inspiration and energy that has kept me going every day. It would not have been possible to complete this journey without their love and support.

Last but not least, I appreciate the help and support I have received from all the academic and non-academic staff in JNCASR.

Synopsis

Genetic diversity is the frequency of genetically non-identical individuals in a population, and is the raw material on which evolution acts. For this reason, it plays a key role in understanding the evolutionary history and future evolutionary changes that may occur in the population. For example, when there is a change in the environment, genetic diversity helps the population to avoid the risk of extinction by providing opportunities for adaptation to occur. However, a large body of work assumes that the environment remains constant over evolutionary time scale, but since the natural environments continually change with time (for example, due to seasonal variations), it is vital to understand the effect of changing environment on genetic diversity and is the subject of this thesis.

An important statistic for measuring the within-population genetic diversity is the site frequency spectrum, which gives unnormalised allele frequency distribution across polymorphic sites in the genome, and is related to the heterozygosity, which is the frequency of individuals carrying a different gene (relative to the reference) at the same position in the chromosome. In this thesis we study these two quantities in detail when the environment is changing in time. The thesis is divided into five chapters, and a brief description of each chapter is given below.

Chapter 1 introduces the primary evolutionary forces, namely, natural selection, mutation, and random genetic drift, that affect genetic diversity. We discuss three theoretical models, i) birth-death process, ii) the infinite sites model, and iii) the Wright-Fisher model, that are pertinent to the discussion in this thesis and are commonly used in the population genetics literature to quantify the combined effects of evolutionary forces. We also describe the mathematical framework of diffusion theory (Fokker-Planck equation) which can be used for any type of mutation, and a semi-deterministic approximation for the beneficial mutation. Besides numerical simulations, these mathematical frameworks

are used to obtain the fixation time, the site frequency spectrum, and heterozygosity in the following chapters.

Chapter 2 deals primarily with the reduction in the genetic diversity at a linked neutral locus due to the fixation of beneficial or deleterious mutant (selective sweep) where the selection coefficient changes periodically with time. We consider a constant size, randomly mating, diploid population in which one locus is under selection and model its dynamics using a birth-death process. We study beneficial sweeps using semi-deterministic approximation and deleterious sweeps using the diffusion theory framework. It is known that in the static environment, the conditional mean fixation time for a codominant mutant is identical for the beneficial and deleterious mutant (Maruyama-Kimura symmetry). We find that the Maruyama-Kimura symmetry does not hold in the changing selective environment, even when the environment changes slowly. For a meaningful comparison with the results in the static environment, we focus on the slowly changing environment, and find that i) the conditional mean fixation time for the beneficial mutant in changing environment is well approximated by the corresponding result in the constant environment; ii) but for the recessive deleterious mutant, the fixation time is substantially different from that in the static environment. Since the diversity patterns are intimately related to the fixation times, the changing environment strongly impacts the diversity patterns due to the deleterious sweeps.

In **Chapter 3**, we relax the constant population size assumption and the population is modeled according to the infinite sites model. We study the joint effect of time-dependent selection coefficient and demography on the site frequency spectrum (SFS) and mean heterozygosity. We derive simple analytical expressions for the equilibrium SFS in the static environment, and the time-averaged SFS when the environment changes either slowly or rapidly compared to the other time scales in the model. The main results are, in the slowly changing environment, when selection strength is strong and the mutant

experiences both positive and negative selection, i) the time-averaged SFS is significantly different than the equilibrium SFS in the static environment with time-averaged selection coefficient, and ii) the difference depends upon the amount of time spent by the population in the deleterious part of the cycle, and the phase difference between the population size and selection coefficient.

In **Chapter 4**, we relax the random mating assumption in previous chapters, and discuss the reduction in genetic diversity at the linked neutral site due to the fixation of the mutant at the selected locus when inbreeding occurs. We find that the effect of slowly changing environment on the mean conditional fixation time due to deleterious and beneficial sweeps is strongest for randomly mating populations. Furthermore, we explore the effect of changing environment on strongly deleterious mutations. Due to strong purifying selection, deleterious mutations are eliminated and result in a reduction in linked neutral genetic diversity (background selection). We therefore study the effect of recurrent deleterious mutations on the linked neutral sites in an asexual population which is described by a Wright Fisher model. We find that for the slowly changing environment, i) when the population fluctuates between zero and negative selection coefficient, the time-averaged SFS and mean heterozygosity are always larger than that in the static environment with the time-averaged selection coefficient, and ii) when the selection coefficient changes but is always negative, the time-averaged SFS and mean heterozygosity are larger (smaller) than that in the static environment with time-averaged selection coefficient when the population is in slow (fast) Muller's ratchet regime.

Chapter 5 concludes the thesis with a discussion of main results and some interesting open questions related to our study.

List of publications

- S. Kaushik and K. Jain. Time to fixation in changing environments. *Genetics* **219**: iyab148, 2021.
- K. Jain and S. Kaushik. Joint effect of changing selection and demography on the site frequency spectrum. *Theor. Popul. Biol.* **146**: 46–60, 2022
- S. Kaushik. Effect of beneficial sweeps and background selection on genetic diversity in changing environments. *bioRxiv* 2022.08.29.505661, 2022

Materials in Chapters 2,3 and 4 have been reproduced from publication numbers 1,2 and 3, respectively.

Table of contents

List of figures	xvii
1 Introduction	1
1.1 Introduction	1
1.1.1 What shapes genetic diversity?	2
1.1.2 Lewontin’s paradox	3
1.2 Models	5
1.2.1 Birth-Death process	6
1.2.2 Infinite sites model	7
1.2.3 Wright-Fisher model	7
1.3 Mathematical frameworks	9
1.3.1 Diffusion theory	9
1.3.2 Semi-deterministic theory	10
1.4 Quantities of Interest	11
1.4.1 Conditional mean fixation time	11
1.4.2 Site frequency spectrum	12
1.5 Overview of the thesis	13
Bibliography	16

2	Time to fixation in changing environments	19
2.1	Introduction	19
2.2	Model	21
2.3	Conditional fixation time in changing environments	24
2.4	Diffusion theory in slowly changing environments	27
2.4.1	Weak selection	29
2.4.2	Moderate selection	31
2.5	Semi-deterministic theory for on-average beneficial mutants in a large population	34
2.5.1	Time-inhomogeneous Feller process	36
2.5.2	Fixation probability	37
2.5.3	Mean fixation time in slowly changing environments	38
2.6	Discussion	42
	Bibliography	47
3	Joint effect of changing selection and demography on the site frequency spectrum	51
3.1	Introduction	51
3.2	Model	53
3.3	Site frequency spectrum: diffusion theory	54
3.4	Results	56
3.4.1	Static environment	56
3.4.2	Slowly changing environment	60
3.4.3	Rapidly changing environment	70
3.5	Discussion	76
	Bibliography	80

4	Effect of beneficial sweeps and background selection on genetic diversity in changing environments	85
4.1	Introduction	85
4.2	Selective sweeps in changing environments	87
4.2.1	Diffusion theory	88
4.2.2	Semi-deterministic theory: strongly beneficial mutations	90
4.3	Background selection in changing environment	93
4.3.1	Multilocus Wright-Fisher model	94
4.3.2	Fluctuating environment between deleterious mutations	95
4.3.3	Fluctuating environment between neutral and deleterious mutations	99
4.4	Discussion	101
	Bibliography	104
5	Summary and open questions	108
	Appendix A	112
A.1	Diffusion theory for time-inhomogeneous process	112
A.2	Mean fixation time in slowly changing environments	114
A.3	Feller process with time-dependent coefficients	115
A.4	Distribution of the fixation time of on-average beneficial mutant	116
A.5	Mean fixation time of on-average beneficial mutant	117
A.6	Mean fixation time of initially deleterious mutant under strong selection	118
	Appendix Bibliography	122
	Appendix B	123
B.1	Equilibrium site frequency spectrum for strong selection	123
B.2	Eigenfunction expansion of the site frequency spectrum	125

- B.3 Nonequilibrium site frequency spectrum for neutral population 126
- B.4 Nonequilibrium site frequency spectrum for weak selection 130
- B.5 Mean heterozygosity for co-dominant, on-average neutral mutant 132
- B.6 Nonequilibrium site frequency spectrum for strong selection 133

- Appendix Bibliography** **136**

- Appendix C** **137**

 - C.1 Diffusion Theory 137
 - C.2 Semi-deterministic approximation 139

- Appendix Bibliography** **141**

List of figures

- 1.1 Random genetic drift: Either the allele A spreads into the entire population or it gets extinct eventually due to stochastic fluctuations in the allele frequencies. 2
- 1.2 Selective sweep: When the mutation ● spreads in the population and the recombination rate is small, the linked neutral mutants ● also get fixed, thereby reducing the genetic variation in the population. 3
- 1.3 Maruyama-Kimura symmetry: The conditional mean fixation times for a mutant with selection coefficient $-s$ (red) and s (green) are identical. . . 4
- 1.4 Background selection: The genetic sequences carrying strongly deleterious mutation ● are eliminated from the population, thereby reducing the neutral (●) genetic variation in the population. 5
- 1.5 Birth-Death Model: Birth-Death process is a continuous-time Markov process in which an individual gives birth and another one dies. The birth ($r_{b,i}$) and death rates ($r_{d,i}$) may depend on the population size and the number of individuals of type i 6
- 1.6 Wright-Fisher Model: Consider a population with constant size N where the individuals are of two types, (●) and (○). Each individual in offspring generation picks a parent with probability proportional to the parent's fitness. 8

-
- 1.7 Dynamical phases: The wild type frequency (red) and beneficial mutant allele frequency (green) initially evolve stochastically (phase A), followed by deterministic evolution (phase B), and again evolve stochastically when they are close to absorption (phase C). 10
- 1.8 Site frequency spectrum: Consider the genomic data set of 6 individuals and 8 sites where 0 and 1, respectively, denote the wild type and mutant allele. The number of observed polymorphic sites (S_i) with allele frequency i are then represented by a histogram which is generated by counting the number of mutant alleles present at each site. 12
- 1.9 Site frequency spectrum: Under the infinite sites model, the equilibrium site frequency spectrum decreases monotonically for neutral and deleterious mutations. For beneficial mutations, the site frequency spectrum is U-shaped because the beneficial mutation can rise to high frequency in the population due to their advantageous effect. 13

- 2.1 Top panel: Conditional mean fixation time \bar{T}_c of a dominant mutant ($h = 0.7$, \blacktriangle) and a recessive mutant ($h = 0.3$, \blacktriangledown) with selection coefficient $s(t) = \sigma \sin(\omega t + \pi/4)$ and $-s(t)$, respectively (see inset for $s(0)$) to show that the symmetry between the conditional mean fixation time for dominant beneficial and recessive deleterious mutant in static environments is not preserved in changing environments. The data are obtained by numerical simulations (closed symbols) and numerically integrating the diffusion theory equations (2.4) and (2.5) (open symbols) for small cycling frequencies in a population of size N . The conditional mean fixation time in the static environment with selection $|s(0)|$ (solid line) and in the neutral environment, given by $2N$, (dashed line) are obtained from diffusion theory and shown for comparison. Middle and Bottom panel: Mutant allele trajectories for (b) the initially deleterious mutant ($h = 0.3$) and cycling frequencies, $\omega = 5 \times 10^{-4}$ (black), 2×10^{-3} (magenta), and 4×10^{-3} (blue), and (c) the initially beneficial mutant ($h = 0.7$) for $\omega = 0$ (blue), 5×10^{-4} (black), and 2×10^{-3} (magenta). The allele trajectories in each panel are almost same at short times as they were started with the same random seed in the computer simulation. The smooth curves show the selection coefficient for the corresponding cycling frequency. In all the panels, $\sigma = 0.01$ and $N = 500$ so that $\alpha = N|s(0)| \approx 3.5$ 24

- 2.2 Conditional mean fixation time of an initially deleterious (\blacktriangledown) and initially beneficial (\blacktriangle) mutant with selection coefficient $s(t) = \bar{s} + \sigma \sin(\omega t + 3\pi/4)$ and $-s(t)$, respectively, obtained by numerical simulations in a population of size N . The conditional fixation time in static environment with selection $|s(0)|$ (solid) and in the time-averaged environment with selection $|\bar{s}|$ (dashed) are also shown for comparison, and obtained using diffusion theory. In both cases, $\bar{s} = 0.01, \sigma = 0.007$ and $N = 500$. The qualitative behavior of these curves can be understood using the arguments given in the text. 25
- 2.3 Scaled conditional mean fixation time, $v_c = \bar{T}_c/2N$ for a co-dominant mutant under weak selection in slowly changing, on-average neutral environment in which the selection coefficient $s(t) = \sigma \sin(\omega t + \theta)$. The parameter $\alpha = N|s(0)|$ was varied with selection amplitude σ , keeping the population size N and the initial phase $\theta = \pi/4$ (for positive α) and $5\pi/4$ (for negative α) fixed. The inset shows the variation of the conditional mean fixation time with dominance in static environment (solid) and slowly varying environment for initial phase $\theta = \pi/4$ (dotted) and $5\pi/4$ (dashed), and $\sigma = 0.000158$. In both plots, $N = 2 \times 10^3$ and $N\omega = 0.08$, and the lines show the analytical expressions (2.9) and (2.10) and the points show the numerical solution of (2.4)-(2.7). 30

- 2.4 Conditional mean fixation time \bar{T}_c for moderate selection in static (\bullet) and slowly changing environment with selection coefficient $s(t) = \sigma \sin(\omega t + \theta)$ (diamonds) and $-s(t)$ (squares) for different dominance coefficients and the initial phase $\theta = \pi/4$ (open symbols) and $3\pi/4$ (closed symbols) in a population of size N . The other parameters are $N = 2 \times 10^3$, $N\omega = 0.05$ and $\sigma = 0.01$. The inset depicts the arrival time of the mutant in all the cases. The data are obtained within the framework of diffusion theory by numerically solving (2.4)-(2.7). 32
- 2.5 Absolute value of the deviation between the conditional mean fixation times in slowly changing environment with selection coefficient $s(t) = \sigma \sin(\omega t + \theta)$ and static environment with selection coefficient $s(0)$ for beneficial ($\theta = \pi/4$, \blacklozenge) and deleterious mutants ($\theta = 5\pi/4$, \blacksquare) as a function of $|\alpha| = N|s(0)|$ for fixed dominance coefficient $h = 1/2$ and $\sigma = 0.01$ obtained within the framework of diffusion theory by numerically solving (2.4)-(2.7). 33
- 2.6 Top panel: Conditional mean fixation time \bar{T}_c for a mutant under strong positive selection at all times with selection coefficient $s(t) = \bar{s} + \sigma \sin(\omega t + \theta)$ when the mutant is beneficial at all times to show that, except for strongly recessive or dominant mutations, it depends weakly on the dominance coefficient h . The points are obtained by numerically calculating (2.18) for dominance coefficient $h = 0.1(\blacklozenge)$, $0.3(*)$, $0.5(\blacksquare)$, $0.7(\circ)$, $0.9(\bullet)$ in a population of size N . The other parameters are $N = 10^5$, $\bar{s} = 0.01$, $\sigma = 0.007$, $\theta = \pi/4$. The bottom panel shows the comparison between (2.18) (points) and (2.20) (line) for the deviation in the conditional mean fixation time in a slowly changing environment where $\omega \ll s(0) \approx 0.014$ 35

- 2.7 Fixation time distribution, $P(T_c)$ in a finite population of size N when the mutant is beneficial at all times and has a changing selection coefficient given by $s(t) = \bar{s} + \sigma(\omega t + \theta)$. The points and the curves are obtained, respectively, from numerical simulations and the semi-deterministic result (2.15) for cycling frequency below, above and close to the resonance frequency ω_r , and given by $\omega = 10^{-4}$ (\bullet , solid line), 0.1 (\circ , dashed line) and 0.002 (\square , dotted line), respectively. The other parameters are $N = 10^5$, $\theta = 0$, $\bar{s} = 0.01$, $\sigma = 0.007$, $h = 0.5$. Note that while the distribution is bell-shaped away from the resonance frequency, it is bimodal close to ω_r which results in a large variance in the conditional fixation time (see inset). For the cycling frequency ω_r , the selection coefficient completes a full cycle and the mutant typically fixes when selection is increasing resulting in the bimodal character of the distribution. For the same reason, the distribution at high frequency also has multiple modes but with very small amplitude. 39

- 2.8 Distribution of the conditional fixation time T_c in a finite population of size N when the mutant is beneficial at all times and has a changing selection coefficient given by $s(t) = \bar{s} + \sigma \sin(\omega t + \theta)$. The data are obtained within a semi-deterministic approximation and given by (2.15) when a mutant is beneficial at all times to show that it does not have the $h \leftrightarrow 1 - h$ symmetry in fast changing environments. The overlapping left curves for $\omega = 10^{-4}$ show the distribution for $h = 0.7$ (solid blue) and 0.3 (dashed red), and the right curves are for $\omega = 0.002$ where $h = 0.7$ (dashdotted blue) and 0.3 (dotted red). The other parameters are $N = 10^5, \theta = \pi/4, \bar{s} = 0.01, \sigma = 0.007$. Inset: Distribution of the fixation time of an initially beneficial mutant when the time-averaged selection is zero and the environment changes slowly. The solid curve is obtained from (2.15) and the points are generated from simulations. The parameters are $N = 10^5, \sigma = 0.007, \omega = 10^{-6}, h = 0.3, \theta = \pi/4$ 41

- 2.9 Reduction in the mean heterozygosity at the linked neutral locus due to the fixation of dominant ($h = 0.7, \blacktriangle$) and recessive mutant ($h = 0.3, \blacktriangledown$) with selection coefficient $s(t) = \sigma \sin(\omega t + \pi/4)$ and $-s(t)$, respectively, to show that the heterozygosity following the fixation of deleterious and beneficial mutant is significantly different in the changing environment, unlike in the static environment where they are identical. The reduction in the mean heterozygosity in the static environment with selection (solid line) and in the neutral environment (dashed line) are also shown for comparison. All the data are obtained by numerical simulations. The inset shows the heterozygosity reduction due to the fixation of dominant mutant ($h = 0.9, \blacktriangle$) and recessive mutant ($h = 0.1, \blacktriangledown$) in the slowly changing environment with selection coefficient $s(t) = \sigma \sin(\omega t + 3\pi/4)$ and $-s(t)$, respectively. The other parameters for both figures are $\sigma = 0.01$, $c = s(0)/10$, $Q_0 = 0$, $R_0 = 0.2002$ and $p_0 = 1/2N$, where c is the recombination probability, Q_0 and R_0 are, respectively, initial relative frequencies of a neutral allele in a chromosome with a and A allele on the first locus, and the population size $N = 500$ (main) and 1000 (inset). . . 45

- 3.1 Static environment: The top panel shows the site frequency spectrum of a large population for selection strength (a) $\bar{N}\bar{s} = 30$, and (b) $\bar{N}\bar{s} = -10$. The solid lines show the exact expression (3.12), while the dashed lines show the approximate expression (3.14a) and (3.14b) for beneficial and deleterious mutations, respectively. The bottom panel shows the mean heterozygosity calculated numerically using (3.12) for (c) a positively selected mutant which approaches the asymptotic value (3.16b) depicted by dashed lines, and (d) a negatively selected mutant that decreases towards zero as $1/\bar{N}|s|$ (black curve) for all h in accordance with (3.16c). In all the plots, the scaled mutation rate $\theta = 1$ 58
- 3.2 Changing environment and on-average neutral selection: The time-dependent site frequency spectrum (solid line) in a large population when the population size and selection coefficient (dashed line) change periodically in time with equal cycling frequency obtained by numerically solving (3.5) for (a) $\bar{N}\omega = 0.01$ and (b) $\bar{N}\omega = 500$. The result (3.18) obtained within adiabatic approximation (dotted line) is also shown for comparison in the slowly changing environment. The other parameters are $x = 0.2, \nu = 0.3, h = 1/2, \theta = 1$ and $\bar{N}\sigma = 100$ 62
- 3.3 Changing population size and no selection: The time-averaged sample site frequency spectrum when population size changes with time for sample size $n = 20$. The bars are obtained by numerically solving (3.5) and (3.9) for different cycling frequencies while the crosses and triangles show (B.3.17) and (B.3.24) for small and large cycling frequencies, respectively. 63

- 3.4 Changing environment and positive selection at all times: The time-averaged heterozygosity, \bar{H} as a function of selection strength when either the selection coefficient (indicated in the legend by $\bar{N}, s(t)$) or population size ($N(t), \bar{s}$) or both ($N(t), s(t)$) vary with time. The top panel shows the results in slowly changing environments ($\omega, \Omega \ll \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 0.01$, and the bottom panel in rapidly changing environments ($\omega, \Omega \gg \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 1400$. The points are obtained by numerically solving (3.5) and (3.10) for different scenarios as indicated in the legend for $\nu = 0.7$, $\sigma = \bar{s}/8$, $\theta = 1$, and $h = 1/2$. The broken line shows the results in the static environment that are obtained using (3.12). The solid lines show the analytical approximation (3.21) and (3.25) for weak and strong selection, respectively, in slowly changing environment, and (3.35) for weak selection in rapidly changing environment. 65
- 3.5 Changing environment and on-average neutral selection: The time-averaged mean heterozygosity, \bar{H} as a function of selection strength when either only the selection coefficient (indicated in the legend by $\bar{N}, s(t)$) or both selection coefficient and population size ($N(t), s(t)$) vary with time. The top panel shows the results in slowly changing environments ($\omega, \Omega \ll \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 0.01$, and the bottom panel in rapidly changing environments ($\omega, \Omega \gg \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 1400$. The points are obtained by numerically solving (3.5) and (3.10) for different scenarios as indicated in the legend for $\nu = 0.7$, $\theta = 1$, $h = 1/2$ and no phase difference between selection and population size variation. The broken line shows the results in the static environment that are obtained using (3.12). The solid lines show the analytical approximation (3.21) for weak selection, and (3.28) and (3.29) for strong selection in slowly changing environment. 67

- 3.6 Changing environment and on-average neutral selection: The time-averaged sample site frequency spectrum represented by bar chart and obtained by numerically solving (3.5) and (3.9). The top panel shows the results in slowly changing environment for $\bar{N}\omega = 0.01$ that are compared with equilibrium SFS for neutral and positively selected mutant in constant environment. The bottom panel shows the sample SFS in rapidly changing environments for two cycling frequencies which is compared with the equilibrium neutral SFS. The points in the top panel are obtained using (3.31). The other parameters are $\nu = 0.7$, $\omega = \Omega$, $\theta = 1$, $h = 1/2$, $\bar{N}\sigma = 100$, $n = 20$ and no phase difference between selection and population size variation. 71
- 3.7 Changing environment and on-average neutral selection: The time-averaged mean heterozygosity as a function of scaled environmental frequency for the dominance parameter $h = 0.3(\blacktriangledown)$, $0.5(\blacksquare)$ and $0.7(\blacktriangle)$ when both selection coefficient and population size change in time with equal frequency. The points are obtained by numerically solving (3.5) and (3.10), and the solid lines on the left show the analytical approximation (3.31) for slowly changing environment while the dashed line on the right is the equilibrium mean heterozygosity θ in the time-averaged environment. The other parameters are $\nu = 0.7$, $\theta = 1$, and $\bar{N}\sigma = 100$. For comparison, note that the mean heterozygosity in neutral constant environment is equal to $\theta = 1$ 76

-
- 4.1 The variation of conditional mean fixation time \bar{T}_c with dominance parameter (h) for moderate selection in static (\bullet) and slowly changing environment where selection coefficient is on-average neutral ($\bar{s} = 0$), and given by $s(t) = \sigma \sin(\omega t + \theta)$ (diamonds) and $-s(t)$ (squares) respectively. The other parameters are $N_e = 2 \times 10^3$, $N_e \omega = 0.05$, $\sigma = 0.01$, $\theta = \pi/4$, (a) $f = 0.2$ and (b) $f = 0.8$. When inbreeding is high, the conditional mean fixation time for both beneficial and deleterious is mildly affected in the slowly changing environment. 89
- 4.2 Strongly beneficial mutations: The deviation in the conditional mean fixation time, $\bar{T}_{1,c}$, in the slowly changing environment with selection coefficient $s(t) = (\bar{s} + \sigma \sin(\omega t + \pi/4))$ when the mutation is beneficial at all times for $f = 0$ (black), $f = 0.2$ (green), and $f = 0.8$ (magenta) to show that the deviation is maximum for randomly mating population. The points are obtained by numerically solving (C.2.1). The other parameters are $N_e = 10^5$, $\bar{s} = 0.01$, $\sigma = 0.007$ and $h = 0.5$ respectively. 91

- 4.3 Strictly deleterious mutations: The time-averaged site frequency spectrum for the static environment (blue) and for the slowly changing environment (red) when the selection coefficient is always negative and varies periodically with time period T_p between $\bar{s} + \sigma$ in first half of the cycle and $\bar{s} - \sigma$ in the second half of the cycle. The parameters are $N = 1000$, $Nu_n = 1$, $u_d = 0.08$, $\sigma = \bar{s}/2$, $T_p = 12000$, (a) $\bar{s} = 0.1$ and (b) $\bar{s} = 0.05$. The solid blue line is $(2Nu_n e^{-u_d/\bar{s}})/x$. The inset shows the time-dependent heterozygosity for the slowly changing environment (red) and static environment (blue), the solid lines represent heterozygosity for the neutral population with the corresponding effective population sizes. The population evolves under drift neutral mutations equilibrium for SFS till $t = 3000$, and after that, deleterious mutations are introduced. 95
- 4.4 Strictly deleterious mutations: The mean heterozygosity as a function of scaled selection coefficient for static environment (circles) and slowly changing environment with selection coefficient (triangles) varies between $\bar{s} + \sigma$ and $\bar{s} - \sigma$ with time period T_p . The other parameters are $N = 1000$ (open symbols), $N = 2000$ (closed symbols), $Nu_n = 1$, $u_d = 0.08$, $\sigma = \bar{s}/2$ and $T_p = 12000$ respectively. The solid lines represent mean heterozygosity for the neutral population with effective population size as $2Nu_n e^{-u_d/\bar{s}}$. The inset shows the relative change in the mean heterozygosity due to the changing environment as compared to the static environment for $N = 1000$. 98

- 4.5 Deleterious and neutral mutations: The time-averaged SFS, when the environment is static (blue) with selection coefficient (\bar{s}) and changes slowly (red) with selection coefficient varies between 0 (neutral) and $2\bar{s}$ (deleterious), and the Muller's ratchet clicks slowly in the deleterious cycle. The blue solid and blue dashed line represents SFS for the neutral mutation with population size N and $N_{eff} = Ne^{-u_d/\bar{s}}$ respectively. The red solid line represents the time-averaged SFS given as, $\frac{\theta}{2x} (1 + e^{-u_d/\bar{s}})$. The other parameters are $N = 1000$, $u_d = 0.08$, $u_n = 0.001$, $\bar{s} = 0.06$, and $\theta = 2Nu_n$. The inset shows the time-dependent heterozygosity in the slowly changing environment (red) which oscillates between the corresponding heterozygosity in the neutral population with size N and N_{eff} respectively. For the static environment (blue), the heterozygosity in the equilibrium population is given by $\theta e^{-u_d/\bar{s}}$. The solid lines from top to bottom represents θ , $\theta e^{-u_d/2\bar{s}}$ and $\theta e^{-u_d/\bar{s}}$ respectively. The population evolves under drift neutral mutations equilibrium for SFS till $t = 3000$, and after that, deleterious mutations are introduced. 99
- 4.6 Deleterious and neutral mutations: The relative change in the mean heterozygosity in the changing environment (\bar{H}) from the static environment (H) when selection coefficient varies periodically between zero (neutral) and $2\bar{s}$ (deleterious) with time period T_p . The other parameters are $N = 1000$, $Nu_n = 1$, $u_d = 0.8$ and $T_p = 12000$ respectively. The inset shows that the time-averaged heterozygosity behaves non-monotonically with the scaled selection coefficient and the minimum of the mean heterozygosity scales linearly with $N\bar{s}$. The other parameters are $N = 1000$ (\bullet), $N = 1500$ (\blacktriangle) and $N = 2000$ (\blacktriangledown) respectively. The solid lines represents $2Nu_n e^{-u_d/\bar{s}}$ for $N = 1000, 1500$, and 2000 from top to bottom. 101

Chapter 1

Introduction

1.1 Introduction

Natural populations are typically subject to two opposing forces: on the one hand, the population has a tendency to march towards specific types that are advantageous in the given environment, and on the other hand, the population should have genetic diversity in order to provide the potential to cope with changing environments. A classic example where genetic diversity prevented the extinction of the population is observed in peppered moths during the industrial revolution [1].

The causes and maintenance of genetic variation can be studied in the framework of population genetics which involves modeling the temporal and spatial changes in the allele frequencies over a period of time. A central goal of population genetics is to examine the genetic make up of the population under the effect of selection, mutation, recombination, random genetic drift and other evolutionary factors. The development of mathematical models in population genetics has significantly refined our understanding of how evolution works at the genetic level, and provided insights into the mechanisms of the evolution which are sometimes very far from the intuitive expectations. One such

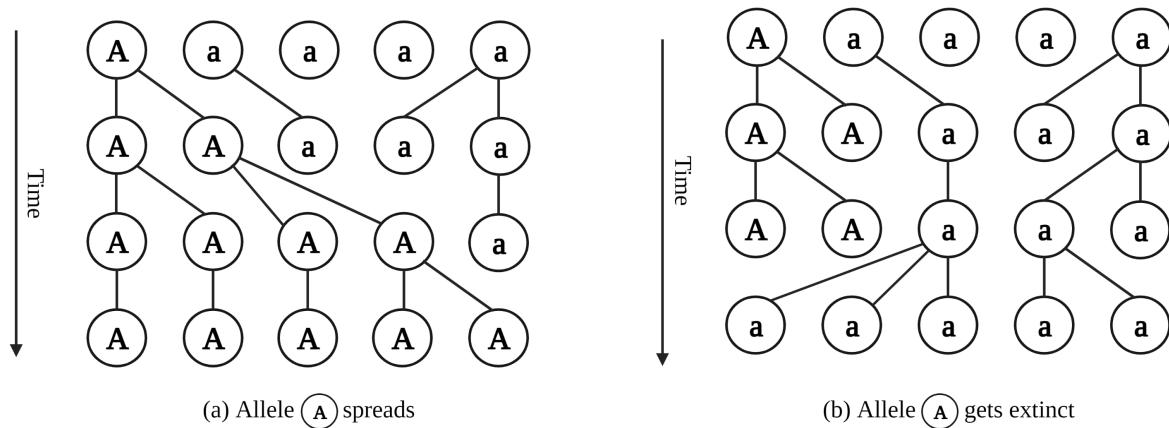


Fig. 1.1 Random genetic drift: Either the allele A spreads into the entire population or it gets extinct eventually due to stochastic fluctuations in the allele frequencies.

counterintuitive result related to the time taken by the mutations to spread in the entire population is discussed in Chapter 2.

1.1.1 What shapes genetic diversity?

While mutations and genetic recombination increase the genetic variation within a population, natural selection and genetic drift decrease the genetic diversity, as explained below.

Mutation is the cause of new variation that can arise into the population, and considered as the ultimate source of genetic variation. Mutations can arise due to occasional random errors in DNA replication or due to the physical damage to the DNA. The mutational effects on the fitness of an individual can be harmful (deleterious mutations), advantageous (beneficial mutations), or null (neutral mutations).

Genetic recombination is the mixing of the maternal and paternal chromosomes during meiosis, and introduces genetic variation in the population.

Natural selection is the difference in the reproductive output among the individuals. The individuals which are fit in a particular environment are most likely to survive and

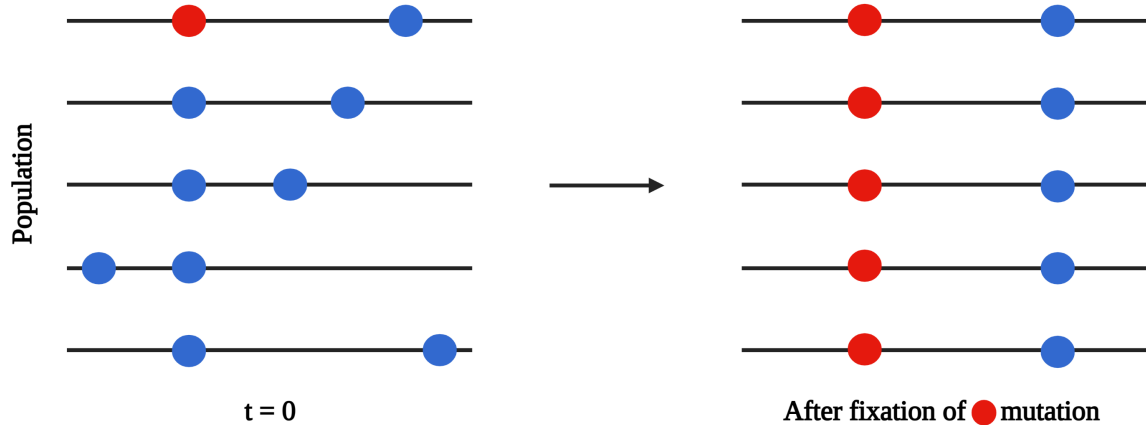


Fig. 1.2 Selective sweep: When the mutation ● spreads in the population and the recombination rate is small, the linked neutral mutants ● also get fixed, thereby reducing the genetic variation in the population.

produce more offspring as compared to the individuals which are less fit. Thus natural selection reduces the genetic diversity by favoring or disfavoring a specific allele.

Random genetic drift is the change in the allele frequencies due to random sampling effects. The fact that the two equally fit individuals can have different number of offspring makes the randomness an important component of the population genetics theory. The consequence of random genetic drift is illustrated in Fig. 1.1. Consider a population of constant size N with haploid individuals, and two segregating neutral alleles, a and A . If there are no mutations, eventually either the neutral allele a spreads into the entire population (fixation) and replaces the wild type allele A , or it is lost from the population. Thus, random genetic drift results in the reduction of the genetic variation initially present in the population.

1.1.2 Lewontin's paradox

In the absence of selection, the new variation is introduced in a population through mutations at rate μ , and the random genetic drift eliminates the new mutations at a rate $1/N$ in a population with size N . Under the action of these two opposite forces, the

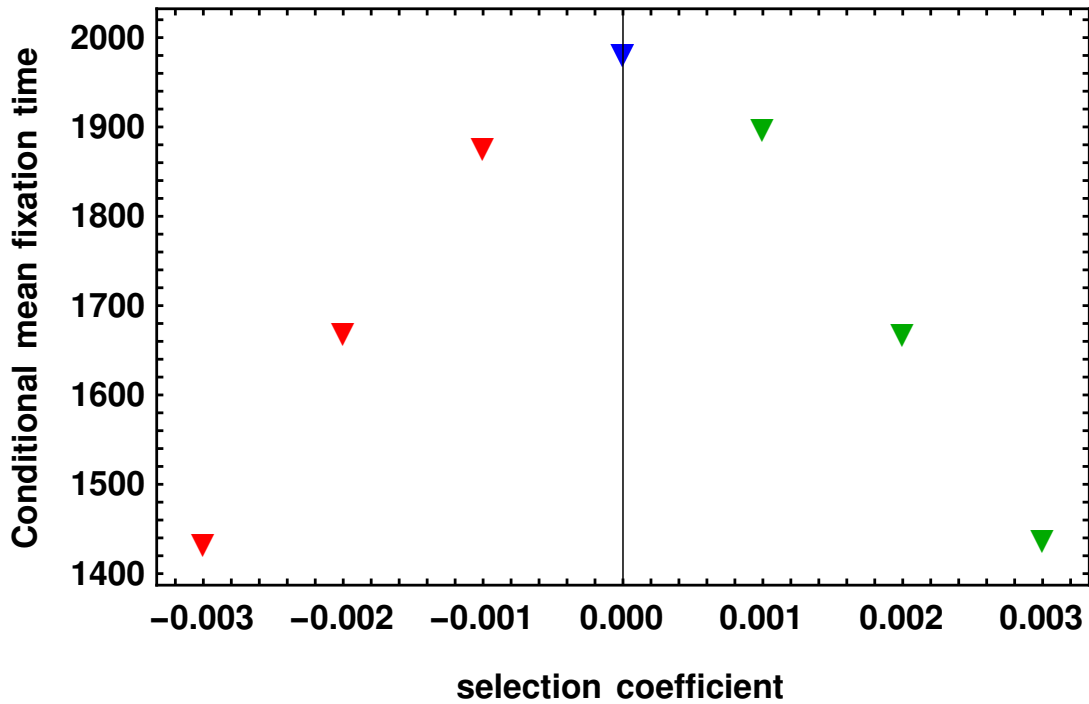


Fig. 1.3 Maruyama-Kimura symmetry: The conditional mean fixation times for a mutant with selection coefficient $-s$ (red) and s (green) are identical.

genetic diversity reaches an equilibrium and the pairwise neutral diversity at a locus is given by $\pi \approx 4N\mu$ [2]. But empirical studies of over a thousand species show that the genetic variability is much smaller than predicted by the neutral model. This observation is known as Lewontin's paradox [3–6].

Several explanations have been proposed to explain the observed narrow range of neutral diversity. Selective sweep [7–11] and/or the background selection [12–14] at selected sites are considered important mechanisms which can potentially reduce the variation at linked neutral sites.

Selective Sweeps: When a new mutation arises in a population, either selection or genetic drift can increase its frequency. In an asexual or weakly recombining population, if the new mutation goes to the fixation, the linked neutral alleles hitchhike along with it, thereby reducing the neutral genetic diversity. Furthermore, in regions of low recombination, the reduction in neutral genetic diversity depends on the conditional

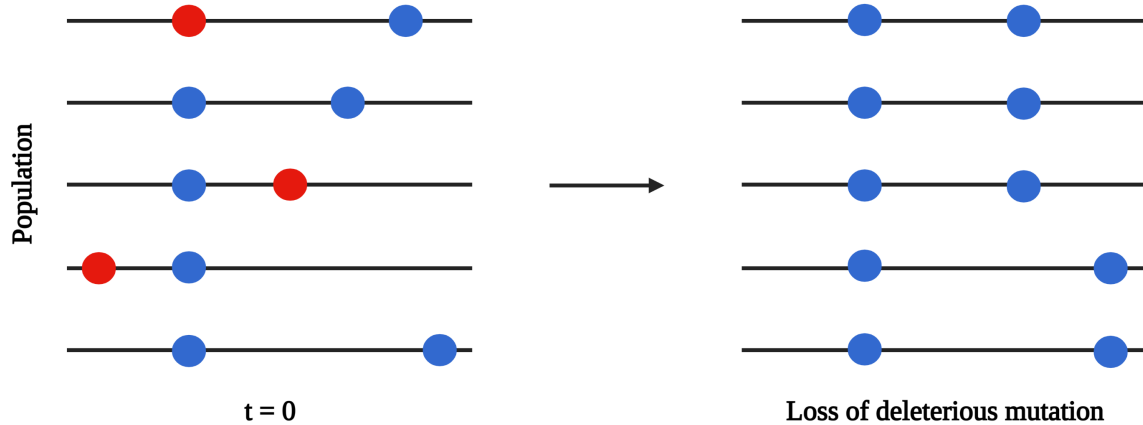


Fig. 1.4 Background selection: The genetic sequences carrying strongly deleterious mutation ● are eliminated from the population, thereby reducing the neutral (●) genetic variation in the population.

mean fixation time which is defined as the mean fixation time divided by the probability of fixation. The conditional mean fixation time for a codominant mutant is the same for a beneficial mutation with selection coefficient $+s$, and deleterious mutant with selection coefficient $-s$ (Maruyama-Kimura symmetry) [15, 16] as depicted in Fig. 1.3. As a consequence of this symmetry the variability at the linked neutral sites is the same due to beneficial and deleterious sweeps.

Background selection: The genetic diversity at the linked neutral sites can also be reduced due to selection against the recurrent deleterious mutations in the genomic regions of low recombination as depicted in Fig. 1.4. This process is known as background selection [12].

1.2 Models

In population genetics, the allele frequency dynamics are described by deterministic or stochastic models. In deterministic models, the evolutionary dynamics are described solely by mean allele frequencies and fluctuations in the allele frequencies are ignored,

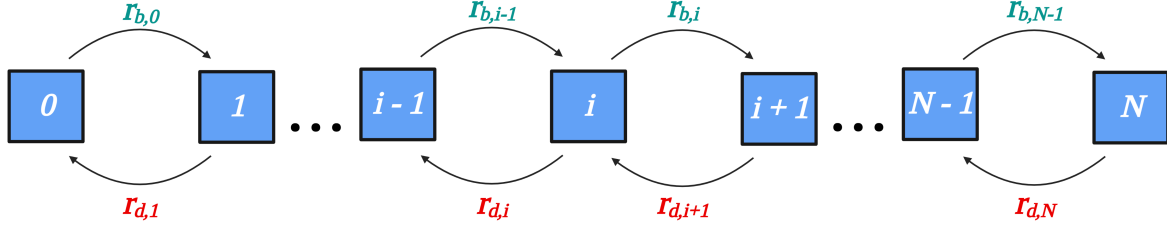


Fig. 1.5 Birth-Death Model: Birth-Death process is a continuous-time Markov process in which an individual gives birth and another one dies. The birth ($r_{b,i}$) and death rates ($r_{d,i}$) may depend on the population size and the number of individuals of type i .

whereas in the stochastic models, the fluctuations are considered which arise due to the finite size of the population. In this thesis, we have used three stochastic models that are described below.

1.2.1 Birth-Death process

A birth-death process is a class of continuous-time Markov process where the outcome at the next step depends only upon the state at the previous step. In birth-death process [17–19], the system makes a transition from state i to $i + 1$ (birth), or state i to $i - 1$ (death) according to the birth rate $r_{b,i}$, and the death rate $r_{d,i}$, respectively, as depicted in Fig. 1.5. Once the system enters the state with i individuals, it stays in this state for a time t (sojourn time), which is distributed exponentially with rate $r_{b,i} + r_{d,i}$. The probability, $P_i(t)$ that the system is in state i at time t satisfies the following differential equations:

$$\frac{\partial P_0(t)}{\partial t} = r_{d,1}P_1(t) \quad (1.1)$$

$$\frac{\partial P_i(t)}{\partial t} = r_{b,i-1}P_{i-1}(t) + r_{d,i+1}P_{i+1}(t) - (r_{b,i} + r_{d,i})P_{i,t} \quad (1.2)$$

$$\frac{\partial P_N(t)}{\partial t} = r_{b,N-1}P_{N-1}(t) - r_{d,N}P_{N,t}(t) \quad (1.3)$$

where $1 \leq i \leq N - 1$. We are interested in the special class of birth-death processes where once the system reaches the states with $i = 0$ or $i = N$, it is trapped in them forever (absorbing states). We have used this model in Chapter 2 to address a fundamental question in population genetics regarding the time taken by a newly arisen mutation to fix into the population when the environment is time-dependent.

1.2.2 Infinite sites model

In the infinite sites model, the genetic sequence is considered to be very long so that the new mutation arises at a new site along the stretch of the DNA which has not mutated before [20–22]. The second assumption is that the recombination is free, allowing allele frequencies to evolve independently at each site. In the most basic case of neutral mutations, genetic diversity is maintained in the population through the combined action of genetic drift and recurrent mutations: the genetic drift removes variation from the population, whereas mutations introduce variation at new sites. This model is used in Chapter 3 to calculate important measures of the genetic diversity such as the site frequency spectrum (described in Sec. 4.2) and heterozygosity, which is the frequency of individuals carrying different alleles.

1.2.3 Wright-Fisher model

The Wright-Fisher model is a discrete time Markov chain that describes the evolution of allele frequency under the influence of various evolutionary forces [23, 24]. In the neutral Wright-Fisher model without mutations, as depicted in Fig. 1.6, the offspring generation is obtained from the parent generation by the following steps: i) in the offspring generation, each individual chooses a parent at random from the parent generation, and ii) the previous step is repeated until the number of individuals in the offspring generation equals the number of individuals in the parent generation. If selection is present, the

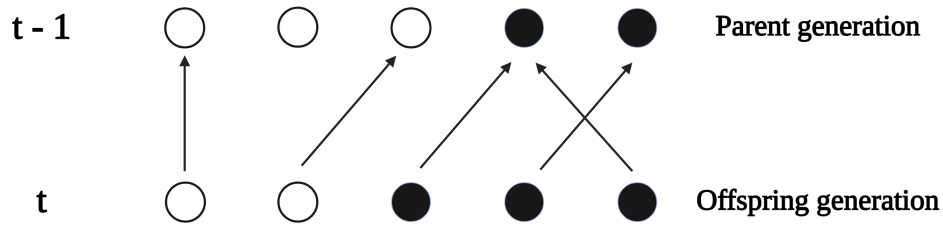


Fig. 1.6 Wright-Fisher Model: Consider a population with constant size N where the individuals are of two types, (●) and (○). Each individual in offspring generation picks a parent with probability proportional to the parent's fitness.

chance that offspring picks a particular parent is proportional to the parent's fitness. Here the time is measured in the units of lifetime of an individual. To illustrate this, consider a small population of 5 haploid individuals (see Fig. 1.6) which are of two types, represented by filled circles and open circles. For simplicity, we consider a neutral population where the fitness of both types of individuals is the same, and there are no further mutations. The offspring generation at time t is generated from the given parent generation at time $t - 1$ according to the steps described above. Each offspring is connected to its parent through an arrow, which indicates the parent from which the offspring inherited its genetic information. The probability that in generation t , $m \leq 5$ open circle type individuals are present is given by

$$P(m) = \binom{5}{m} \left(\frac{3}{5}\right)^m \left(\frac{2}{5}\right)^{(5-m)} \quad (1.4)$$

The $m = 0$ and $m = 5$ are the absorbing states; if the population becomes of one type of individual, then it will remain so for the subsequent generations. If there are more than two type of individuals, the above binomial sampling is replaced by the multinomial sampling. The Wright-Fisher model is used in Chapter 4 to describe the change in the neutral diversity due to background selection when selection coefficient is time-dependent.

1.3 Mathematical frameworks

In this thesis, besides performing numerical simulations of the models described in the last section, we also obtain simple, analytical results using a diffusion theory, and a semi-deterministic theory.

1.3.1 Diffusion theory

As mentioned in Sec. 1.1, evolutionary processes are not deterministic, in general. If the change in the allele frequency is regarded as a continuous-time stochastic process, then the probability density of allele frequency x at time t , given that it starts with x_0 at time t_0 obeys the following equation [25]:

$$\frac{\partial \phi(x, t|x_0, t_0)}{\partial t} = -\frac{\partial}{\partial x} (M(x)\phi(x, t|x_0, t_0)) + \frac{\partial^2}{\partial x^2} (V(x)\phi(x, t|x_0, t_0)) \quad (1.5)$$

where $M(x)$ is the average change in the allele frequency per generation due to selection and mutation and $V(x)$ is the variance in the allele frequency change per generation due to random genetic drift. The above equation is the Kolmogorov forward equation, and in the physics literature, it is known as the forward Fokker-Planck equation and used to study dynamics or non-trivial stationary states. When the population eventually reaches an absorbing state, it is comparatively easier to calculate the quantities of interest using the following Kolmogorov backward equation:

$$-\frac{\partial \phi(x, t|x_0, t_0)}{\partial t_0} = M(x_0, t_0) \frac{\partial}{\partial x_0} \phi(x, t|x_0, t_0) + V(x_0, t_0) \frac{\partial^2}{\partial x_0^2} \phi(x, t|x_0, t_0) \quad (1.6)$$

The above equations (1.5), and (1.6) assume that the mutation rate (μ) and the selection coefficient (s) are small, and that the population size N is large, but the scaled selection coefficient (Ns), and scaled mutation rate ($N\mu$) are finite.

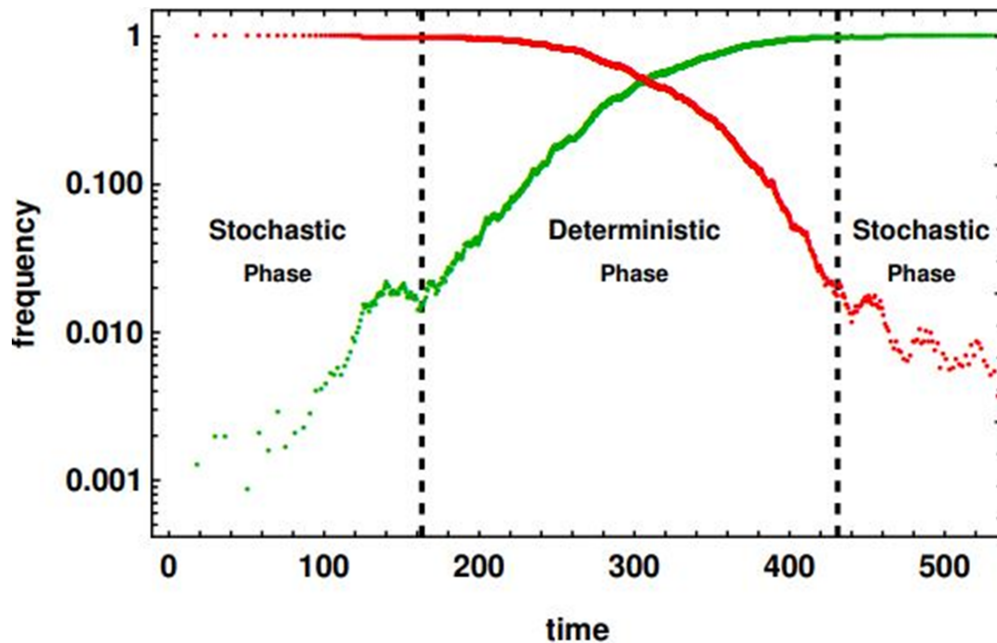


Fig. 1.7 Dynamical phases: The wild type frequency (red) and beneficial mutant allele frequency (green) initially evolve stochastically (phase A), followed by deterministic evolution (phase B), and again evolve stochastically when they are close to absorption (phase C).

1.3.2 Semi-deterministic theory

Unfortunately, it is not possible to solve the diffusion equations (1.5) and (1.6) in a closed form in the presence of selection. However, for beneficial mutations, we can use a semi-deterministic theory [26, 27]. Given that the mutant allele fixes in the population, the dynamics of its frequency can be divided into three phases as shown in Fig. 1.7: i) initially, the mutant allele frequency is very low so that it evolves stochastically (phase A), ii) once the mutant allele frequency becomes substantial, it evolves deterministically (phase B), and iii) when the mutant allele is close to fixation, so that the wild type frequency is very low, it again evolves stochastically once again (phase C).

In phase A, as the mutant allele frequency is low, the forward Fokker Planck equation can be simplified to the Feller differential equation [28] in the limit $x \rightarrow 0$. The Feller

equation can be solved exactly for the allele frequency distribution, and conditioned on fixation, the allele frequency distribution reaches a stationary state at long times. In these conditional trajectories, the further evolution of the allele frequency is treated deterministically (phase B). In phase C, the wild type frequency is low and a Feller equation similar to that in phase A can be written for the wild type. The average allele frequencies in the stochastic phases and the deterministic phase are matched at the boundaries. These ideas are used in Chapter 2 to find the full fixation time distribution of the mutant allele in time-dependent environment.

1.4 Quantities of Interest

In this thesis, we will primarily focus on fixation time and site frequency spectrum that are described below in detail.

1.4.1 Conditional mean fixation time

We discussed in Sec. 1.3 that the reduction in the genetic diversity at the linked neutral sites depends on the conditional mean fixation time which is the mean fixation time calculated only from those allele trajectories that are destined to fix. The unconditional mean fixation time for the mutant which starts with frequency x_0 can be calculated using (1.5) as $\bar{T} = \int_0^\infty dt t \phi(x \rightarrow 1, t | x_0, 0)$. The conditional mean fixation time is given by $\bar{T}_c = \frac{\bar{T}}{u(x_0)}$ where $u(x_0)$ is the eventual fixation probability and given by setting $u(x) = \phi(x \rightarrow 1, t \rightarrow \infty | x_0, t_0)$ in (1.5). In Chapter 2, we calculate the conditional mean fixation time for the more complicated scenario where the selection coefficient is time-dependent.

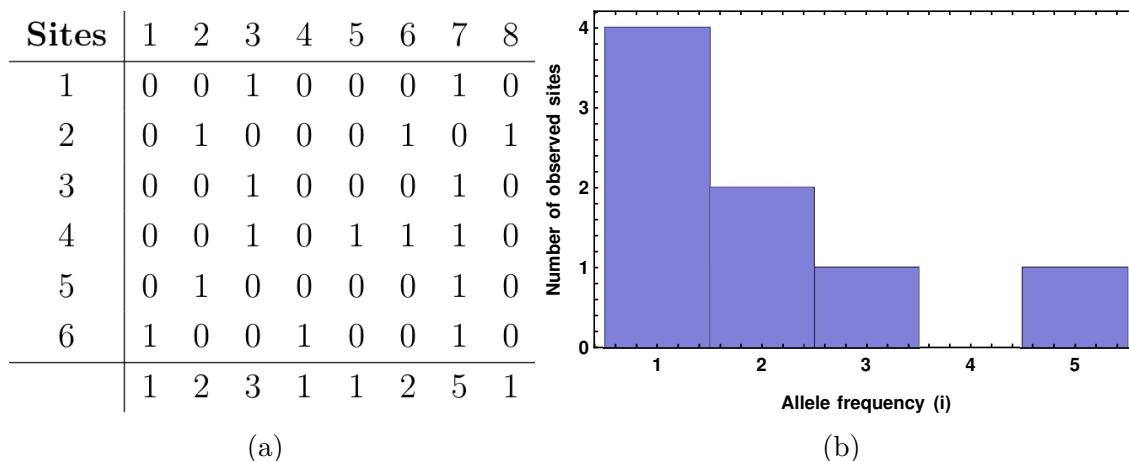


Fig. 1.8 Site frequency spectrum: Consider the genomic data set of 6 individuals and 8 sites where 0 and 1, respectively, denote the wild type and mutant allele. The number of observed polymorphic sites (S_i) with allele frequency i are then represented by a histogram which is generated by counting the number of mutant alleles present at each site.

1.4.2 Site frequency spectrum

In order to make computational or analytical progress, we must reduce the big genomic dataset using some summary statistics that can encapsulate the variation within the population. One such powerful statistics is site frequency spectrum (S_i), which is defined as the number of sites at which exactly i individuals carry the mutation.

Figure 1.8a illustrates a simple example to measure the site frequency spectrum from the genomic data set. Consider a population of 6 individuals where each individual DNA sequence has 8 sites. The mutant alleles are represented by 1, and wild types are represented by 0. At each site, we count the total number of mutant alleles present among all individuals and represent it by vector (1, 2, 3, 1, 1, 2, 5, 1). The histogram of this vector yields the site frequency spectrum as depicted in Fig. 1.8b.

The shape of the site frequency spectrum provides information about the evolutionary forces under which a population may be evolving. In static environment, the site frequency spectrum approaches a stationary distribution under the action of mutation, selection and genetic drift [29], and is shown for neutral, deleterious and beneficial mutations

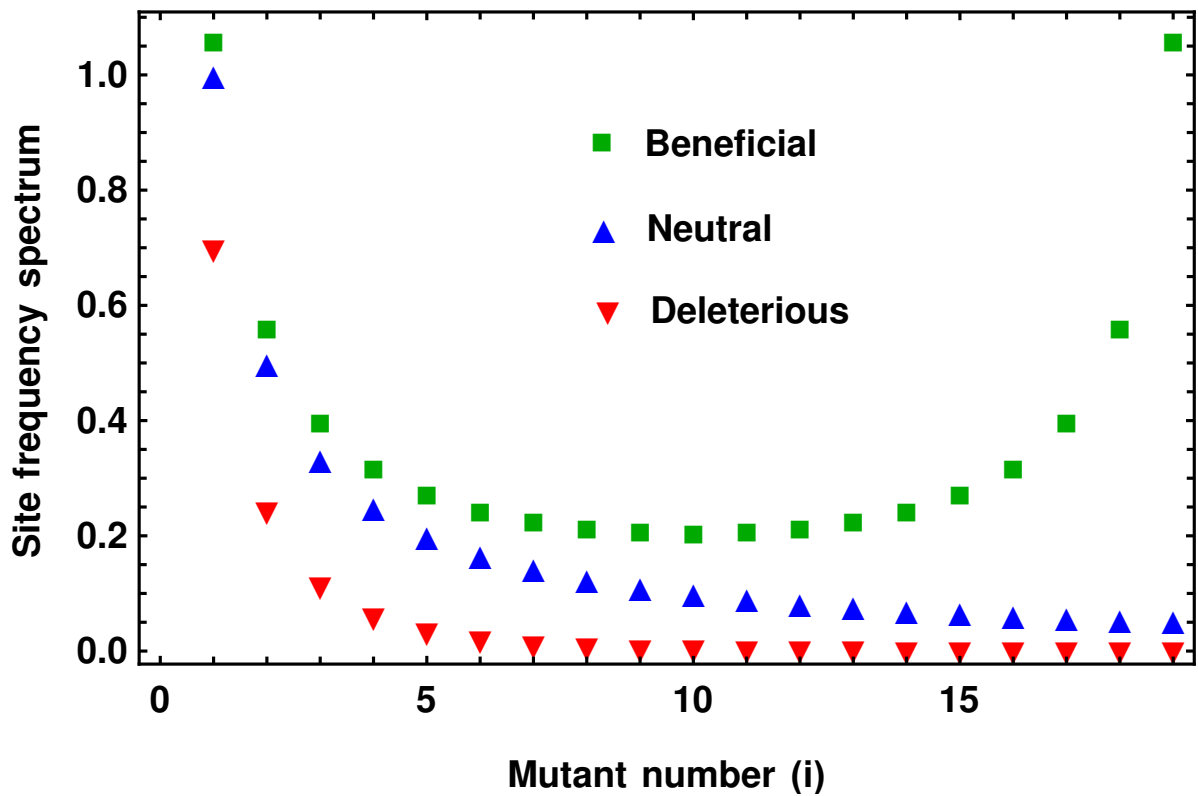


Fig. 1.9 Site frequency spectrum: Under the infinite sites model, the equilibrium site frequency spectrum decreases monotonically for neutral and deleterious mutations. For beneficial mutations, the site frequency spectrum is U-shaped because the beneficial mutation can rise to high frequency in the population due to their advantageous effect.

in Fig. 1.9. When the mutations are neutral, the equilibrium site frequency spectrum decreases monotonically as $\frac{\theta}{x}$, where $\theta = 2N\mu$ is the scaled mutation rate. But it decreases exponentially fast for deleterious mutations, and it is non-monotonic in the case of beneficial mutations. In Chapters 3 and 4, we find the site frequency spectrum when the population size and selection coefficient vary with time.

1.5 Overview of the thesis

In this thesis, we focus on how various measures of genetic diversity are shaped in changing environments under the influence of various evolutionary forces described in

section 1.1. We consider the change in environment i) due to time-varying selection coefficient, and ii) time-varying population size.

In chapter 2, we study the reduction in the linked neutral variation due to the fixation of mutation at the selected site in the changing environment when population size is constant but selection coefficient varies periodically with time using a continuous-time birth-death model. We find that the Maruyama-Kimura symmetry is not preserved in the changing environment when selection coefficient varies with time resulting in different neutral genetic diversity levels due to the beneficial and the deleterious sweep. Even in the slowly changing environment, the conditional mean fixation time is found to be significantly different than that in the static environment for recessive deleterious mutations while the fixation time for beneficial mutation is hardly affected.

In chapter 3, we consider the joint effect of changing population size and changing selection coefficient on the measures of genetic diversity using infinite sites model. Our most interesting result is for on-average neutral mutations in slowly changing environment, where the time-averaged site frequency spectrum is found to have a qualitative different shape as compared to equilibrium site frequency spectrum in the static environment, and mimics the site frequency spectrum for the beneficial mutations in the static environment. As in chapter 2, the deleterious mutations are found to be the reason for this qualitatively different behavior of the site frequency spectrum.

In chapter 4, we further explore the effect of changing selection on selective sweeps in inbreeding population, and the changing environment effects are strongest for the randomly mating population. We also study the effect of background selection on the linked neutral variation in the changing environment using discrete time Wright-Fisher model, and find our results to be significantly different than that in the static environment.

In chapter 5, we conclude the thesis with the main results and discuss some interesting open questions.

Glossary

Term	Description
Allele	variant forms of a gene.
Allele frequency	fraction of an allele.
Background selection	process of elimination of deleterious mutations due to selection bias that reduces genetic diversity
Dominance	one allele has a stronger effect on an individual's traits than another allele for the same gene.
Fitness	refers to the relative ability of an individual to survive and reproduce
Fixation	when the mutant allele frequency becomes 1
Heterozygosity	frequency of different alleles at a particular locus.
Locus	position on a chromosome where a particular gene is located.
Linkage	tendency of closely positioned genes to be inherited together
Selective sweeps	process by which a mutant sweeps through the population reducing the genetic variation at the linked neutral sites
Site frequency spectrum (SFS)	distribution of allele frequencies across the polymorphic sites.

Bibliography

- [1] COOK, L. M. and I. J. SACCHERI, 2013 The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* **110**: 207–212.
- [2] KIMURA, M., 1984 The neutral theory of molecular evolution. Cambridge University Press.
- [3] LEWONTIN, R. C., 1974 The genetic basis of evolutionary change. New York (NY): Columbia University Press.
- [4] ROBERTS, R. G., 2015 Lewontin’s paradox resolved? In larger populations, stronger selection erases more diversity. *PLoS Biol.* **13**: e1002113.
- [5] BUFFALO, V., 2021 Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin’s paradox. *eLife* **10**: e67509.
- [6] CHARLESWORTH, B. and J. D. JENSEN, 2022 How can we resolve Lewontin’s paradox? *Gen. Biol. and Evol.* **14**: evac096.
- [7] MAYNARD SMITH, J. and J. HAIGH, 1974 Hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.

-
- [8] BEGUN, D. J. and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**: 519–520.
- [9] STEPHAN, W., T. H. E. WIEHE, and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Pop. Bio.* **41**: 237–254.
- [10] STEPHAN, W., 1995 An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. *Mol. Bio. and Evol.* **12**: 959–962.
- [11] GILLESPIE, J. H., 2000 Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**: 909–919.
- [12] CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- [13] KAPLAN, N. L., R. R. HUDSON, and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. *Genetics* **123**: 887–889.
- [14] NORDBOG, M., B. CHARLESWORTH, and D. CHARLESWORTH, 1996 The effect of recombination on background selection. *Genet. Res.* **67**: 159–174.
- [15] MARUYAMA, T., 1974 The age of an allele in a finite population. *Genet. Res.* **23**: 137–143.
- [16] MARUYAMA, T. and M. KIMURA, 1974 A note on the speed of gene frequency changes in reverse directions in a finite population. *Evolution* **28**: 161–163.
- [17] YULE, G. U., 1924 A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. *Phil. Trans. R. Soc. Lond. B* **213**: 21–87.

-
- [18] KENDALL, D. G., 1948 On the generalized birth-and-death process. *Ann. Math. Stat.* **19**: 1–15.
- [19] KARLIN, S., and H. M. TAYLOR, 1975 A first course in stochastic processes. California: Academic Press.
- [20] KIMURA, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**: 893–903.
- [21] TAJIMA, F., 1996 Infinite-allele model and infinite-site model in population genetics. *Journal of Genetics* **75**: 27–31.
- [22] WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Pop. Bio.* **7**: 256–276.
- [23] FISHER, R. A., 1930 The genetical theory of natural selection. Oxford: Clarendon Press.
- [24] WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- [25] RISKEN, H., 1966 The Fokker Planck equation methods of solution and applications. Berlin: Springer.
- [26] COHN, H. and P. JAGERS, 1994 General branching processes in varying environment. *Ann. Applied Prob.* **4**: 184–193.
- [27] MARTIN, G. and A. LAMBERT, 2015 A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theo Pop Biol* **101**: 40–46.
- [28] FELLER, W., 1951b Two singular diffusion problems. *Ann. Math.* **54**: 173–182
- [29] KIMURA, M., 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**: 247–269

Chapter 2

Time to fixation in changing environments

2.1 Introduction

In this Chapter, we investigate the reduction in genetic diversity at a linked neutral locus caused by the fixation of beneficial or deleterious mutants (selective sweep) in the periodically changing selective environment. Here, we focus on conditional fixation time because the reduction in linked genetic diversity in low recombination regimes is proportional to it.

In a finite, recombining population where a selected locus is linked to neutral loci, if a new advantageous mutation fixes faster than the time it takes for neutral loci to get dissociated via recombination, the neutral genetic diversity in the neighborhood of the selected locus is reduced (beneficial sweep) [1, 2]; a similar pattern arises when a mildly deleterious mutation reaches fixation due to genetic drift (deleterious sweep). Thus, the time of fixation is intimately related to the level and patterns of neutral diversity [3]. It is important to note that the fixation time under discussion here is obtained from a stochastic process that is conditioned on fixation; this is because at the end of selective

sweep, one is observing only those trajectories of the new allele in which fixation has occurred and not all the trajectories in which it had appeared [4, 5].

Theoretical models of sweeps and their genomic applications assume the selective environment to be constant in time; however, environmental variation is ubiquitous in nature, and may potentially affect the fixation time. For example, suppose a mutant arises while selection is positive and increasing. In this case, the mean fixation time, conditional on fixation, is expected to be smaller than when the selection pressure remains the same as that when the mutant arose, and can result in a larger reduction in the neutral diversity. One may then ask: how much does the fixation time in a changing environment (especially, if it varies slowly) differ from that in a static environment?

Furthermore, in static environments, the conditional mean fixation time has the important property of being the same for a mutant with selection coefficient s and dominance coefficient h and a mutant with respective parameters, $-s$ and $1 - h$ [6, 7], as a result of which it may be difficult to distinguish between the diversity patterns due to positive and negative selection [8]. In a changing environment, on general grounds, this symmetry can be expected to be absent, and one may delineate the parameter space where the lack of this symmetry has a strong effect on variability patterns.

As a first step towards an understanding of selective sweeps in changing environments, here we study the properties of the conditional fixation time of a mutant in a finite, diploid population when the selection coefficient is time-dependent. To the best of our knowledge, except for a preliminary study [9], the fixation time in a changing environment has not been investigated in detail. We consider the evolution in an environment that changes periodically due to, for example, seasonal cycles [10], and study how the fixation time is affected by the rate of environmental change, the time of appearance of the mutant, the strength of selection and the dominance coefficient. Throughout the Chapter, we assume random mating and autosomal inheritance.

Our results are obtained analytically using a diffusion theory for time-inhomogeneous processes when selection of either sign is weak or moderate and a semi-deterministic theory for strongly selected beneficial mutants, and are supplemented and checked by numerical simulations. Our main finding is that in slowly changing environments, the conditional mean fixation time of an initially beneficial mutant with intermediate dominance is well-approximated by that in a static environment, and the same holds true for an initially deleterious mutant under moderate selection. However, if an initially deleterious mutant is recessive, its conditional mean fixation time is considerably longer or shorter in a slowly changing environment than in a static environment. In other words, the symmetry property for the conditional mean fixation time mentioned above [6, 7] does not hold between recessive deleterious and dominant beneficial mutants. Since by virtue of Haldane’s sieve, which operates in both static [11] and slowly varying environments [12], most deleterious mutations are recessive and beneficial ones are dominant, the results obtained here are relevant to an understanding of selective sweeps in changing environments (see DISCUSSION for details).

2.2 Model

We consider the model in DEVI and JAIN [12] that deals with a randomly mating population of size N . We assume that a single biallelic locus is under selection and the three genotypes, aa , Aa and AA have the fitness $1 + s(t)$, $1 + hs(t)$ and 1 , respectively. Here $0 < h < 1$ is the dominance parameter and $s(t) = \bar{s} + \sigma \sin(\omega t + \theta)$, $t \geq 0$ is the time-dependent selection coefficient that varies periodically with cycling frequency ω . Without loss of generality, we assume that the oscillation amplitude $\sigma > 0$ but the time-averaged selection coefficient \bar{s} is arbitrary, and the initial phase $0 \leq \theta < 2\pi$.

For large population size and small selection coefficient, instead of genotypic frequencies, we can work with the allelic frequencies [13]. We start with a single mutant allele in

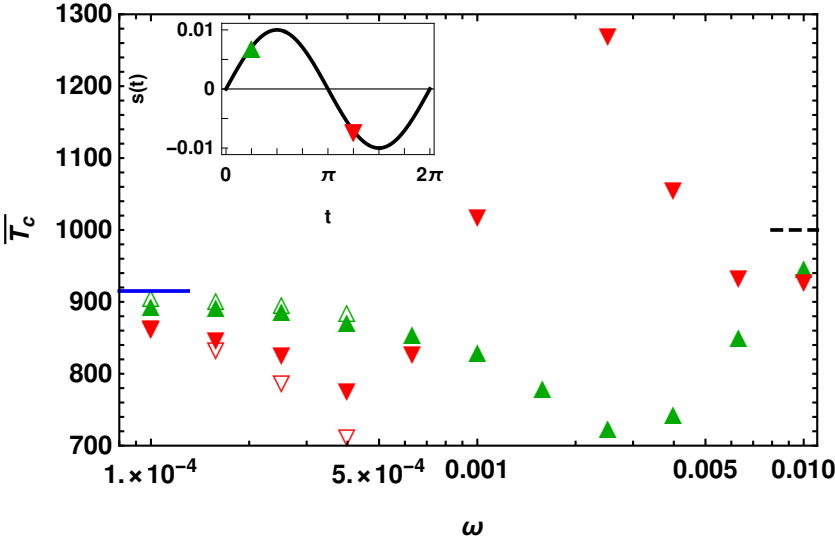
the population and ignore any further mutations. The evolution of the population under selection and random genetic drift is modeled by a continuous time birth-death process (Chapter 4, KARLIN and TAYLOR [14]) in which the number i of alleles a increase or decrease by one at rate $r_b(t)$ or $r_d(t)$, respectively. These rates are given by

$$r_b(t) = \frac{2Niw_a(t)}{iw_a(t) + (2N - i)w_A(t)} \times \frac{2N - i}{2N} \quad (2.1)$$

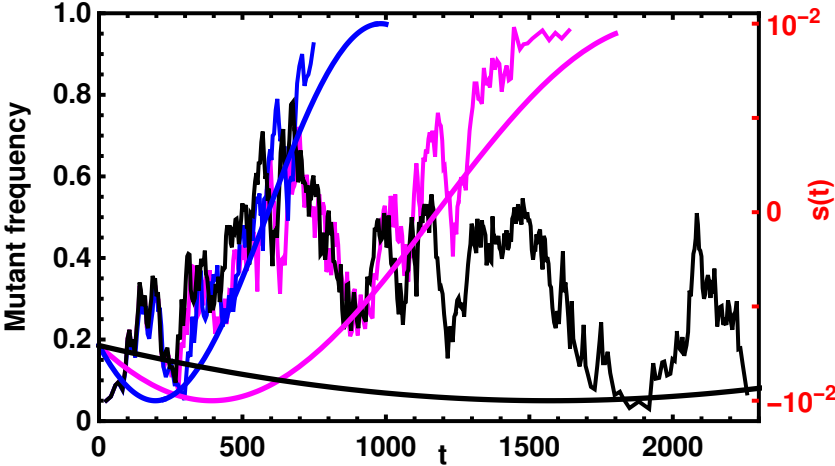
$$r_d(t) = \frac{2N(2N - i)w_A(t)}{iw_a(t) + (2N - i)w_A(t)} \times \frac{i}{2N} \quad (2.2)$$

where $w_a(t) = (1 + s(t))i + (1 + hs(t))(2N - i)$ and $w_A(t) = (2N - i) + (1 + hs(t))i$ are, respectively, the marginal fitness of allele a and A . The allele numbers at time t are updated at time $t + \delta t$ where the interval δt is chosen from the probability distribution $p(\delta t) = r(t + \delta t)e^{-\int_t^{t+\delta t} dt' r(t')}$ with $r(t) = r_b(t) + r_d(t)$ being the total rate at which either birth or death events occur.

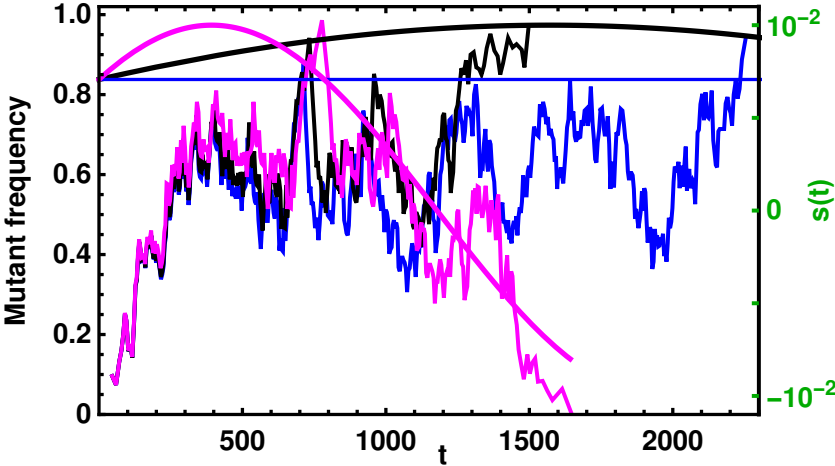
For computational efficiency, numerical simulations of the above model were carried out assuming that the birth and death rates remain constant at $r(t)$ during the interval δt . Then it follows that δt obeys an exponential distribution with rate $r(t)$; however, we have checked that our results do not change if we relax this assumption. In our simulations, $10^6 - 10^8$ independent trajectories of the mutant allele were generated but the data for the fixation time were averaged over only those trajectories that lead to the fixation of the mutant. The conditional mean fixation time was thus obtained by averaging over about 10^3 and 10^4 fixation events for deleterious and beneficial mutants, respectively. The standard error on the conditional mean fixation time were also calculated for some representative parameters and found to be at most 2% of the mean value.



(a)



(b)

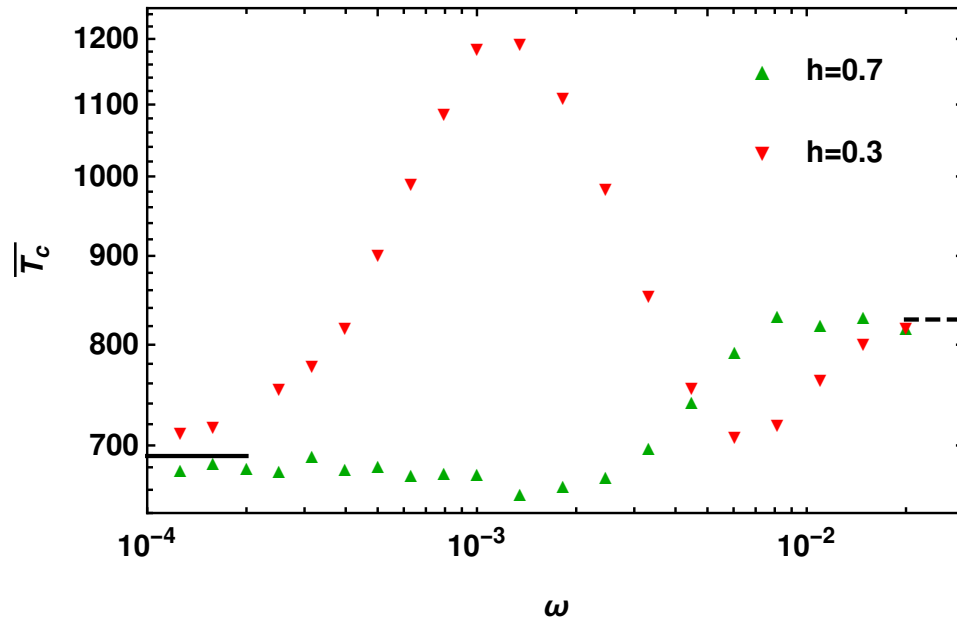


(c)

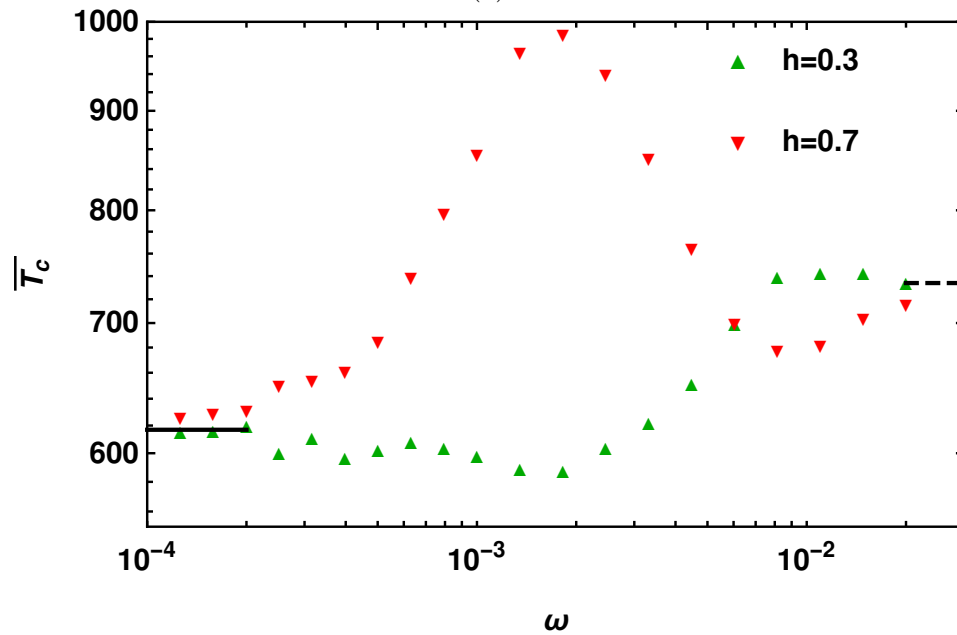
Fig. 2.1 Top panel: Conditional mean fixation time \bar{T}_c of a dominant mutant ($h = 0.7$, \blacktriangle) and a recessive mutant ($h = 0.3$, \blacktriangledown) with selection coefficient $s(t) = \sigma \sin(\omega t + \pi/4)$ and $-s(t)$, respectively (see inset for $s(0)$) to show that the symmetry between the conditional mean fixation time for dominant beneficial and recessive deleterious mutant in static environments is not preserved in changing environments. The data are obtained by numerical simulations (closed symbols) and numerically integrating the diffusion theory equations (2.4) and (2.5) (open symbols) for small cycling frequencies in a population of size N . The conditional mean fixation time in the static environment with selection $|s(0)|$ (solid line) and in the neutral environment, given by $2N$, (dashed line) are obtained from diffusion theory and shown for comparison. Middle and Bottom panel: Mutant allele trajectories for (b) the initially deleterious mutant ($h = 0.3$) and cycling frequencies, $\omega = 5 \times 10^{-4}$ (black), 2×10^{-3} (magenta), and 4×10^{-3} (blue), and (c) the initially beneficial mutant ($h = 0.7$) for $\omega = 0$ (blue), 5×10^{-4} (black), and 2×10^{-3} (magenta). The allele trajectories in each panel are almost same at short times as they were started with the same random seed in the computer simulation. The smooth curves show the selection coefficient for the corresponding cycling frequency. In all the panels, $\sigma = 0.01$ and $N = 500$ so that $\alpha = N|s(0)| \approx 3.5$.

2.3 Conditional fixation time in changing environments

In a constant environment, the expected fixation time of a new mutant that is destined to fix decreases with the magnitude of its selection coefficient s since a strongly deleterious mutant fixes soon to avoid extinction, whereas a strongly beneficial mutant that has a low chance of extinction grows fast [15–17]. In a population of size N , this result holds only for a strongly selected mutant ($N|s| \gg 1$) but for weak selection ($N|s| \ll 1$), the conditional mean fixation time can vary nonmonotonically with s [18]. Interestingly, for any selection strength, diffusion theory predicts that the conditional mean fixation time of a single mutant has a remarkable property: it is the same for a beneficial mutant with selective coefficient s and dominance coefficient h and a deleterious mutant with respective parameters $-s$ and $1 - h$ [6, 7]. For a codominant mutant ($h = 1/2$), this symmetry is even stronger in that it holds for any initial mutant number $0 < i < 2N$ (p. 170, EWENS [19]).



(a)



(b)

Fig. 2.2 Conditional mean fixation time of an initially deleterious (\blacktriangledown) and initially beneficial (\blacktriangle) mutant with selection coefficient $s(t) = \overline{s} + \sigma \sin(\omega t + 3\pi/4)$ and $-s(t)$, respectively, obtained by numerical simulations in a population of size N . The conditional fixation time in static environment with selection $|s(0)|$ (solid) and in the time-averaged environment with selection $|\overline{s}|$ (dashed) are also shown for comparison, and obtained using diffusion theory. In both cases, $\overline{s} = 0.01$, $\sigma = 0.007$ and $N = 500$. The qualitative behavior of these curves can be understood using the arguments given in the text.

In order to test whether the Maruyama-Kimura symmetry mentioned above also holds in a periodically changing environment, we need to compare the fixation time of mutants whose selection coefficients are of opposite sign at *all* times. Figure 2.1a shows the conditional mean fixation time \bar{T}_c of mutants with dominance parameter h and $1 - h$ and selection coefficient $s(t)$ and $-s(t)$, respectively, when they are under moderate selection and on-average neutral (for nonzero average selection coefficient, see Fig. 2.2). It is clear from these figures that the Maruyama-Kimura symmetry does not hold when selection is time-dependent.

To understand the qualitative behavior of the conditional mean fixation time \bar{T}_c in Fig. 2.1a, we first consider the fixation time of the initially deleterious mutant in a slowly deteriorating environment. If this mutant segregates in the population for too long, it is at a risk of extinction even if it manages to reach a high allele frequency (see mutant allele trajectory for cycling frequency $\omega = 5 \times 10^{-4}$ in Fig. 2.1b). For this reason, in Fig. 2.1a, below $\omega \approx \pi/(4 \times 900) = 8 \times 10^{-4}$, the fixation time is smaller than the corresponding result in the static environment. The fixation time of the initially beneficial mutant in a slowly improving environment is also smaller than that in the constant environment but for a different reason: here, as exemplified by the allele trajectories for $\omega = 0$ (static environment) and 5×10^{-4} in Fig. 2.1c, the mutant in the latter case, by virtue of its larger selection coefficient, grows faster than the one in the static environment; therefore, for frequencies below $\omega \approx \pi/(2 \times 900) = 2 \times 10^{-3}$, the fixation time decreases.

However, if the environment changes fast enough ($\omega > 8 \times 10^{-4}$ in Fig. 2.1a) so that an initially deleterious mutant experiences a relatively better environment ($|ds/dt| > 0$) on reaching a high frequency, fixation can occur at a later time than in a static environment (see allele trajectory for $\omega = 2 \times 10^{-3}$ in Fig. 2.1b). For a further increase in cycling frequency ($\omega > 2 \times 10^{-3}$ in Fig. 2.1a), the mutant can experience strong positive selection at late times (refer allele trajectory for $\omega = 4 \times 10^{-3}$ in Fig. 2.1b), which, as explained

above, results in a faster growth and a decrease in the fixation time. For the initially beneficial mutant, as the cycling frequency increases ($\omega > 8 \times 10^{-4}$), the mutant allele experiences decreasing selection and must fix soon to avoid extinction (see allele trajectory for ω_r in Fig. 2.1c). As the cycling frequency is further increased ($\omega > 2 \times 10^{-3}$), the mutant population sees negative but improving selection at late time, and hence runs a lower risk of extinction which results in an increase of the conditional mean fixation time. At even higher frequencies, as selection changes sign and direction several times, the fixation time during the fixation process, the fixation time for both mutants approaches the value in the time-averaged environment.

We also note that in Fig. 2.1a, an extremum in the fixation time occurs at the *resonance frequency*, $\omega_r = 2 \times 10^{-3}$ which is defined as the cycling frequency at which the environment changes at a rate proportional to a frequency scale in the population when the oscillation amplitude $\sigma = 0$. In Fig. 2.1a, the resonance frequency ω_r is proportional to the reciprocal of the fixation time $2N$ in the neutral environment [15]. Whether this extremum in the fixation time is a maximum or a minimum is determined by the initial phase θ (for a discussion of the resonance frequency for fixation probability in changing environments, refer to DEVI and JAIN [12]).

2.4 Diffusion theory in slowly changing environments

To explore and better understand the qualitative observations discussed above, we now develop a diffusion theory for time-dependent selection coefficients. As explained in Appendix A.1 the probability distribution $\Phi_b(x, t|p, t_0)$ that the mutant frequency is x at time t , given that it was p at time $t_0 < t$ obeys the following backward Kolmogorov equation [20]:

$$-\frac{\partial \Phi_b(x, t|p, t_0)}{\partial t_0} = s(t_0)g(p) \frac{\partial \Phi_b(x, t|p, t_0)}{\partial p} + \frac{pq}{2N} \frac{\partial^2 \Phi_b(x, t|p, t_0)}{\partial p^2} \quad (2.3)$$

where $s(t_0) = \bar{s} + \sigma \sin(\omega t_0 + \theta)$ and $g(p) = pq(p + h(1 - 2p))$. Using (2.3), it can be shown that the unconditional mean fixation time $\bar{T}(p, t_0)$ and the eventual fixation probability $u(p, t_0)$, respectively, obey (A.1.4) and (A.1.5). But, unfortunately, these equations do not appear to be solvable for the full range of parameters. For slow and fast changing environments, simple expressions for the eventual fixation probability have been obtained in DEVI and JAIN [12] using a perturbation theory. Below, using the same method, we find the mean fixation time in slowly changing environments.

In environments that change at a rate $\omega \ll N^{-1}$, $s(0)$ with arbitrary $Ns(0)$, the unconditional mean fixation time $\bar{T} \approx \bar{T}_0 + N\omega\bar{T}_1$ and the eventual fixation probability, $u \approx u_0 + N\omega u_1$, where the subscript 0 and 1, respectively, denote quantities in a static and slowly changing environment, respectively. As described in Appendix A.2, u_1 and $v_1 = \bar{T}_1/(2N)$ obey the following ordinary differential equations,

$$\frac{pq}{2}u_1'' + \alpha g(p)u_1' = -\frac{\partial u_0}{\partial \theta} \quad (2.4)$$

$$\frac{pq}{2}v_1'' + \alpha g(p)v_1' = -Nu_1 - \frac{\partial v_0}{\partial \theta} \quad (2.5)$$

and the quantities u_0 and $v_0 = \bar{T}_0/(2N)$ in the static environment obey [15]

$$\frac{pq}{2}u_0'' + \alpha g(p)u_0' = 0 \quad (2.6)$$

$$\frac{pq}{2}v_0'' + \alpha g(p)v_0' = -Nu_0 \quad (2.7)$$

In the above equations, prime denotes the derivative with respect to p and $\alpha = Ns(0)$ is the scaled selection strength. Equations (2.4)-(2.7) are subject to boundary conditions $u_0(1, t_0) = 1$ and $u_i(0, t_0) = v_i(0, t_0) = v_i(1, t_0) = 0, i = 0, 1$. The conditional mean fixation time \bar{T}_c scaled by the mean fixation time $2N$ in the neutral environment is then given by

$$\frac{\bar{T}_c}{2N} = \frac{\bar{T}}{2Nu} \approx \frac{v_0}{u_0} + N\omega \left(\frac{v_1}{u_0} - v_{0,c} \frac{u_1}{u_0} \right) \quad (2.8)$$

where $v_{0,c} = v_0/u_0$ (see Appendix A.2 for details).

Although a formal solution of (2.4)-(2.7) can be written down, it appears difficult to obtain a simple analytical expression for the fixation time using these results in (2.8). However, (2.4)-(2.7) can be easily integrated numerically, and as attested by Fig. 2.1a, these numerical results are in good agreement with those obtained from the simulations at small cycling frequencies.

2.4.1 Weak selection

In a static environment, the conditional mean fixation time of a beneficial mutant under weak selection ($N|s| \ll 1$) and with dominance coefficient $h > 1/2$ *increases* with the selection pressure and can be larger than the fixation time of a neutral mutant. This may be understood by noting that although the mutant population is subject to strong random fluctuations, the fixation probability of a beneficial mutant increases with the level of dominance [21] and therefore a beneficial dominant mutant can counter the risk of extinction at late times. By the Maruyama-Kimura symmetry, an analogous result is obtained for a deleterious mutant with dominance coefficient $h < 1/2$ [18].

To see this result quantitatively, for small $\alpha = Ns$, we expand the fixation probability u_0 and the fixation time v_0 in a power series about $\alpha = 0$ up to order α^2 , and substitute them in (2.6) and (2.7). Collecting terms with the same power of α on both sides of these equations, we get a set of second order ordinary differential equations which can be solved straightforwardly, and we finally obtain

$$v_{0,c} \approx 1 + \frac{\alpha H}{9} - \frac{\alpha^2}{72} \quad (2.9)$$

where $H = h - (1/2)$ is the deviation from codominance. The above result shows that the conditional mean fixation time (relative to the neutral fixation time) is a non-monotonic function of α with a maximum at $\alpha^* = 4H$ and the value at the maximum,

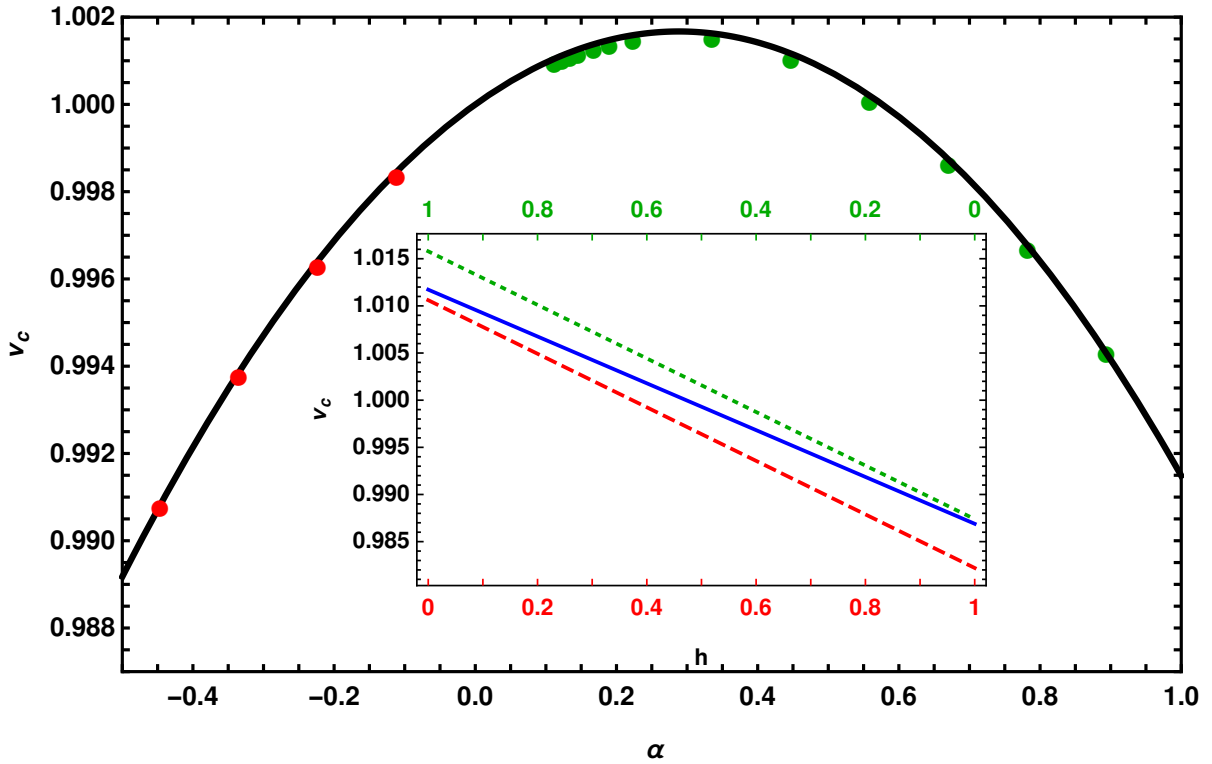


Fig. 2.3 Scaled conditional mean fixation time, $v_c = \bar{T}_c/2N$ for a co-dominant mutant under weak selection in slowly changing, on-average neutral environment in which the selection coefficient $s(t) = \sigma \sin(\omega t + \theta)$. The parameter $\alpha = N|s(0)|$ was varied with selection amplitude σ , keeping the population size N and the initial phase $\theta = \pi/4$ (for positive α) and $5\pi/4$ (for negative α) fixed. The inset shows the variation of the conditional mean fixation time with dominance in static environment (solid) and slowly varying environment for initial phase $\theta = \pi/4$ (dotted) and $5\pi/4$ (dashed), and $\sigma = 0.000158$. In both plots, $N = 2 \times 10^3$ and $N\omega = 0.08$, and the lines show the analytical expressions (2.9) and (2.10) and the points show the numerical solution of (2.4)-(2.7).

$v_{0,c}(\alpha^*) = 1 + (2H^2/9)$. As $H^2 < 1/4$, the conditional mean fixation time of the selected mutant can exceed that of the neutral mutant at most by $\sim 5\%$, as observed numerically in MAFESSONI and LACHMANN [18].

In a slowly changing environment, unlike in the last section where $\alpha > 1$, the fixation time of an initially beneficial (deleterious) mutant in an improving (deteriorating) environment increases (decreases) with selection strength when $\alpha \ll 1$; this is due to a slight increase (decrease) in the fixation probability from the neutral value (see Fig. 2.3).

But with increasing selection strength, the fixation time in either case eventually decreases. Proceeding in a similar fashion as for the static environment, we expand $u_1(p)$ and $v_1(p)$ in a power series in α to quadratic orders, and plug them in (2.4) and (2.5). We then find the change in the fixation time due to slowly changing environment to be

$$\begin{aligned} v_{1,c} &\approx \frac{\alpha}{54} \left[(70 - 6\pi^2) H + 9(\pi^2 - 9) \right] \cot \theta \\ &+ \frac{\alpha^2}{108} (3\pi^2 - 38) \cot \theta \end{aligned} \quad (2.10)$$

for $\theta \neq 0, \pi$. Using this result, the maximum in the total conditional mean fixation time is found to occur at

$$\alpha^* \approx 4H + \frac{2}{3} N\omega \cot \theta \left[(6\pi^2 - 82)H + 9\pi^2 - 81 \right] \quad (2.11)$$

Equations (2.9) and (2.10) and also Fig. 2.3 show that the conditional mean fixation time continues to be a non-monotonic function of the selection coefficient in changing environments. But, for a codominant mutant, while $v_{0,c}$ is symmetric about $\alpha = 0$ in a constant environment (see (2.9)), due to the lack of Maruyama-Kimura symmetry in the changing environment, the maximum in the fixation time occurs at a nonzero α^* , as predicted by (2.11). For small α , from (2.10), we have $v_{1,c} \approx 0.2h\alpha \cot \theta$ which shows that the changing environment has the strongest effect when the mutant of either sign is dominant; however, the magnitude of these effects is quite small, see inset of Fig. 2.3.

2.4.2 Moderate selection

We now consider the parameter regime where selection is moderately strong and the deleterious mutant has a significant chance of fixation ($1 \ll |\alpha| \lesssim 20$). As in the last subsection, one would like to obtain simple analytical expressions for the time \bar{T}_c but, unfortunately, it is generally not possible to develop consistent approximations when the

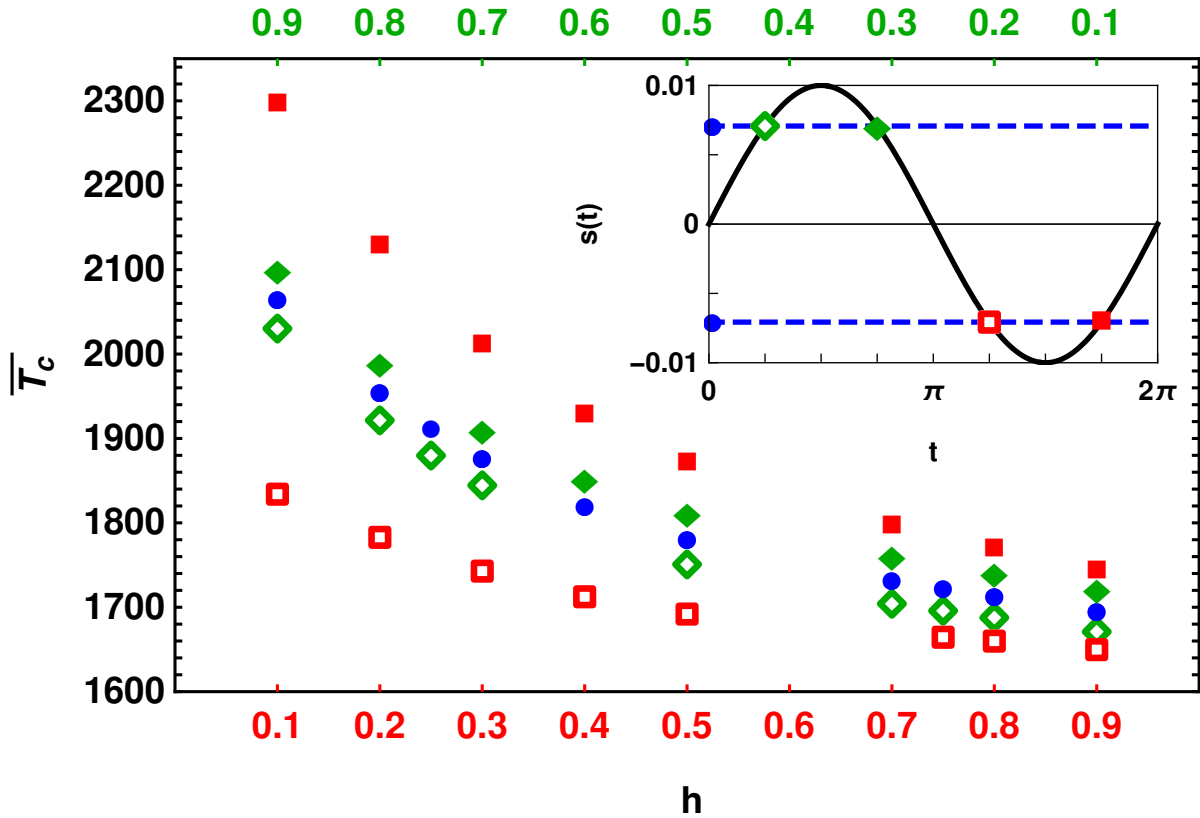


Fig. 2.4 Conditional mean fixation time \bar{T}_c for moderate selection in static (\bullet) and slowly changing environment with selection coefficient $s(t) = \sigma \sin(\omega t + \theta)$ (diamonds) and $-s(t)$ (squares) for different dominance coefficients and the initial phase $\theta = \pi/4$ (open symbols) and $3\pi/4$ (closed symbols) in a population of size N . The other parameters are $N = 2 \times 10^3$, $N\omega = 0.05$ and $\sigma = 0.01$. The inset depicts the arrival time of the mutant in all the cases. The data are obtained within the framework of diffusion theory by numerically solving (2.4)-(2.7).

parameters are of moderate size. Below, we therefore discuss the results obtained by numerically integrating (2.4)-(2.7).

To understand the results shown in Fig. 2.4, we first consider the fixation time of the initially deleterious mutant in a deteriorating environment and the initially beneficial mutant in an improving environment (both denoted by open symbols) for a given dominance coefficient. On account of larger scaled selection strength, the former has a lower fixation probability than a deleterious mutant in the constant environment, and therefore should fix sooner to avoid extinction; the initially beneficial mutant, on

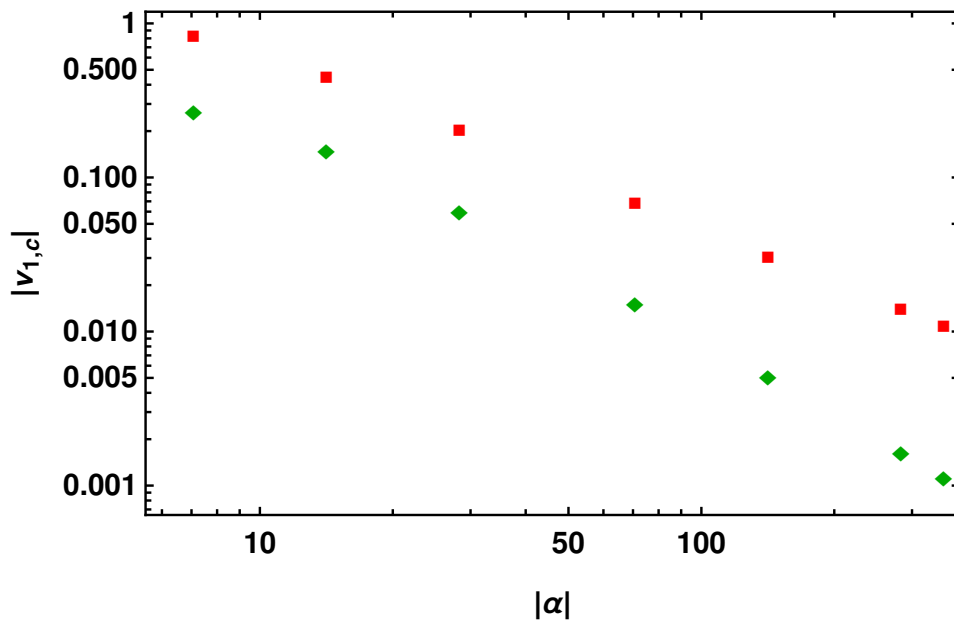


Fig. 2.5 Absolute value of the deviation between the conditional mean fixation times in slowly changing environment with selection coefficient $s(t) = \sigma \sin(\omega t + \theta)$ and static environment with selection coefficient $s(0)$ for beneficial ($\theta = \pi/4$, \blacklozenge) and deleterious mutants ($\theta = 5\pi/4$, \blacksquare) as a function of $|\alpha| = N|s(0)|$ for fixed dominance coefficient $h = 1/2$ and $\sigma = 0.01$ obtained within the framework of diffusion theory by numerically solving (2.4)-(2.7).

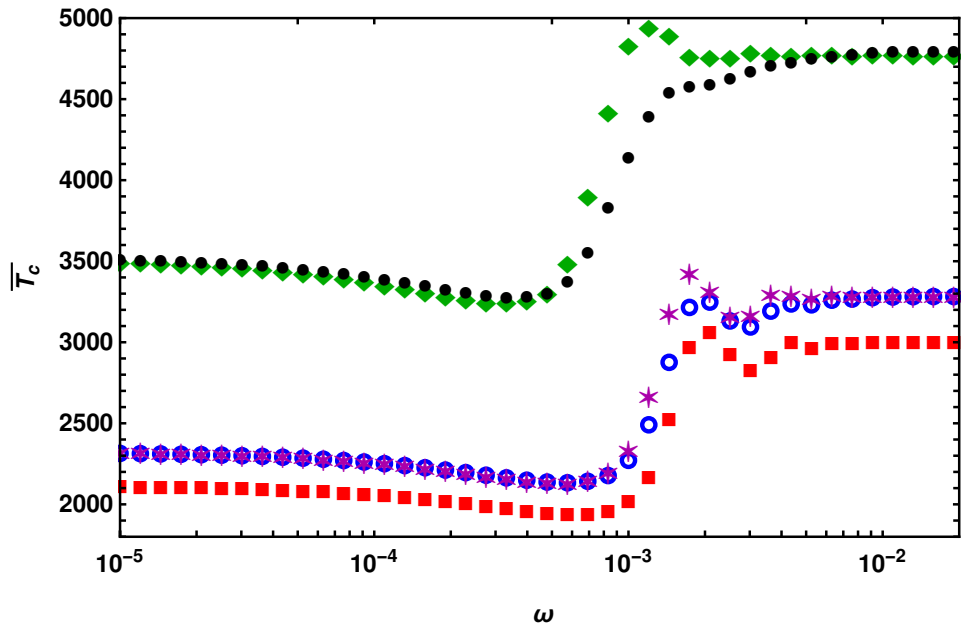
the other hand, grows faster in an improving environment than in the corresponding static environment. In either case, the fixation time is smaller than that in the static environment. For similar reasons, the fixation time of the initially beneficial (deleterious) mutant in a deteriorating (improving) environment is larger than that in the static environment. Crucially, however, as Fig. 2.4 shows, the magnitude of the deviation between the fixation time in the changing and the constant environment is larger for the initially deleterious mutant than that for the initially beneficial mutant. The reason underlying for this behavior is the strong asymmetry between the fixation probabilities of beneficial and deleterious mutants. In a static environment, for moderate-to-strong selection, a small change in the selection coefficient affects the chance of fixation of a beneficial mutant only by a small amount, but the fixation probability of a deleterious mutant changes by an exponential factor [21]. This strong asymmetry holds even in

slowly changing environments (refer Fig. 2 for moderate selection and equation (11) for strong selection of DEVI and JAIN [12]). As a result, the fixation time of an initially deleterious mutant is more strongly affected by a changing environment. Although a strongly deleterious mutation has a negligible chance of fixation, it is possible to obtain some analytical understanding of its fixation time. In Appendix A.6, we find that $v_{1,c}$ which captures the effect of changing environment on the fixation time decays slowly, as $|\alpha|^{-1}$ for a deleterious mutant, while as shown in the following section for a mutant under strong positive selection, $v_{1,c} \sim \alpha^{-2}$ (see (2.20)). These results again emphasize that the changing environment has a much more stronger impact on deleterious mutations (see also Fig. 2.5).

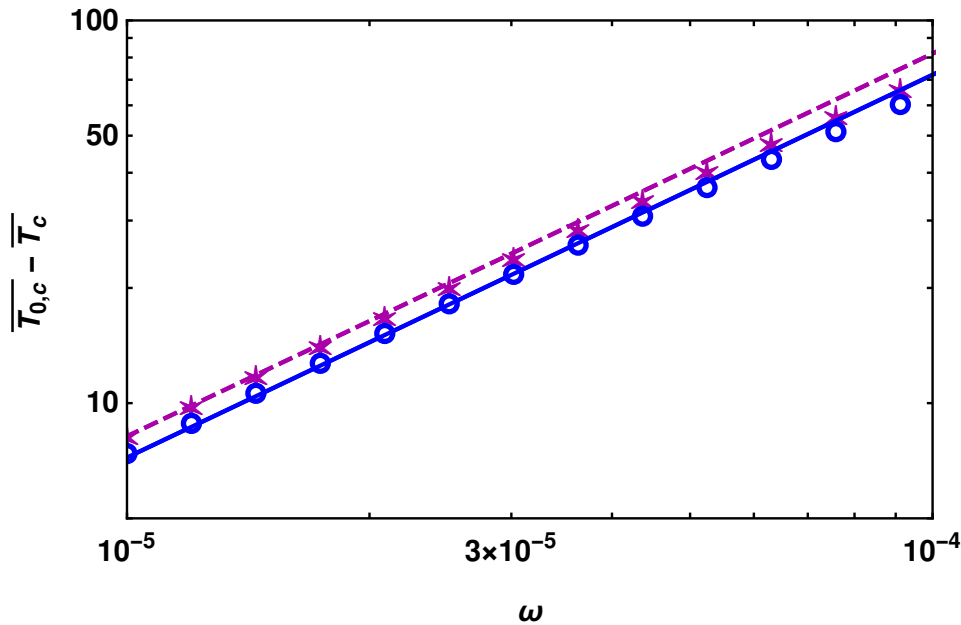
To understand the dominance-dependence of the fixation time, we first recall that in the static environment, the fixation probability increases (decreases) with the dominance level for mutants under positive (negative) selection. This result known as the Haldane's sieve [11] operates in slowly changing environments also [12], and explains the increasing (decreasing) fixation time of the initially beneficial (deleterious) mutant with increasing h . For the initially deleterious mutant, since the recessive mutant (as compared to the dominant mutant) spends more time in the varying environment, its trajectories are more influenced by the changing selection coefficient and the fixation time differs considerably from that in the static environment.

2.5 Semi-deterministic theory for on-average beneficial mutants in a large population

In the preceding discussion, we assumed the mutant to be under weak-to-moderate selection as strongly deleterious mutations are unlikely to fix [21]. Here we study the conditional mean fixation time of a mutant that is under strong positive selection



(a)



(b)

Fig. 2.6 Top panel: Conditional mean fixation time \bar{T}_c for a mutant under strong positive selection at all times with selection coefficient $s(t) = \bar{s} + \sigma \sin(\omega t + \theta)$ when the mutant is beneficial at all times to show that, except for strongly recessive or dominant mutations, it depends weakly on the dominance coefficient h . The points are obtained by numerically calculating (2.18) for dominance coefficient $h = 0.1$ (\blacklozenge), 0.3 ($*$), 0.5 (\blacksquare), 0.7 (\circ), 0.9 (\bullet) in a population of size N . The other parameters are $N = 10^5$, $\bar{s} = 0.01$, $\sigma = 0.007$, $\theta = \pi/4$. The bottom panel shows the comparison between (2.18) (points) and (2.20) (line) for the deviation in the conditional mean fixation time in a slowly changing environment where $\omega \ll s(0) \approx 0.014$.

($|\alpha| \gg 10^2$) at all times. Since the frequency of a mutant will rise faster (slower) if its selection coefficient remains larger (smaller) than $s(0)$ until it fixes, the fixation time shown in Fig. 2.6a initially decreases, then increases and finally approaches the fixation time in the time-averaged environment exhibiting oscillations with decreasing amplitude. For nonzero \bar{s} , the resonance frequency ω_r is inversely proportional to the fixation time when the selection coefficient is $|\bar{s}|$ (which, for strong positive selection, is given by (2.19) on replacing $s(0)$ by \bar{s}), and we verify that the data in Fig. 2.6a is consistent with this assertion.

Below we study the dependence of \bar{T}_c on the dominance coefficient and the rate of environmental change within a semi-deterministic theory [22]. This approach has been recently used to find the distribution of the conditional fixation time in a constant environment [23]; here, we are interested in generalizing these results to time-dependent environments.

Starting at a low initial frequency, if it escapes stochastic loss, the mutant population evolves stochastically until a time t_1 when it reaches a finite frequency (phase *A*). For such allele trajectories, it is a good approximation to treat the further evolution of the mutant population deterministically (phase *B*). However, at a time $t_2 (> t_1)$, when the mutant frequency is close to one, as the wildtypes are in low number, they are subject to stochastic fluctuations and go extinct at a time T_c (phase *C*). The stochastic phases *A* and *C* can be described by a Feller process, as discussed below.

2.5.1 Time-inhomogeneous Feller process

In a time-dependent environment, the allele frequency distribution $\Phi_f(p, t|p_0, 0)$ which describes the probability that the mutant frequency is p given that its initial frequency

is p_0 obeys the following forward Kolmogorov equation [20],

$$\frac{\partial \Phi_f(p, t|p_0, 0)}{\partial t} = -s(t) \frac{\partial}{\partial p} [g(p) \Phi_f(p, t|p_0, 0)] + \frac{\partial^2}{\partial p^2} \left[\frac{pq \Phi_f(p, t|p_0, 0)}{2N} \right] \quad (2.12)$$

where, as before, $g(p) = pq(p + h(1 - 2p))$ and $s(t) = \bar{s} + \sigma \sin(\omega t + \theta)$, $t \geq 0$. At short times where the mutant frequency is low ($p \rightarrow 0$), the frequency distribution $\Phi_f \rightarrow \mathcal{F}$, and (2.12) reduces to

$$\frac{\partial \mathcal{F}(p, t|p_0, 0)}{\partial t} = -hs(t) \frac{\partial}{\partial p} [p \mathcal{F}(p, t|p_0, 0)] + \frac{1}{2N} \frac{\partial^2}{\partial p^2} [p \mathcal{F}(p, t|p_0, 0)] \quad (2.13)$$

where $\mathcal{F}(p, t|p_0, 0)$ is the probability distribution of a Feller process [24, 25]. This process describes the mutant frequency dynamics when the lineages can be assumed to grow independently, and is a continuous analogue of classical branching process that is defined in discrete time and deals with the number of individuals. The Feller diffusion equation above can be easily generalized to include mutations and time-dependent population size [26–29]. In the later discussion, we will use the Feller process to describe the wildtype dynamics also at large times where the wildtype frequency is low. As detailed in Appendix A.3, the exact solution of (2.13) is given by (A.3.4).

2.5.2 Fixation probability

Since the probability that the mutant dies out by time t is equal to $1 - \int_0^\infty dp \mathcal{F}(p, t)$, its eventual fixation probability, $u = \text{Lim}_{t \rightarrow \infty} \int_0^\infty dp \mathcal{F}(p, t)$ (see also Appendix A.3); using (A.3.4), we then obtain

$$u(p_0, 0) = 1 - \exp \left[- \frac{2Np_0}{\int_0^\infty dt e^{-h \int_0^t dt' s(t')}} \right] \quad (2.14)$$

Equation (2.14) shows that u is nonzero provided the integral $J = \int_0^\infty dt e^{-h} \int_0^t dt' s(t')$ is finite. We verify that for constant selection and single initial mutant, (2.14) gives $u = 1 - e^{-hs} \approx hs, h > 0$ for a beneficial mutant [11].

Before proceeding further, we compare the result (2.14) with that obtained using a birth-death process in earlier studies [30, 9, 12]. While $u^{(\text{Feller})} = 1 - e^{-1/J}$, the probability $u^{(\text{birth-death})} = (1 + J)^{-1}$ (refer to (4) of DEVI and JAIN [12]) for a single mutant. For small selection coefficients, as the fixation probability is expected to be small, J must be large. Then, it follows that to leading order in $1/J$, both the processes yield the fixation probability to be $1/J$.

2.5.3 Mean fixation time in slowly changing environments

As described in Appendix A.4, the distribution of the conditional fixation time for a mutant with initial phase θ is given by

$$P(T_c; \theta) = \frac{h\Upsilon_A}{h-1} \frac{d\Upsilon_C}{dT_c} \int_0^\infty dq q^{-\frac{h}{1-h}} e^{-\Upsilon_C q} e^{-\Upsilon_A q^{-\frac{h}{1-h}}} \quad (2.15)$$

where

$$\Upsilon_C = \frac{2N}{\int_0^{T_c} dt e^{(1-h)t} \int_0^t dt' s(t')} \quad (2.16)$$

$$\Upsilon_A = 2N e^{\frac{1-2h}{1-h} \ln\left(\frac{h}{1-h}\right)} u(\theta) \quad (2.17)$$

and the eventual fixation probability u is given by (2.14) for a single mutant. Figure 2.7 shows a comparison between the expression (2.15) and the results obtained using numerical simulations when the cycling frequency is below, above and close to the resonance frequency ω_r , and we find a good agreement in all the three cases. For constant selection, we find that the generating function for the conditional fixation time obtained using (2.15) reduces to (A.11) of MARTIN and LAMBERT [23].

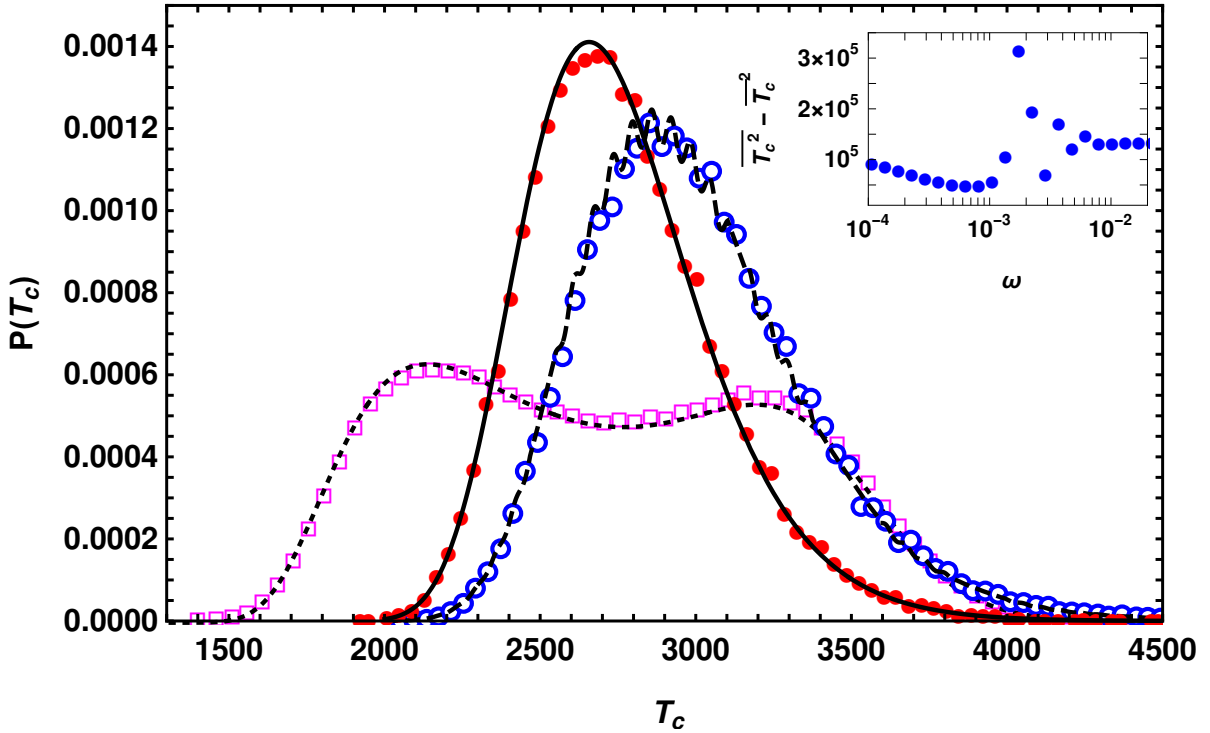


Fig. 2.7 Fixation time distribution, $P(T_c)$ in a finite population of size N when the mutant is beneficial at all times and has a changing selection coefficient given by $s(t) = \bar{s} + \sigma(\omega t + \theta)$. The points and the curves are obtained, respectively, from numerical simulations and the semi-deterministic result (2.15) for cycling frequency below, above and close to the resonance frequency ω_r , and given by $\omega = 10^{-4}$ (\bullet , solid line), 0.1 (\circ , dashed line) and 0.002 (\square , dotted line), respectively. The other parameters are $N = 10^5$, $\theta = 0$, $\bar{s} = 0.01$, $\sigma = 0.007$, $h = 0.5$. Note that while the distribution is bell-shaped away from the resonance frequency, it is bimodal close to ω_r which results in a large variance in the conditional fixation time (see inset). For the cycling frequency ω_r , the selection coefficient completes a full cycle and the mutant typically fixes when selection is increasing resulting in the bimodal character of the distribution. For the same reason, the distribution at high frequency also has multiple modes but with very small amplitude.

Figure 2.6a shows that, except for strongly recessive or dominant mutants, the conditional mean fixation time $\bar{T}_c = \int_0^\infty dT_c T_c P(T_c)$ depends weakly on dominance for arbitrary rate of environmental change. Figure 2.6a also suggests that for small and large cycling frequencies, the conditional mean fixation time $\bar{T}_c(h, s) \approx \bar{T}_c(1 - h, s)$. To

understand this result, using (2.15), we rewrite the time \bar{T}_c as

$$\bar{T}_c = \frac{h\Upsilon_A}{h-1} \int_0^\infty dq e^{-\Upsilon_A q} q^{-\frac{h}{1-h}} \int_0^\infty d\Upsilon_C T_c(\Upsilon_C) e^{-\Upsilon_C q} \quad (2.18)$$

and analyze it for slowly changing environments using a perturbation theory. As explained in Appendix A.5, for $\omega \ll s(0)$, we get $\bar{T}_c \approx \bar{T}_{0,c} + (\omega/s(0))\bar{T}_{1,c}$ where,

$$\frac{\bar{T}_{0,c}}{2N} \approx \frac{\ln(2\alpha)}{2h(1-h)\alpha} + \frac{\gamma + (2-3h)\ln h + (3h-1)\ln(1-h)}{2h(1-h)\alpha} \quad (2.19)$$

$$\begin{aligned} \frac{\bar{T}_{1,c}}{2N} &\approx -\frac{\sigma \cos \theta (\ln \alpha)^2}{4h^2(1-h)^2 s(0)\alpha} \\ &\times \left[1 + \frac{2(\ln 2 - h + \gamma + (2-3h)\ln h + (3h-1)\ln(1-h))}{\ln \alpha} \right] \end{aligned} \quad (2.20)$$

In the above equations, $\gamma \approx 0.577$ is the Euler constant and, as before, $\alpha = Ns(0)$. Note that while the fixation time \bar{T}_c was expanded in powers of $N\omega$ in the last section, here the expansion parameter is $\omega/s(0)$.

The semi-deterministic theory described in this section is not a systematic, controlled approximation (unlike various perturbation theories) and it is not clear how good this approximation is; however, here we find that (2.19) matches exactly with (4.2) of EWING *et al.* [31] (on replacing N in the above expression by $2N$) which is obtained using a diffusion theory, and shows that the conditional mean fixation time in a population with dominance coefficient h is approximately equal to that in a population with corresponding parameter $1-h$. Note that this result holds for large α ($\gtrsim 10^3$) while as shown in Fig. 2.4 for moderate selection, the dominant mutant takes longer than the recessive one to fix (see also TESHIMA and PRZEWORSKI [16]).

Equation (2.20) captures the effect of a slowly changing environment on the conditional mean fixation time and matches well with the data obtained by numerically integrating (2.18) as shown in Fig. 2.6b. The leading term on the right-hand side (RHS) of (2.20) is symmetric about $h = 1/2$ pointing to the approximate symmetry, $\bar{T}_c(h, s) \approx \bar{T}_c(1-h, s)$

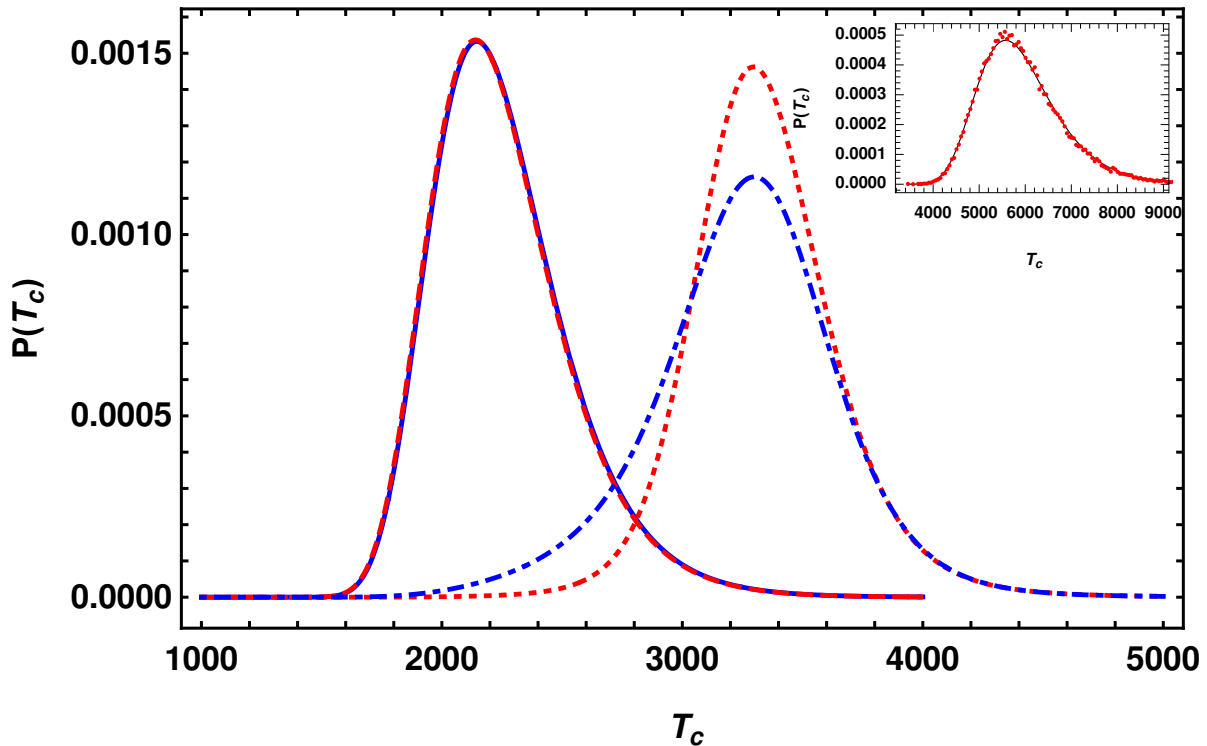


Fig. 2.8 Distribution of the conditional fixation time T_c in a finite population of size N when the mutant is beneficial at all times and has a changing selection coefficient given by $s(t) = \bar{s} + \sigma \sin(\omega t + \theta)$. The data are obtained within a semi-deterministic approximation and given by (2.15) when a mutant is beneficial at all times to show that it does not have the $h \leftrightarrow 1 - h$ symmetry in fast changing environments. The overlapping left curves for $\omega = 10^{-4}$ show the distribution for $h = 0.7$ (solid blue) and 0.3 (dashed red), and the right curves are for $\omega = 0.002$ where $h = 0.7$ (dashdotted blue) and 0.3 (dotted red). The other parameters are $N = 10^5$, $\theta = \pi/4$, $\bar{s} = 0.01$, $\sigma = 0.007$. Inset: Distribution of the fixation time of an initially beneficial mutant when the time-averaged selection is zero and the environment changes slowly. The solid curve is obtained from (2.15) and the points are generated from simulations. The parameters are $N = 10^5$, $\sigma = 0.007$, $\omega = 10^{-6}$, $h = 0.3$, $\theta = \pi/4$.

discussed above. While the subleading correction does not have $h \leftrightarrow 1 - h$ symmetry but its effect is small compared to the leading term for intermediate dominance. Figure 2.8 further suggests that the distribution of the fixation time has $h \leftrightarrow 1 - h$ approximate symmetry for small cycling frequencies but not for frequencies close to the resonance frequency, in accordance with the behavior of the mean fixation time shown in Fig. 2.6a.

Equation (2.20) also shows that the mean fixation time decreases (increases) if the beneficial allele arises when the selection gradient (in time) is positive (negative), as intuitively expected; furthermore, both $\bar{T}_{0,c}$ and $\bar{T}_{1,c}$ decay with increasing selection. Finally, we mention that at the beginning of this section, we had assumed that $\sigma < \bar{s}$. But in a slowly changing environment, the semi-deterministic approximation may be expected to work for $\sigma > \bar{s}$ and $0 < \theta < \pi$; this is indeed confirmed in Fig. 2.8.

2.6 Discussion

In this Chapter, we have studied how a selective environment that is varying periodically and predictably in time affects the fixation time of a mutant in a finite, diploid population. **Effect of the environmental parameters:** We find that if the environment changes fast, the fixation time in the temporally varying environment differs considerably from that in the static environment, as can be seen in Figs. 2.1a and 2.6a at intermediate cycling frequencies. But for a meaningful comparison with the body of work on the fixation time in constant environments [17], most of our analysis has focused on the effect of slowly changing environments.

It should be noted that for time-dependent selection coefficients, the stochastic process is time-inhomogeneous and therefore the fixation time depends on the time at which the mutant arose and whether the environment is improving or deteriorating [9, 12]. If an initially deleterious mutant on the way to fixation experiences a more favorable environment, its chance of extinction reduces and such a mutant can be expected to have a larger time of fixation than in an environment that remained unfavorable. On the other hand, if an initially deleterious (or beneficial) mutant faces an even more unfavorable environment, due to the higher risk of extinction at late times, the mutant is likely to fix sooner. The dynamics of an initially beneficial mutant that remains beneficial until fixation are, however, less affected by the random genetic drift - if selection increases,

due to the higher growth rate, the mutant fixes sooner than in an environment that had remained constant.

Selection regimes: In static environments, the qualitative behavior of the conditional mean fixation time of a mutant depends on the sign and strength of the scaled selection coefficient, $\alpha = Ns$. For a beneficial mutant, if selection is weak ($0 < \alpha \ll 1$), the fixation time increases with the dominance coefficient h and selection strength α , and can even exceed the fixation time of a neutral mutant [18]. But for moderately strong selection ($1 \ll \alpha \ll 100$), the conditional mean fixation time decreases with α and increases with h [16, 17]. For stronger selection, it decreases with α and is approximately same for two mutants with the same selection coefficient but dominance coefficients h and $1 - h$ [32, 31]. The patterns for deleterious mutations follow on realizing that the conditional fixation time for a beneficial mutant with dominance coefficient h and a deleterious mutant with the same magnitude of selection but dominance level $1 - h$ are equal [6, 7].

Here, we find that all the qualitative patterns described above continue to hold when the environment changes slowly but there are quantitative differences. While a slowly changing environment has only a mild effect on the conditional mean fixation time if the mutant is beneficial, its impact is much stronger for deleterious mutants. This asymmetry can be traced back to the fact that the fixation probability of a deleterious mutant is much more sensitive to a change in selection than the fixation probability of a beneficial mutant. Furthermore, for an initially deleterious mutant, as the fixation probability of the recessives is higher than the dominants, the former can segregate in the population for a longer time and are therefore exposed to the changing environment for a longer duration resulting in a fixation time which is substantially different from that in the constant environment.

Implications: In a constant environment, due to the Maruyama-Kimura symmetry for the conditional mean fixation time [6, 7], similar diversity patterns for beneficial

and deleterious sweeps may be generated [8]. But in a changing environment, due to the lack of the Maruyama-Kimura symmetry, a beneficial mutant in a slowly improving (deteriorating) environment can generate a diversity pattern different from that due to the fixation of a deleterious mutant under the same selection pressure but in a slowly deteriorating (improving) environment.

To ascertain this effect, we have conducted a preliminary study of the effect of hard sweep on linked neutral variation in a two-locus model of a finite, diploid population in which the first locus is modeled as in the Model section but the second biallelic locus is neutral. Recombination is assumed to occur with a probability $c \ll 1$ but recurrent mutations are not allowed. As a result of genetic hitchhiking, the heterozygosity H_{fix} at the neutral locus following the fixation event (relative to the heterozygosity H_0 before the new mutant appeared) is expected to decrease [1]. Figure 2.9 shows the relative heterozygosity as a function of the cycling frequency for the parameters in Fig. 2.1a, and we find that its qualitative behavior is the same as that of the conditional mean fixation time [3]. The inset of Fig. 2.9 emphasizes that the slowly changing selective environment only mildly influences the heterozygosity of an initially beneficial dominant mutant but it has a strong effect on the heterozygosity of an initially deleterious recessive mutant. A more detailed study of the effect of environmental variation on various measures of genetic diversity will be taken up in a future work.

As already mentioned, although the qualitative patterns for the fixation time in a static environment are robust with respect to a slow change in the environment, there are quantitative differences. As a consequence, the effect of varying environment may be interpreted as an effective selection coefficient or dominance parameter. For example, in Fig. 2.4, the fixation time in slowly changing environment, and for dominance coefficient $h = 1/2$ and selection strength $\alpha = -14.14$ ($\theta = 5\pi/4$) is about 1691. But if one assumes

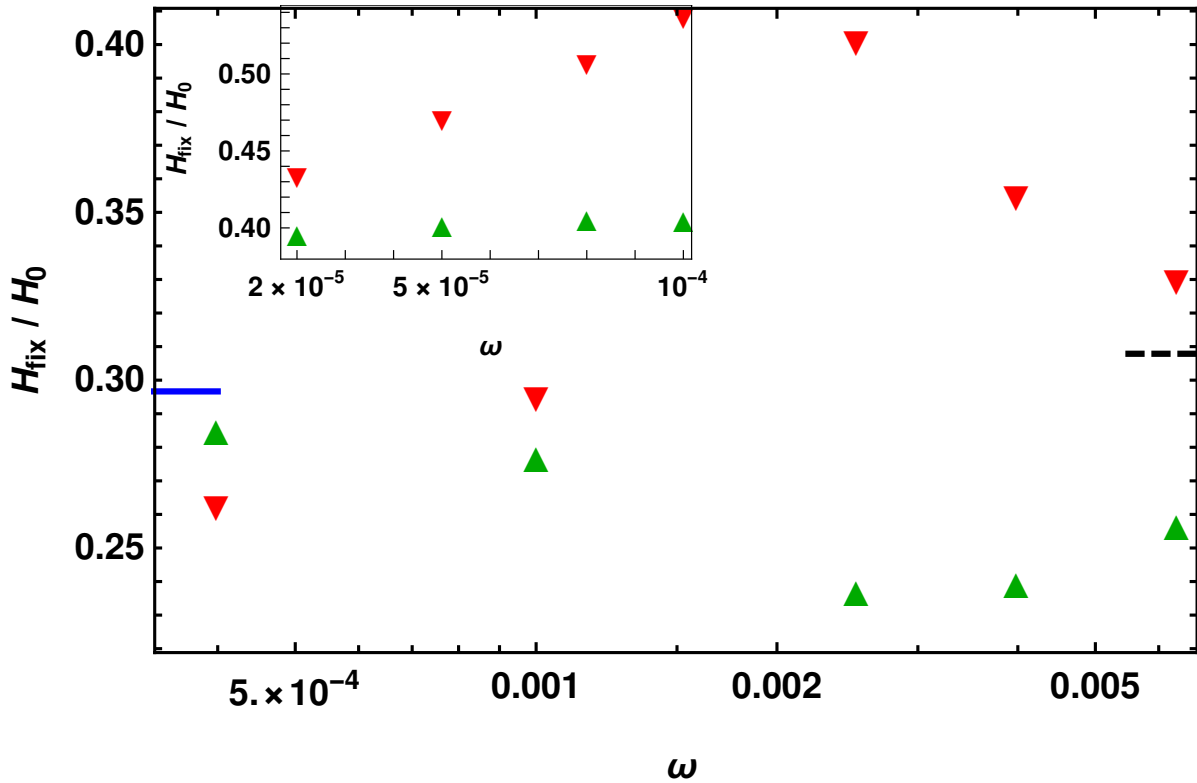


Fig. 2.9 Reduction in the mean heterozygosity at the linked neutral locus due to the fixation of dominant ($h = 0.7$, \blacktriangle) and recessive mutant ($h = 0.3$, \blacktriangledown) with selection coefficient $s(t) = \sigma \sin(\omega t + \pi/4)$ and $-s(t)$, respectively, to show that the heterozygosity following the fixation of deleterious and beneficial mutant is significantly different in the changing environment, unlike in the static environment where they are identical. The reduction in the mean heterozygosity in the static environment with selection (solid line) and in the neutral environment (dashed line) are also shown for comparison. All the data are obtained by numerical simulations. The inset shows the heterozygosity reduction due to the fixation of dominant mutant ($h = 0.9$, \blacktriangle) and recessive mutant ($h = 0.1$, \blacktriangledown) in the slowly changing environment with selection coefficient $s(t) = \sigma \sin(\omega t + 3\pi/4)$ and $-s(t)$, respectively. The other parameters for both figures are $\sigma = 0.01$, $c = s(0)/10$, $Q_0 = 0$, $R_0 = 0.2002$ and $p_0 = 1/2N$, where c is the recombination probability, Q_0 and R_0 are, respectively, initial relative frequencies of a neutral allele in a chromosome with a and A allele on the first locus, and the population size $N = 500$ (main) and 1000 (inset).

a constant environment, the same fixation time is obtained for $\alpha = -15.27$ which implies an 8% increase in the selection coefficient.

We therefore suggest to include the effect of changing environment in theoretical models of selective sweeps as this can potentially allow one to distinguish between the

sign of selection, detect deleterious sweeps and correctly estimate the model parameters. Generalizing the above results to include the effect of inbreeding and sex-linked inheritance [33, 34] could also help to assess the importance of changing environment in evolutionary dynamics.

Copyright and Permission

The results of this chapter have been published in:

Kaushik, S., and K. Jain, 2021 Time to fixation in changing environments, *Genetics* **219**, iyab 148.

©Oxford University Press. Reproduced with permission. All rights reserved.

We have reproduced the materials here with kind permission from the Oxford University Press. See the publication link below:

<https://academic.oup.com/genetics/article-abstract/219/3/iyab148/6369518?redirectedFrom=fulltext>

Bibliography

- [1] MAYNARD SMITH, J. and J. HAIGH, 1974 Hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- [2] STEPHAN, W., 2016 Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol Ecol* **25**: 79–88.
- [3] TAJIMA, F., 1990 Relationship between DNA polymorphism and fixation time. *Genetics* **125**: 447–454.
- [4] EWENS, W., 1973 Conditional diffusion processes in population genetics. *Theo. Pop Biol* **4**: 21–30.
- [5] ZHAO, L., M. LASCoux, A. D. J. OVERALL, and D. WAXMAN, 2013 The characteristic trajectory of a fixing allele: a consequence of fictitious selection that arises from conditioning. *Genetics* **195**: 993–1006.
- [6] MARUYAMA, T., 1974 The age of an allele in a finite population. *Genetics. Res. Camb.* **23**: 137–143.
- [7] MARUYAMA, T. and M. KIMURA, 1974 A note on the speed of gene frequency changes in reverse directions in a finite population. *Evolution* **28**: 161–163.

-
- [8] JOHRI, P., B. CHARLESWORTH, E. K. HOWELL, and M. LYNCH, 2021 Revisiting the notion of deleterious sweeps. *Genetics* **219**:3 iyab094.
- [9] UECKER, H. and J. HERMISSON, 2011 On the fixation process of a beneficial mutation in a variable environment. *Genetics* **188**: 915–930.
- [10] WILLIAMS, C. M., G. J. RAGLAND, G. BETINI, L. B. BUCKLEY, Z. A. CHEVIRON, K. DONOHUE, J. HEREFORD, M. M. HUMPHRIES, S. LISOVSKI, K. E. MARSHALL, P. S. SCHMIDT, K. S. SHELDON, O. VARPE, and M. E. VISSER, 2017 Understanding evolutionary impacts of seasonality: an introduction to the symposium. *Integrative and Comparative Biology* **57**: 921–933.
- [11] HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection. V. *Proc. Camb. Philos. Soc.* **23**: 838–844.
- [12] DEVI, A. and K. JAIN, 2020 The impact of dominance on adaptation in changing environments. *Genetics* **216**: 227–240.
- [13] NAGYLAKI, T., 1992 *Introduction to Theoretical Population Genetics*. Springer-Verlag.
- [14] KARLIN, S. and H. M. TAYLOR, 1975 *A first course in stochastic processes*. Academic Press.
- [15] KIMURA, M. and T. OHTA, 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763–771.
- [16] TESHIMA, K. M. and M. PRZEWORSKI, 2006 Directional positive selection on an allele of arbitrary dominance. *Genetics* **172**: 713–718.
- [17] CHARLESWORTH, B., 2020 How long does it take to fix a favorable mutation, and why should we care? *Am. Nat.* **195**: 753–771.

-
- [18] MAFESSONI, F. and M. LACHMANN, 2015 Selective strolls: fixation and extinction in diploids are slower for weakly selected mutations than for neutral ones. *Genetics* **201**: 1581–1589.
- [19] EWENS, W., 2004 *Mathematical Population Genetics*. Springer, Berlin.
- [20] RISKEN, H., 1996 *The Fokker Planck equation. Methods of solution and applications*. Springer, Berlin.
- [21] KIMURA, M., 1957 Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**: 882–901.
- [22] COHN, H. and P. JAGERS, 1994 General branching processes in varying environment. *Ann. Applied Prob.* **4**: 184–193.
- [23] MARTIN, G. and A. LAMBERT, 2015 A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theo Pop Biol* **101**: 40–46.
- [24] FELLER, W., 1951a Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 227–246.
- [25] FELLER, W., 1951b Two singular diffusion problems. *Annals of Mathematics* **54**: 173–182.
- [26] CATTIAUX, P., P. COLLET, A. LAMBERT, S. MARTINEZ, S. S MÉLÉARD, and J. SAN MARTIN, 2009 Quasi-stationary distributions and diffusion models in population dynamics. *The Annals of Probability* **37**: 1926–1969.
- [27] MASOLIVER, J. and J. PERELLÓ, 2012 First-passage and escape problems in the Feller process. *Phys. Rev. E* **86**: 041116.
- [28] GAN, X. and D. WAXMAN, 2015 Singular solution of the Feller diffusion equation via a spectral decomposition. *Phys. Rev. E* **91**: 012123.

-
- [29] MASOLIVER, J., 2016 Nonstationary Feller process with time-varying coefficients. *Phys. Rev. E* **93**: 012122.
- [30] KENDALL, D. G., 1948 On the generalized “birth-and-death” process. *Ann. Math. Stat.* **19**: 1–15.
- [31] EWING, G., J. HERMISSON, P. PFAFFELHUBER, and J. RUDOLF, 2011 Selective sweeps for recessive alleles and for other modes of dominance. *J. Math. Biol.* **63**: 399–431.
- [32] VAN HERWAARDEN, O. A. and N. J. VAN DER WAL, 2002 Extinction time and age of an allele in a large finite population. *Theo. Pop. Biol.* **61**: 311–318.
- [33] GLÉMIN, S., 2012 Extinction and fixation times with dominance and inbreeding. *Theo. Pop. Biol.* **81**: 310–316.
- [34] HARTFIELD, M. and T. BATAILLON, 2020 Selective sweeps under dominance and inbreeding. *G3: Genes, Genomes, Genetics* **10**: 1063–1075.

Chapter 3

Joint effect of changing selection and demography on the site frequency spectrum

3.1 Introduction

In the previous chapter, we assumed that population size was constant; in this chapter, we relax that assumption. Here, we will investigate the site frequency spectrum in a randomly mating diploid population where both population size and selection coefficient change periodically over time.

Despite several decades of intense research, the factors that determine the genetic variation in a population are not completely understood [1, 2]. An important statistic for measuring the within-population genetic diversity is the site frequency spectrum (SFS) which gives the (unnormalized) distribution of allele frequency across polymorphic loci in the genome. It is known that in a neutral population of constant size, the equilibrium SFS decreases with the allele frequency of the derived allele whereas it is a U-shaped function when the mutant allele is under constant, positive selection [3, 4]. However, these

classical results can change due to complex evolutionary processes such as hitchhiking of neutral alleles with a beneficial mutant [5, 6], and in populations with high mutation rate [7, 8], fat-tailed offspring distribution [9], seed bank [10], etc..

The above discussion assumes that the environment remains constant over evolutionary time scales. But in a changing environment, in general, there is no stationary state and the nonequilibrium SFS can differ significantly from the equilibrium SFS in a constant environment. For example, in an expanding neutral population, there is an excess of rare variants as compared to when the population size remains constant at its initial size, while a population bottleneck leads to a deficit in the low-frequency alleles due to the stronger effect of random genetic drift [11]. The SFS is also related to other measures of genetic diversity such as mean heterozygosity and mean number of segregating sites which have been extensively studied when the population size is variable and selection is absent [12–14], and in recent years, there has been some progress in understanding the effect of demography on SFS in a population under constant selection [15–18]. However, the impact of temporally changing selection on the nonequilibrium SFS is little studied [19, 20]) and to our knowledge, the *joint* effect of time-dependent selection and demography has not been investigated so far.

Furthermore, most studies on the nonequilibrium SFS are concerned with historical demography such as population expansion in the human population at the end of the last ice age [17] or population bottleneck in the non-African population of *D. melanogaster* around 6000 years ago [21]. However, variation in population size occurs over short time scales also, for which different demographic models are needed. An example is the oscillatory change in the prey and predator population size over the course of a few generations (see, for e.g., Chapter 3 of MURRAY [22]).

To better understand the SFS in populations of varying size and in changing selective environments, here we consider a randomly mating, diploid population in which either

population size and/or the selection coefficient change with time in a periodic fashion. We will explore the effect of selection strength, frequency of environmental variation, and genetic dominance on the time-dependent SFS, and focus on how the time-averaged SFS deviates from the equilibrium SFS in a constant environment.

3.2 Model

We consider a randomly mating, diploid population of size $N(t)$ at time t in which an individual's genome is described by an infinite-sites model [23] that makes the following key assumptions: first, a large number of sites are available for mutation which occurs with a small rate μ per site so that every mutation occurs (irreversibly) at a new site, and second, recombination is free so that at each site, allele frequency dynamics occur independently.

Furthermore, at each site, the homozygotes AA and aa have fitness 1 and $1 + s(t)$, respectively, and the fitness of the heterozygote aA is $1 + hs(t)$ where the dominance coefficient $0 < h < 1$. The frequency x of the mutant allele a increases or decreases at rate r_b or r_d respectively, which are given by [24, 25]

$$r_b(t) = 2N(t) \frac{x(1-x)w_a(t)}{\bar{w}(t)} \quad (3.1)$$

$$r_d(t) = 2N(t) \frac{x(1-x)w_A(t)}{\bar{w}(t)} \quad (3.2)$$

where $w_a(t) = (1 + s(t))x + (1 + hs(t))(1 - x)$ and $w_A(t) = (1 - x) + (1 + hs(t))x$ are the marginal fitness of allele a and A , respectively, and $\bar{w}(t) = w_a(t)x + w_A(t)(1 - x)$ is the average fitness. The population size and the selection coefficient are assumed to vary in time in a periodic fashion. We therefore write the selection coefficient as

$$s(t) = \bar{s} + \sigma \sin(\omega t) \quad (3.3)$$

where the average \bar{s} is arbitrary but the amplitude $\sigma \geq 0$. The population size varies deterministically as

$$N(t) = \bar{N}(1 + \nu \sin(\Omega t)) \quad (3.4)$$

where $0 \leq \nu < 1$ to ensure that the total population size remains strictly positive at all times. In general, the variation in population size and selection can have a phase difference also. But, unless specified otherwise, we will assume that these two parameters vary in phase.

3.3 Site frequency spectrum: diffusion theory

To obtain an analytical understanding of the SFS and other measures of genetic diversity such as the mean heterozygosity, we work in the framework of the diffusion theory. It has been shown that in a large population, the average (unnormalized) density $f(x, t)$ of polymorphic sites with mutant allele frequency x at time t obeys the following partial differential equation: $\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x}[a(x, t)f(x, t)] + \frac{\partial^2}{\partial x^2}[b(x, t)f(x, t)]$ where $a(x, t)$ and $b(x, t)$ are, respectively, the mean and variance of the transition probability [26].

For the model detailed in Section 3.2, these coefficients have been derived in our earlier work [25], from which we find that

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} &= -s(t) \frac{\partial}{\partial x} [x(1-x)(x+h(1-2x))f(x, t)] \\ &+ \frac{\partial^2}{\partial x^2} \left[\frac{x(1-x)}{2N(t)} f(x, t) \right]. \end{aligned} \quad (3.5)$$

The above equation is subject to the initial condition $f(x, 0)$, and boundary conditions, $\lim_{x \rightarrow 0} xf(x, t) = 2N(t)\mu$ and $f(1, t)$ finite [26]. The first and second term on the RHS of (3.5) describe the effect of selection and random genetic drift, respectively, and the boundary condition at low allele frequency models the balance between the loss of polymorphism due to the absorption of the mutant allele by genetic drift alone and the

gain by new mutations (for other modeling approaches, see ŽIVKOVIĆ *et al.* [18], GRAVEL [27]).

However, for both numerical and analytical purposes, it is easier to work with the density $g(x, t) = x(1 - x)f(x, t)$ which obeys [26],

$$\begin{aligned} \frac{\partial g(x, t)}{\partial t} &= -s(t)x(1 - x)\frac{\partial}{\partial x} [(x + h(1 - 2x))g(x, t)] \\ &+ \frac{x(1 - x)}{2N(t)}\frac{\partial^2 g(x, t)}{\partial x^2} \end{aligned} \quad (3.6)$$

with initial condition $g(x, 0) = x(1 - x)f(x, 0)$, and boundary conditions

$$g(0, t) = \theta\rho(t), \quad g(1, t) = 0 \quad (3.7)$$

where

$$\theta \equiv 2\bar{N}\mu, \quad \rho(t) \equiv \frac{N(t)}{\bar{N}} \quad (3.8)$$

A standard approach to solve partial differential equations such as (3.6) is to carry out an eigenfunction expansion of its solution; unfortunately, the required eigenfunctions are not known in a closed form when selection is present [28]. However, as detailed in the following section, it is possible to make analytical progress even when the mutant allele is under selection in the limiting cases where the environment is changing either slowly or rapidly [29].

In the following, we will also study the unfolded sample frequency spectrum defined as the expected number of sites at which exactly $1 \leq i \leq n - 1$ individuals in a population sample of size $n \ll N$ carry a mutation which is given by [30]

$$f_{n,i}(t) = \binom{n}{i} \int_0^1 dx x^i (1 - x)^{n-i} f(x, t). \quad (3.9)$$

The sample SFS is also related to the mean heterozygosity,

$$H(t) \equiv f_{2,1}(t) = 2 \int_0^1 dx x(1-x)f(x,t) . \quad (3.10)$$

We will mainly focus on the time-averaged statistics obtained by averaging the time-dependent quantity $Q(t)$ over the slower cycle, that is,

$$\bar{Q} = \frac{\min(\omega, \Omega)}{2\pi} \int_0^{\frac{2\pi}{\min(\omega, \Omega)}} dt Q(t) \quad (3.11)$$

and compare the results with those in the constant environment to ascertain the effect of the changing environment.

3.4 Results

3.4.1 Static environment

We first consider the equilibrium SFS which is obtained at large times when the selection coefficient and population size are time-independent. Although an exact expression for it has been known for a long time, simple analytical approximations have not been obtained and therefore, below we proceed to derive them.

In a constant environment, polymorphic sites are lost due to absorption of the mutant allele and created by new mutations thus resulting in a steady state. Therefore, for $s(t) = \bar{s}$, $N(t) = \bar{N}$, on setting the LHS of (3.5) equal to zero, the stationary state SFS is found to be [3, 31, 4, 23, 16]

$$f^*(x) = \frac{\theta e^{\bar{N}\bar{s}[(1-2h)x^2+2hx]} \int_x^1 dy e^{-\bar{N}\bar{s}[(1-2h)y^2+2hy]}}{x(1-x) \int_0^1 dy e^{-\bar{N}\bar{s}[(1-2h)y^2+2hy]}} . \quad (3.12)$$

Equation 3.12 can also be derived by noting that as θ mutations occur per unit time per site in the population, $f^*(x)dx/\theta$ is the mean sojourn time of the mutant allele between frequency x and $x + dx$ before absorption, when the mutant allele is initially present in a single copy [32].

Weak selection

A power series expansion of (3.12) in scaled selection strength ($\bar{N}\bar{s}$) yields

$$f^*(x) \approx \frac{\theta}{x} - \frac{\theta}{3}\bar{N}\bar{s}[1 - 2x + h(4x - 5)], \bar{N}|\bar{s}| \ll 1 \quad (3.13)$$

which, as for a neutral population, decreases monotonically with the allele frequency. This is because at most sites, single mutant will quickly get lost due to genetic drift, and therefore the average number of loci with very low mutant frequency would be high, and only a small number of loci will have intermediate to high allele frequency.

Strong selection

As explained in Appendix B.1, when selection is strong, the integrals in (3.12) can be estimated using (B.1.2) for $h \neq 0, 1$, and we obtain

$$f^*(x) \approx \begin{cases} \frac{\theta}{x(1-x)} \frac{h}{h + x(1-2h)} & , \bar{N}\bar{s} \gg 1 \\ \frac{\theta e^{-\bar{N}|\bar{s}|(1-\ell(x))}}{x(1-x)} & , \bar{N}\bar{s} \ll -1 \end{cases} \quad (3.14a)$$

$$(3.14b)$$

where $\ell(x) = (1-x)[(1-x) + 2x(1-h)]$ (for completely dominant and recessive mutant allele, see Appendix B.1). Figures 3.1a and 3.1b show that when selection is strong, except for high allele frequencies, the exact equilibrium SFS given by (3.12) is in good agreement with the approximate expressions (3.14a) and (3.14b).

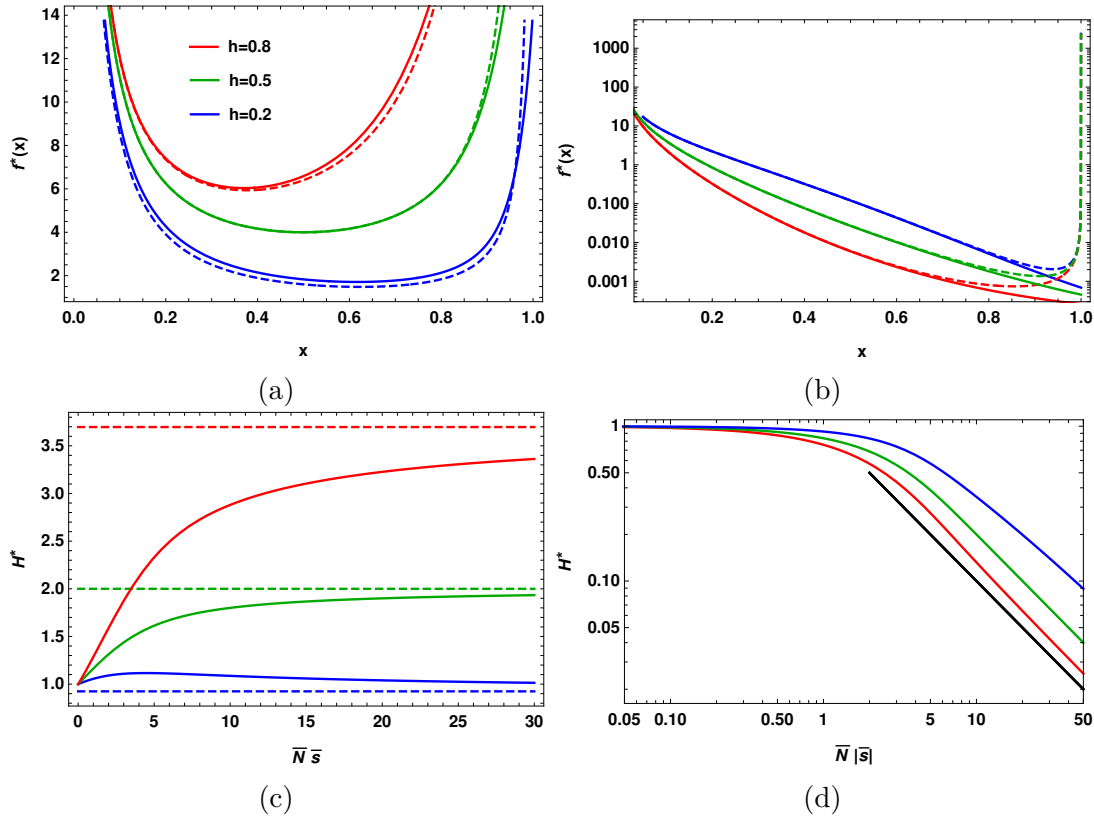


Fig. 3.1 Static environment: The top panel shows the site frequency spectrum of a large population for selection strength (a) $\bar{N}\bar{s} = 30$, and (b) $\bar{N}\bar{s} = -10$. The solid lines show the exact expression (3.12), while the dashed lines show the approximate expression (3.14a) and (3.14b) for beneficial and deleterious mutations, respectively. The bottom panel shows the mean heterozygosity calculated numerically using (3.12) for (c) a positively selected mutant which approaches the asymptotic value (3.16b) depicted by dashed lines, and (d) a negatively selected mutant that decreases towards zero as $1/\bar{N}|\bar{s}|$ (black curve) for all h in accordance with (3.16c). In all the plots, the scaled mutation rate $\theta = 1$.

Equations (3.14a) and (3.14b) show that for $x \ll (\bar{N}|\bar{s}|)^{-1}$, $f^*(x) \approx \theta/x$ so that at low allele frequencies, the SFS for strong selection behaves the same way as that for a neutral population, and the effect of selection becomes apparent only at intermediate and high allele frequencies. For negative selection, the SFS decreases monotonically towards zero with the mutant allele frequency for the same reason as for the neutral population. But for positive selection, it has the characteristic U-shape reflecting the fact that the mean sojourn time is long at high allele frequencies for beneficial mutations [32].

From (3.14a) and (3.14b), we also find that for a given allele frequency, the SFS increases with the dominance coefficient for beneficial mutations and decreases for deleterious ones; this is due to the longer sojourn time to fixation for dominant, advantageous mutation and the fact that the deleterious recessives are more likely than the dominant alleles to reach finite frequencies [16].

Sample SFS

Using the approximate results (3.13), (3.14a) and (3.14b) for a large population in (3.9), we find the sample SFS to be (see Appendix B.1 for details)

$$f_{n,i}^* \approx \begin{cases} \frac{\theta}{i} + \frac{\theta[h(5n-4i+6)+2i-n]\bar{N}\bar{s}}{3(n+1)(n+2)} & , -1 \ll \bar{N}\bar{s} \ll 1 & (3.15a) \\ \frac{n\theta}{i(n-i)} {}_2F_1\left(1, i; n; 2 - \frac{1}{h}\right) & , \bar{N}\bar{s} \gg 1 & (3.15b) \\ \binom{n}{i} \frac{\theta\Gamma(i)}{(2\bar{N}|\bar{s}|h)^i} & , \bar{N}\bar{s} \ll -1 & (3.15c) \end{cases}$$

where $\Gamma(i)$ is the gamma function and ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function [33]. For $i \ll n-1$, the contribution to the integral on the RHS of (3.9) comes from small x , while for large i , variants at high frequency matter; as a result, the sample SFS behaves qualitatively in the same manner as the SFS for a large population discussed above.

The neutral mean heterozygosity in a constant size population is given by θ . From (3.15a)-(3.15c), we find that when the mutant allele is under selection, the mean heterozygosity can be approximated by

$$H^* \approx \begin{cases} \theta + \frac{\theta h}{3} \bar{N}\bar{s} & , -1 \ll \bar{N}\bar{s} \ll 1 & (3.16a) \\ \frac{2h\theta}{1-2h} \ln\left(\frac{1-h}{h}\right) & , \bar{N}\bar{s} \gg 1 & (3.16b) \\ \frac{\theta}{h\bar{N}|\bar{s}|} & , \bar{N}\bar{s} \ll -1 & (3.16c) \end{cases}$$

As positive (negative) selection decreases (increases) the chance of loss of the new mutant allele, one expects the heterozygosity to be enhanced (reduced) with increasing magnitude of selection for beneficial (deleterious) mutants; indeed, (3.16a) and (3.16c) are consistent with this expectation. But, for strong positive selection, as Figure 3.1c shows, the mean heterozygosity does not necessarily increase with selection. In fact, on equating the RHS of (3.16b) to θ , we find that the mean heterozygosity saturates to a value smaller than that for the neutral allele for $h < \frac{1}{1-2W_{-1}\left(-\frac{1}{2\sqrt{e}}\right)} \approx 0.221589$ where $W_{-1}(z)$ is the lower branch of the Lambert W function [34]. We also note that with increasing magnitude of selection, the mean heterozygosity approaches the asymptotic value, viz., (3.16b) for a beneficial mutant and zero for a deleterious one, algebraically slowly (see (B.1.6a) and (B.1.6b), and Fig. 3.1d).

For positively selected, codominant mutant, on taking the limit $h \rightarrow 1/2$ in (3.16b), we find that the mean heterozygosity is equal to 2θ . Equation (3.16b) also shows that H^* for a dominant (recessive) allele is larger (smaller) than that for the co-dominant allele; the mean heterozygosity increases with the dominance level since the mutant's chance of escaping stochastic loss is enhanced with increasing h [35], and hence it can contribute to the population diversity before eventually fixing. Similarly, one can argue that for negative selection, H^* will decrease with increasing h (see Figs. 3.1c and 3.1d).

In the following discussion, we will denote the equilibrium SFS (and, similarly, heterozygosity) for positive and negative selection by $f_p^*(x)$ and $f_n^*(x)$, respectively.

3.4.2 Slowly changing environment

We now turn to the properties of the SFS in a slowly changing environment. If the frequency with which the selective environment and population size change are much smaller than the inverse average population size and average selection coefficient (that is, $\omega, \Omega \ll \bar{N}^{-1}, \bar{s}$), one can obtain the time-dependent SFS within an *adiabatic approxima-*

tion [29]. For this purpose, we first rewrite (3.5) as

$$\begin{aligned} \bar{N}\omega \frac{\partial f(x, \psi)}{\partial \psi} &= -\bar{N}s(\psi) \frac{\partial}{\partial x} [x(1-x)(x+h(1-2x))f(x, \psi)] \\ &+ \frac{1}{2\rho(\Psi)} \frac{\partial^2}{\partial x^2} [x(1-x)f(x, \psi)] \end{aligned} \quad (3.17)$$

where $\psi = \omega t$, $\Psi = \Omega t$ and the ratio $0 < \Omega/\omega < \infty$. For arbitrary scaled selection, on expanding $f(x, \psi)$ as a power series in $\bar{N}\omega$ and substituting it in the above equation, we find that to leading order in $\bar{N}\omega$, the LHS of (3.17) is zero. It is easy to verify that this result also holds when either the population size or the selection coefficient varies with time. Thus in a slowly changing environment, the mean density $f(x, t)$ of sites with allele frequency $0 < x < 1$ obeys the steady state equation at *instantaneous* $N(t)$ and $s(t)$, and one can therefore obtain $f(x, t)$ by simply letting $\bar{N} \rightarrow N(t)$, $\bar{s} \rightarrow s(t)$, $\theta \rightarrow \theta\rho(t)$ in (3.12) and arrive at

$$\frac{f(x, t)}{\theta\rho(t)} \approx \frac{e^{N(t)s(t)[(1-2h)x^2+2hx]}}{x(1-x)} \frac{\int_x^1 dy e^{-N(t)s(t)[(1-2h)y^2+2hy]}}{\int_0^1 dy e^{-N(t)s(t)[(1-2h)y^2+2hy]}}. \quad (3.18)$$

Figure 3.2a shows that (3.18) is in good agreement with the results obtained by numerically integrating (3.5) for small cycling frequencies. We find that for strong positive selection, $f(x, t)$ is nonzero and increases (decreases) with increasing (decreasing) selection strength but it is close to zero when the selection strength is small or negative; this behavior may be understood by appealing to the results (3.14a) and (3.14b) in static environments. It is evident from (3.18) that when either N or s is time-dependent, $f(x, t)$ is a periodic function in time with the period of the varying parameter; if both N and s vary such that $\omega = (n_1/n_2)\Omega$ with n_1, n_2 being integers and $n_1 < n_2$ (say), the period of $f(x, t)$ is given by $2\pi n_1/\Omega$.

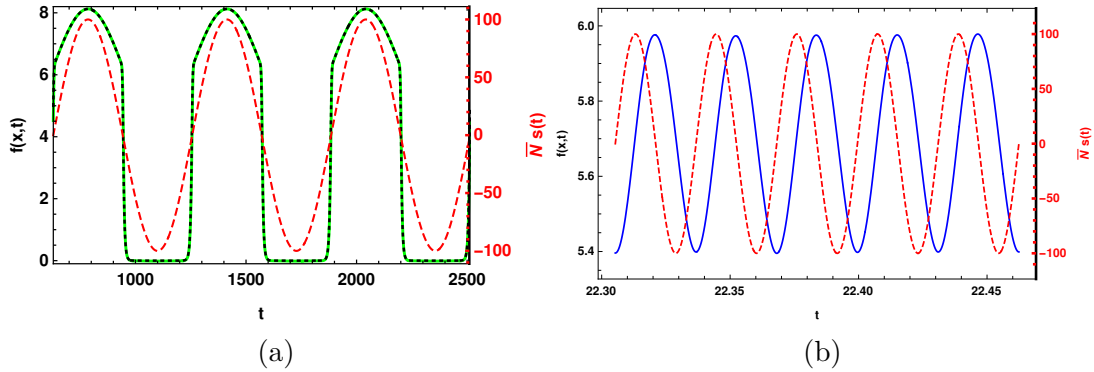


Fig. 3.2 Changing environment and on-average neutral selection: The time-dependent site frequency spectrum (solid line) in a large population when the population size and selection coefficient (dashed line) change periodically in time with equal cycling frequency obtained by numerically solving (3.5) for (a) $\bar{N}\omega = 0.01$ and (b) $\bar{N}\omega = 500$. The result (3.18) obtained within adiabatic approximation (dotted line) is also shown for comparison in the slowly changing environment. The other parameters are $x = 0.2$, $\nu = 0.3$, $h = 1/2$, $\theta = 1$ and $\bar{N}\sigma = 100$.

No selection

In the absence of selection, (3.18) gives $f(x, t) = \theta\rho(t)/x$ from which it immediately follows that the time-averaged SFS, $\bar{f}(x) = \frac{\Omega}{2\pi} \int_0^{2\pi/\Omega} dt' f(x, t')$ is simply θ/x and the time-averaged sample SFS is equal to θ/i . In Appendix B.3, these results are obtained using an eigenfunction expansion of $f(x, t)$ which furthermore shows that the correction to the adiabatic approximation is of order $(\bar{N}\Omega)^2$, see (B.3.17). Figure 3.3 shows the sample SFS for various Ω 's, and we find that it is well approximated by θ/i when the population size changes slowly.

Weak selection

When both population size and selection vary with the same frequency, using (3.11) and (3.13), we find that the time-averaged SFS, $\bar{f}(x)$ differs from $f^*(x)$ by an amount

$$\delta\bar{f} = \bar{f}(x) - f^*(x) = -\frac{\theta}{3}\bar{N}\bar{s} \left(\frac{\nu^2}{2} + \frac{\nu\sigma}{\bar{s}} \right) [1 - 2x + h(4x - 5)]. \quad (3.19)$$

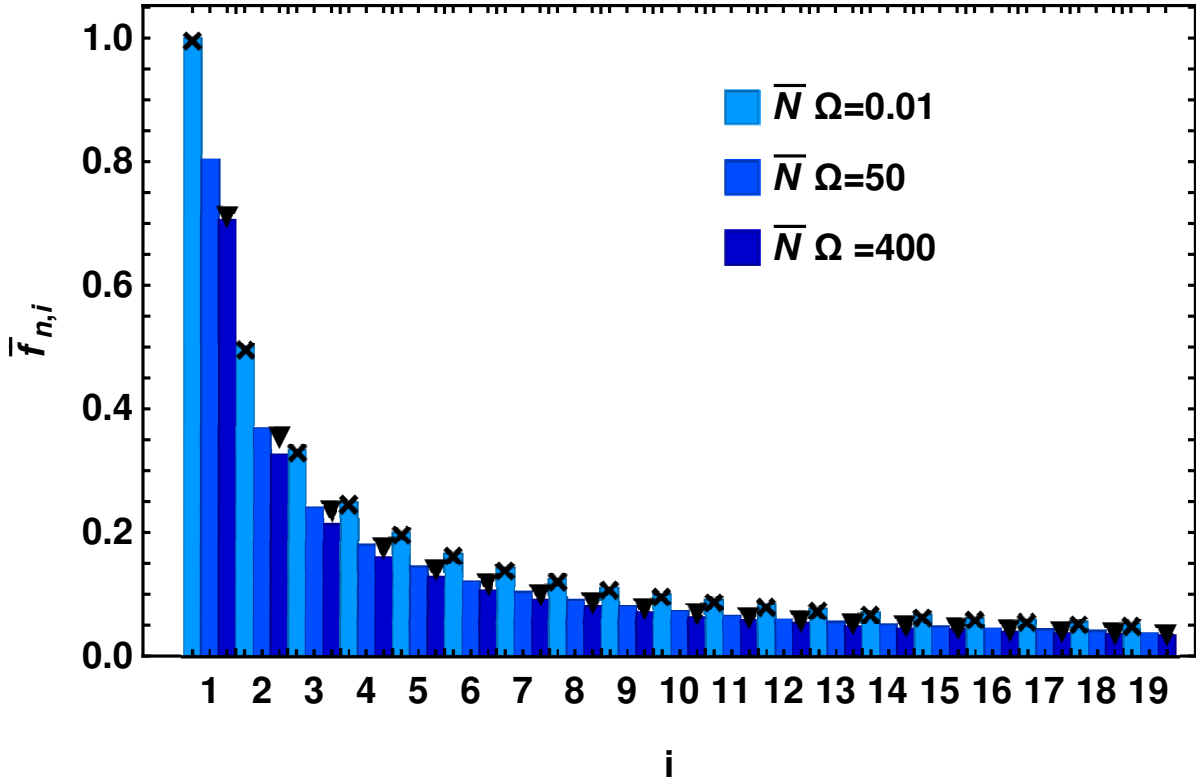


Fig. 3.3 Changing population size and no selection: The time-averaged sample site frequency spectrum when population size changes with time for sample size $n = 20$. The bars are obtained by numerically solving (3.5) and (3.9) for different cycling frequencies while the crosses and triangles show (B.3.17) and (B.3.24) for small and large cycling frequencies, respectively.

Similarly, the change in the time-averaged mean heterozygosity $\delta\bar{H} = \bar{H} - H^*$ due to slowly changing environment is given by

$$\delta\bar{H} = \theta\bar{N}\bar{s}h \left(\frac{\nu^2}{2} + \frac{\nu\sigma}{\bar{s}} \right). \quad (3.20)$$

The factor in the bracket on the RHS of the above equation captures the joint effect of the change in the selection and population size, and shows that (to leading order in $\bar{N}\bar{s}$), changing selection but constant population size results in the same heterozygosity as in the static environment and therefore the mean heterozygosity is affected mainly due to

demography; however, as expected, the change $\delta\bar{H}$ is small for weak selection, as shown in Fig. 3.4a and Fig. 3.5a.

The correction to the adiabatic approximation is obtained in Appendix B.4 when selection is weak and constant but population size varies, and we find that the time-averaged mean heterozygosity is given by (see (B.4.8) and (B.4.11)),

$$\bar{H} \approx \theta + \frac{\theta\bar{N}\bar{s}h}{3} + \frac{\theta\bar{N}\bar{s}h\nu^2}{6} - \frac{\theta\nu^2\bar{N}^2\Omega^2}{2}. \quad (3.21)$$

The first three terms on the RHS of the above equation match with the result from adiabatic theory, and the last term can be ignored for $\bar{N}\Omega \ll \sqrt{\bar{N}\bar{s}}$.

Strong, on-average nonzero selection

On replacing the selection coefficient and population size by the corresponding time-dependent quantities in (3.14a) and (3.14b), we obtain the time-dependent SFS for on-average non-neutral mutants under strong selection to be

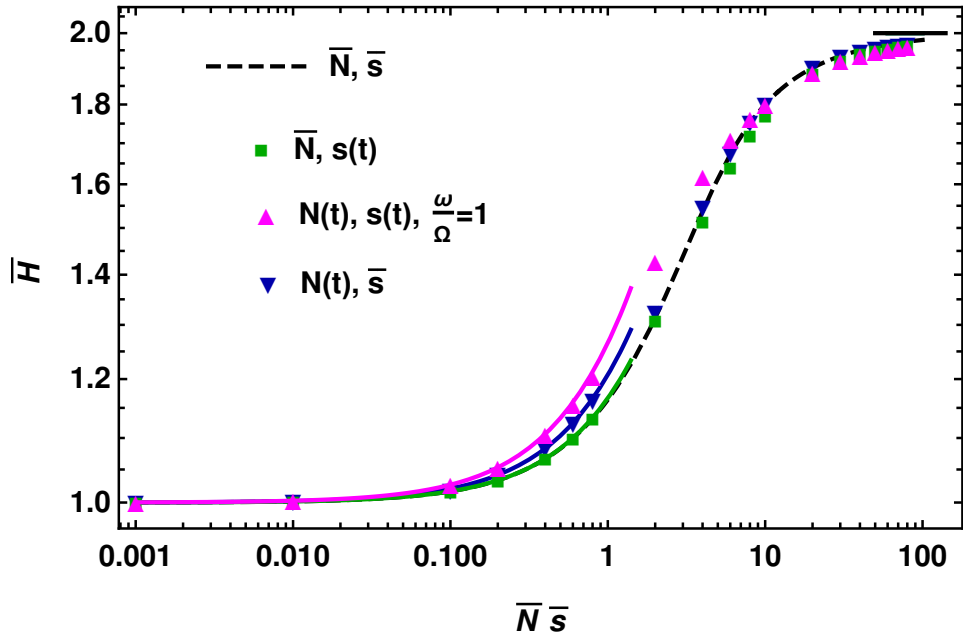
$$f(x, t) \approx \begin{cases} \frac{\theta\rho(t)}{x(1-x)} \frac{h}{h+x(1-2h)} & , s(t) > 0 \\ \frac{\theta\rho(t)}{x(1-x)} e^{-N(t)|s(t)|(1-\ell(x))} & , s(t) < 0 \end{cases} . \quad (3.22a) \quad (3.22b)$$

If both selection and population size change with the same frequency and selection is strong enough so that the exponentially small contribution in selection strength from the negative cycle of selection can be ignored, then the time-averaged SFS is given by

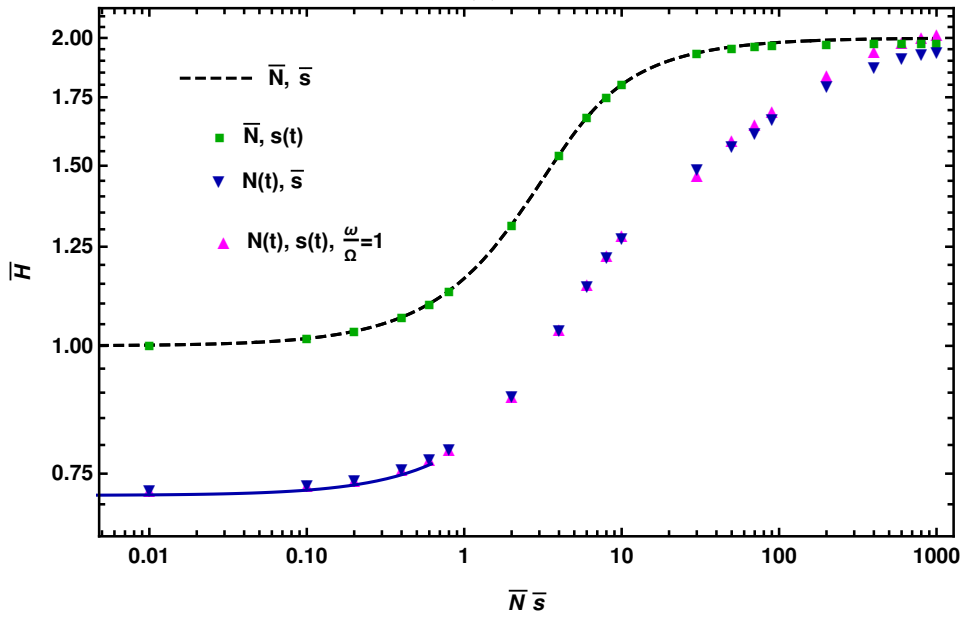
$$\bar{f}(x) \approx \frac{2\hat{N}_e\mu}{x(1-x)} \frac{h}{h+x(1-2h)} \quad (3.23)$$

where

$$\hat{N}_e = \bar{N} \int_0^{2\pi/\omega} dt' \frac{\rho(t')\Theta(s(t'))}{2\pi/\omega} \quad (3.24)$$



(a)



(b)

Fig. 3.4 Changing environment and positive selection at all times: The time-averaged heterozygosity, \bar{H} as a function of selection strength when either the selection coefficient (indicated in the legend by $\bar{N}, s(t)$) or population size ($N(t), \bar{s}$) or both ($N(t), s(t)$) vary with time. The top panel shows the results in slowly changing environments ($\omega, \Omega \ll \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 0.01$, and the bottom panel in rapidly changing environments ($\omega, \Omega \gg \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 1400$. The points are obtained by numerically solving (3.5) and (3.10) for different scenarios as indicated in the legend for $\nu = 0.7$, $\sigma = \bar{s}/8$, $\theta = 1$, and $h = 1/2$. The broken line shows the results in the static environment that are obtained using (3.12). The solid lines show the analytical approximation (3.21) and (3.25) for weak and strong selection, respectively, in slowly changing environment, and (3.35) for weak selection in rapidly changing environment.

and $\Theta(x)$ is the Heaviside step function. The effective population size \hat{N}_e defined above states that the contribution to $\bar{f}(x)$ comes from the average population size during the time selection is positive.

This shows that when both selection and population size are changing but selection remains positive throughout the cycle, $\bar{f}(x)$ is well approximated by the static result given by (3.14a). Figure 3.4a shows the time-averaged heterozygosity when the selection is positive at all times, which, in the light of the above discussion, is given by

$$\bar{H} = H_p^* = \frac{2h\theta}{1-2h} \ln\left(\frac{1-h}{h}\right). \quad (3.25)$$

However, for $\bar{s} > 0$, if $s(t)$ is negative for a part of the cycle, (3.23) shows that $\bar{f}(x)$ is proportional to $\hat{\theta}_e \equiv 2\hat{N}_e\mu$ which is smaller than θ . Thus, in a slowly changing environment, for strong selection, the time-averaged heterozygosity is smaller than H_p^* if the mutant becomes deleterious during a part of the cycle. Similarly, for $\bar{s} < 0$, $\bar{H} > H_n^*$ due to a nonzero contribution from the positive part of the selection cycle.

Strong, on-average zero selection

One can obtain explicit expressions for the time-averaged SFS when the mutant spends equal time in both the positive and negative part of the selection cycle (that is, $\bar{s} = 0$). For strong selection ($\bar{N}\sigma \gg 1$) and equal cycling frequencies ($\omega = \Omega$), from (3.23), we get

$$\bar{f}_{\pm}(x) = \frac{\theta}{x(1-x)} \frac{h}{h+x(1-2h)} \left(\frac{1}{2} \pm \frac{\nu}{\pi}\right) \quad (3.26)$$

for $\rho_{\pm}(t) = 1 \pm \nu \sin(\Omega t)$. Thus we find that unlike in the static, neutral environment where the SFS decays with allele frequency, the time-averaged SFS in an on-average neutral environment is a non-monotonic function of the allele frequency as, up to a scale factor, it is given by the equilibrium SFS for positively selected mutant (see (3.14a)). The

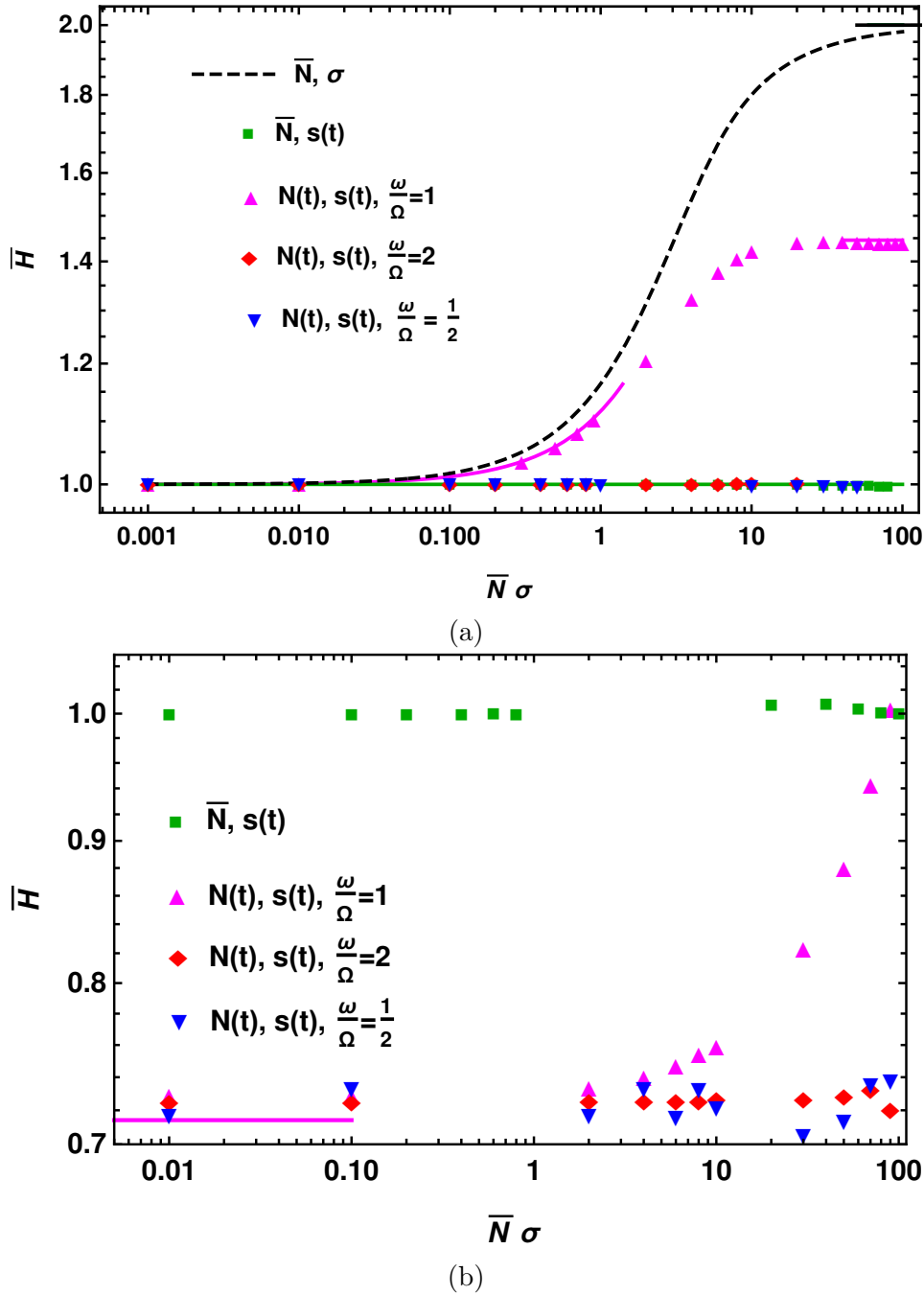


Fig. 3.5 Changing environment and on-average neutral selection: The time-averaged mean heterozygosity, \bar{H} as a function of selection strength when either only the selection coefficient (indicated in the legend by $\bar{N}, s(t)$) or both selection coefficient and population size ($N(t), s(t)$) vary with time. The top panel shows the results in slowly changing environments ($\omega, \Omega \ll \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 0.01$, and the bottom panel in rapidly changing environments ($\omega, \Omega \gg \bar{N}^{-1}, \bar{s}$) for $\bar{N}\omega = 1400$. The points are obtained by numerically solving (3.5) and (3.10) for different scenarios as indicated in the legend for $\nu = 0.7$, $\theta = 1$, $h = 1/2$ and no phase difference between selection and population size variation. The broken line shows the results in the static environment that are obtained using (3.12). The solid lines show the analytical approximation (3.21) for weak selection, and (3.28) and (3.29) for strong selection in slowly changing environment.

factor in the parentheses on the RHS of (3.26) shows that when the population size varies in-phase with selection, as the population size is larger than the average \bar{N} during the positive cycle and stronger positive selection leads to larger heterozygosity, demography ameliorates the effect of deleterious mutations. On the other hand, when selection and population size change out-of-phase, the mean heterozygosity is smaller than in the situation where only selective environment is varied. Quantitatively, from (3.26), we find that the relative mean heterozygosity $0.5 \leq \bar{H}_+/H_p^* < 0.81$ and $0.18 < \bar{H}_-/H_p^* \leq 0.5$ for $0 \leq \nu < 1$.

When the population size is constant ($\nu = 0$), selection is strong ($\bar{N}\sigma \gg 1$) and $h = 1/2$, from (3.26), we find that the time-averaged mean heterozygosity is the same as in the constant, neutral environment. In fact, as shown in Appendix B.5, this result holds exactly for any cycling frequency and selection strength provided the population size remains constant:

$$\bar{H}(h = 1/2) = \theta, \text{ for } \nu = 0, \text{ arbitrary } \bar{N}\omega, \bar{N}\sigma. \quad (3.27)$$

However, for $h \neq 1/2$, from (3.16b) and (3.26), we find that $\bar{H} = H_p^*/2$ which is smaller than θ for $h < 1/2$ and larger for $h > 1/2$.

We now consider the general situation where the population size and selection coefficient vary with different cycling frequencies and have a phase difference. For simplicity, we assume that the ratio of the two frequencies is an integer and consider the average over the smaller frequency (that is, larger time period) as defined in (3.11). If selection changes slower than the population size ($\Omega = K\omega, K = 1, 2, \dots$), and $s(t) = \sigma \sin(\omega t), \rho(t) = 1 + \nu \sin(\Omega t + \Phi), 0 \leq \Phi < 2\pi$, averaging (3.22a) over the selection cycle

and ignoring the contribution from (3.22b) for strong selection, we obtain

$$\frac{\bar{f}(x)}{f_p^*(x)} = \begin{cases} \frac{1}{2} + \frac{\nu \cos \Phi}{\pi K} & , K=\text{odd} \end{cases} \quad (3.28a)$$

$$\frac{\bar{f}(x)}{f_p^*(x)} = \begin{cases} \frac{1}{2} & , K=\text{even} \end{cases} . \quad (3.28b)$$

To understand the K -dependence, we consider the case when $\Phi = 0$. When an even number of cycles of population size occur during one cycle of selection, $K/2$ complete cycles of $N(t)$ lie in the beneficial part of the selection cycle. But as the average population size in each of the $K/2$ cycles is \bar{N} , the final result (3.28b) is found to be independent of K . In contrast, for odd K , the first one half of the $(\frac{K+1}{2})$ th cycle of $N(t)$ contributes more than \bar{N} but its duration shrinks with increasing K as captured by the second term on the RHS of (3.28a). Similarly, when selection changes faster than the population size ($\omega = \kappa\Omega$, $\kappa = 1, 2, \dots$), and $s(t) = \sigma \sin(\omega t)$, $\rho(t) = 1 + \nu \sin(\Omega t + \phi)$, $0 \leq \phi < 2\pi$, we obtain

$$\frac{\bar{f}(x)}{f_p^*(x)} = \begin{cases} \frac{1}{2} + \frac{\nu \cos \phi}{\pi} & , \kappa = 1 \end{cases} \quad (3.29a)$$

$$\frac{\bar{f}(x)}{f_p^*(x)} = \begin{cases} \frac{1}{2} & , \kappa \geq 2 \end{cases} . \quad (3.29b)$$

The above analysis thus demonstrates that unequal cycling frequencies and phase difference can affect the time-averaged SFS. Figure 3.5a summarizes the results obtained above when population size and selection change in phase, and shows that for strong selection, the time-averaged mean heterozygosity in a slowly changing environment is smaller than that in the constant environment with population size \bar{N} and selection coefficient $\sigma > 0$. But, due to (3.28b) and (3.29b) and recalling that the equilibrium heterozygosity of a positively selected, co-dominant mutant is equal to 2θ , we find that \bar{H} is equal to the mean heterozygosity in the constant neutral environment for unequal cycling frequencies. However, \bar{H} can be smaller than θ if the population size and selection coefficient variation have a phase difference.

From (3.15b) and (3.15c), for in-phase variation in population size and selection coefficient with equal cycling frequency, we find the time-averaged sample SFS to be

$$\bar{f}_{n,i} \approx \frac{\omega}{2\pi} \int_0^{\pi/\omega} \frac{\theta \rho(t) n_2 F_1(1, i; n; 2 - \frac{1}{h})}{i(n-i)} dt \quad (3.30)$$

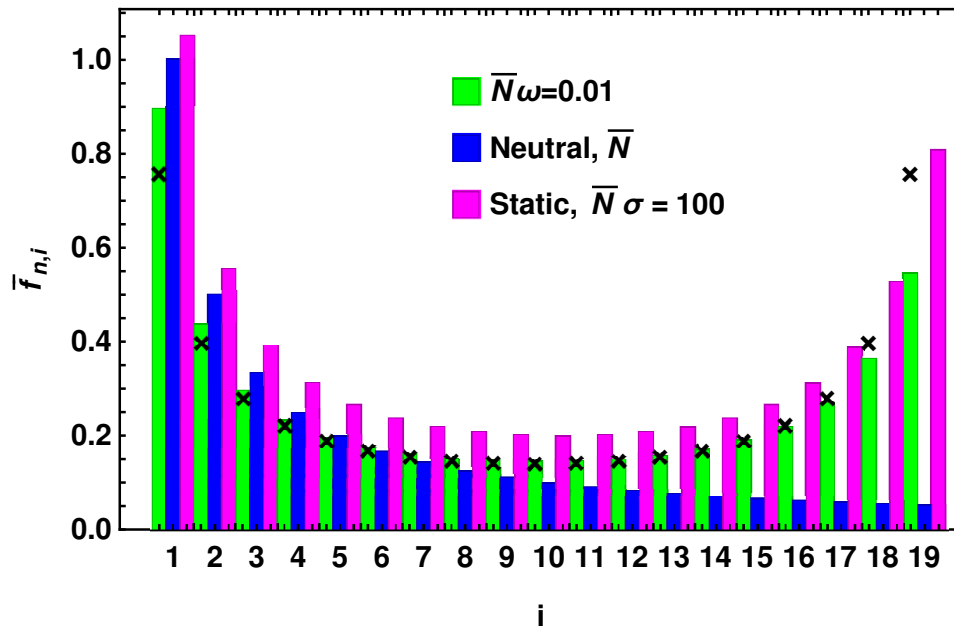
$$= \left(\frac{1}{2} + \frac{\nu}{\pi} \right) \frac{\theta n_2 F_1(1, i; n; 2 - \frac{1}{h})}{i(n-i)}. \quad (3.31)$$

Figure 3.6a shows that in slowly changing environment, the sample SFS obtained by numerically integrating (3.5) and the above analytical approximation are in good agreement except at $i = 1, n - 1$. Thus the sample SFS in on-average neutral environment is not a monotonically decreasing function unlike for equilibrium neutral SFS. In Fig. 3.6a, the nonequilibrium SFS is also compared with the equilibrium SFS for mutant under strong, positive selection with selection strength $\bar{N}\sigma$, and we find that the sample SFS in changing environment is consistently smaller than the equilibrium SFS.

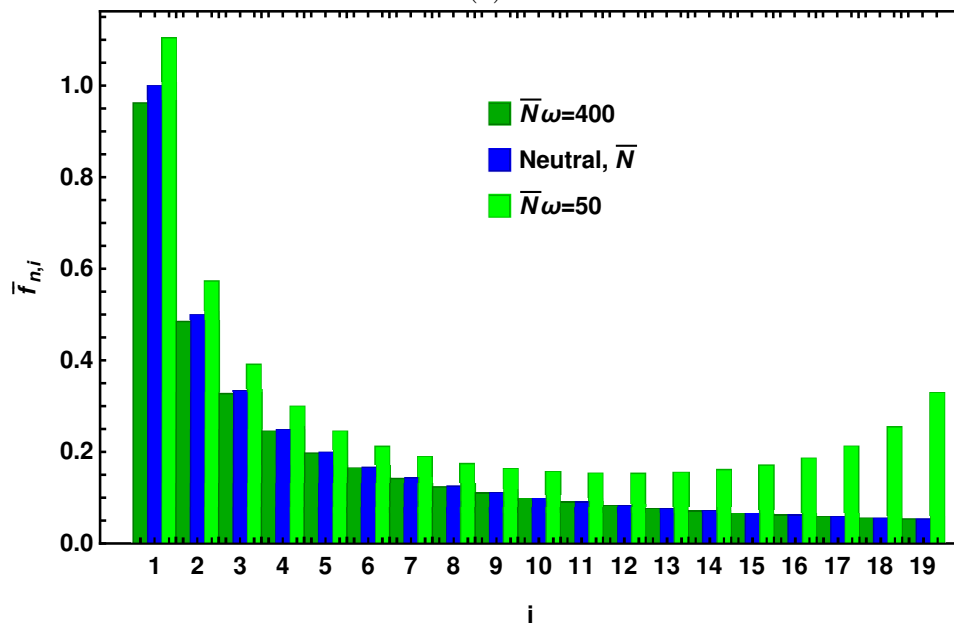
3.4.3 Rapidly changing environment

We now consider the parameter regime where the rate of change of population size and selection is much larger than the inverse population size and average selection coefficient ($\omega, \Omega \gg \bar{N}^{-1}, \bar{s}$). Figure 3.2b shows the periodic variation of the time-dependent SFS with changing environment for large frequencies. We note that, unlike in slowly changing environment, there is a phase difference between the SFS and environment: the SFS increases while the selection is positive and decreases in the deleterious part of the cycle.

In rapidly changing environments, a general expectation is that the population is sensitive only to the time-averaged environment [36]. If the population size remains constant in time but selection varies rapidly, the boundary condition (3.7) for $g(0, t)$ is time-independent, and one can use a standard argument to find $\bar{g}(x)$ [37]. For this



(a)



(b)

Fig. 3.6 Changing environment and on-average neutral selection: The time-averaged sample site frequency spectrum represented by bar chart and obtained by numerically solving (3.5) and (3.9). The top panel shows the results in slowly changing environment for $\bar{N}\omega = 0.01$ that are compared with equilibrium SFS for neutral and positively selected mutant in constant environment. The bottom panel shows the sample SFS in rapidly changing environments for two cycling frequencies which is compared with the equilibrium neutral SFS. The points in the top panel are obtained using (3.31). The other parameters are $\nu = 0.7$, $\omega = \Omega$, $\theta = 1$, $h = 1/2$, $\bar{N}\sigma = 100$, $n = 20$ and no phase difference between selection and population size variation.

purpose, we first rewrite (3.6) as

$$\begin{aligned} \frac{\partial g(x, \psi)}{\partial \psi} &= -\frac{s(\psi)}{\omega} x(1-x) \frac{\partial}{\partial x} [(x + h(1-2x))g(x, \psi)] \\ &+ \frac{x(1-x)}{2\bar{N}\omega} \frac{\partial^2 g(x, \psi)}{\partial x^2} \end{aligned} \quad (3.32)$$

where $\psi = \omega t$, and expand $g(x, \psi)$ as a power series in $(\bar{N}\omega)^{-1}$, that is, write $g(x, \psi) = g_0(x, \psi) + (\bar{N}\omega)^{-1}g_1(x, \psi) + \dots$, and substitute this expansion in the above equation. Keeping terms to leading order in the expansion parameter, we obtain $\frac{\partial g_0(x, \psi)}{\partial \psi} = 0$, which implies that g_0 is a function of x alone and subject to boundary conditions $g_0(0) = \theta, g_0(1) = 0$. As a result, the next order term g_1 obeys the following partial differential equation:

$$\begin{aligned} \frac{\partial g_1(x, \psi)}{\partial \psi} &= -\bar{N}s(\psi)x(1-x) \frac{d}{dx} [(x + h(1-2x))g_0(x)] \\ &+ \frac{x(1-x)}{2} \frac{d^2 g_0(x)}{dx^2}. \end{aligned} \quad (3.33)$$

Integrating both sides of the above equation over the period 2π and using the periodicity property of $g(x, \psi)$, we find that the LHS of the above equation is zero, and $f_0(x) = g_0(x)/(x(1-x))$ is given by (3.12).

Therefore, when the population size is constant, the time-averaged SFS in a rapidly varying environment is simply given by the equilibrium SFS in the time-averaged environment. For this reason, the time-averaged mean heterozygosity shown in Fig. 3.4b and Fig. 3.5b when only selection changes matches with the equilibrium mean heterozygosity in the time-averaged environment for any (nonnegative) selection strength. But when demographic changes also occur, as the boundary condition $g(0, t)$ is time-dependent, it is not clear how to generalize the above argument. As we will discuss below, the time-averaged selective environment is the primary determinant of the behavior of the time-averaged SFS: if the selective environment is neutral or weak on-average, the de-

mography has a strong impact, but if selection is on-average strongly positive, changing population size does not influence the genetic diversity.

No selection

Previous numerical studies [12, 13] with demography and no selection ($\bar{s} = \sigma = 0$) suggest that for large $\bar{N}\Omega$, the time-averaged mean heterozygosity is equal to $2N_e\mu$ where N_e is the effective population size given by the harmonic mean over the period $2\pi/\Omega$,

$$N_e = \bar{N} \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{d\Psi}{\rho(\Psi)} \right]^{-1} = \bar{N} \sqrt{1 - \nu^2} \quad (3.34)$$

which decreases with increasing amplitude of the population size variation. In Appendix B.3, we show that at large cycling frequencies, indeed the effect of demography is to replace \bar{N} by N_e in the static environment results. Thus the time-averaged sample SFS is given by $\bar{f}_{n,i} = 2N_e\mu/i$ (see (B.3.24)) from which it immediately follows that the time-averaged mean heterozygosity, $\bar{H} = 2N_e\mu \equiv \theta_e$. Figure 3.3 shows that the sample SFS is well approximated by θ_e/i when the population size changes rapidly.

Weak selection

In Appendix B.4, we develop a perturbation theory to understand the SFS in rapidly changing environments. In particular, for weak constant selection ($\bar{N}|\bar{s}| \ll 1, \sigma = 0$), from (B.4.8) and (B.4.12), we find that the time-averaged mean heterozygosity is given by

$$\bar{H} = \theta_e + \frac{\theta_e h}{3} N_e \bar{s} \quad (3.35)$$

which is simply the equilibrium result (3.16a) for a weakly selected mutant with population size N_e and matches with the numerical results shown in Fig. 3.4b. We have verified analytically that for a weakly selected, on-average neutral mutant also, the effect of

changing selection and demography is to replace \bar{N} by N_e in the constant neutral environment results, and which yields $\bar{H} = \theta_e$ (refer to Fig. 3.5b).

Thus the above discussion suggests that when the environment changes rapidly and time-averaged selection is neutral or weak (while the population size may or may not vary), the time-averaged sample SFS is given by the expression (3.15a) in a constant environment with selection coefficient \bar{s} and population size N_e ,

$$\bar{f}_{n,i} \approx \frac{\theta_e}{i} + \frac{\theta_e[h(5n - 4i + 6) + 2i - n]N_e\bar{s}}{3(n+1)(n+2)}. \quad (3.36)$$

Strong, on-average positive selection

For strong and positive selection ($\bar{N}\bar{s} \gg 1$), from (3.6), we have

$$\begin{aligned} \frac{1}{\bar{s}} \frac{\partial g(x,t)}{\partial t} &= - \left(1 + \frac{\sigma}{\bar{s}} \sin(\omega t) \right) x(1-x) \frac{\partial}{\partial x} [(x + h(1-2x))g(x,t)] \\ &+ \frac{x(1-x)}{2\bar{N}\bar{s}\rho(t)} \frac{\partial^2 g(x,t)}{\partial x^2} \end{aligned} \quad (3.37)$$

$$\approx -x(1-x) \frac{\partial}{\partial x} [(x + h(1-2x))g(x,t)] \quad (3.38)$$

where the last equation is obtained for large $\bar{N}\bar{s}$ and $\sigma \ll \bar{s}$. Then averaging both sides of (3.38) over the slower cycle (as defined in (3.11) and for integer ratio of the two cycling frequencies), and using the boundary condition $\bar{g}(0) = \theta$, we obtain

$$\bar{g}(x) \approx \frac{\theta h}{x + h(1-2x)} \quad (3.39)$$

which gives the time-averaged SFS to be the same as for the positively selected mutant in constant environment with population size \bar{N} (see (3.14a)). Note that the above argument is valid for arbitrary scaled cycling frequency. Furthermore, due to (3.15b),

the time-averaged sample SFS is given by

$$\bar{f}_{n,i} \approx \frac{n\theta}{i(n-i)} {}_2F_1 \left(1, i; n; 2 - \frac{1}{h} \right) . \quad (3.40)$$

The above discussion shows that for a co-dominant mutant under strong selection, the time-averaged mean heterozygosity is equal to 2θ in both slowly and rapidly changing environments in agreement with the data shown in Figs. 3.4a and 3.4b. Thus, when both selection and population size vary rapidly, while the results for weak selection are given by the equilibrium ones on replacing \bar{N} by N_e , the equilibrium SFS for population size \bar{N} remains valid when selection is strong. This is because in the above deterministic analysis for large $\bar{N}\bar{s}$, the genetic drift term on the RHS of (3.37) whose coefficient depends on the inverse population size is neglected, and the population size enters the expression for SFS through the scaled mutation rate only.

Strong, on-average zero selection

When the selective environment is on-average neutral, the time-averaged mean heterozygosity displayed in Fig. 3.5b for large cycling frequencies is equal to θ for constant population size and changing selection as discussed at the beginning of Section 3.4.3. When both population size and selection vary with equal frequencies, \bar{H} again approaches θ ; however, for unequal frequencies, the results obtained by numerically integrating (3.5) suggest that \bar{H} is independent of the selection strength and given by θ_e . We do not have an analytical understanding of these results; see, however, Appendix B.6. Thus, barring the case when the cycling frequencies are equal, the time-averaged heterozygosity is given by that in the time-averaged selective environment for both weak and strong selection strength and hence the results obtained in Sec. 3.4.3 apply.

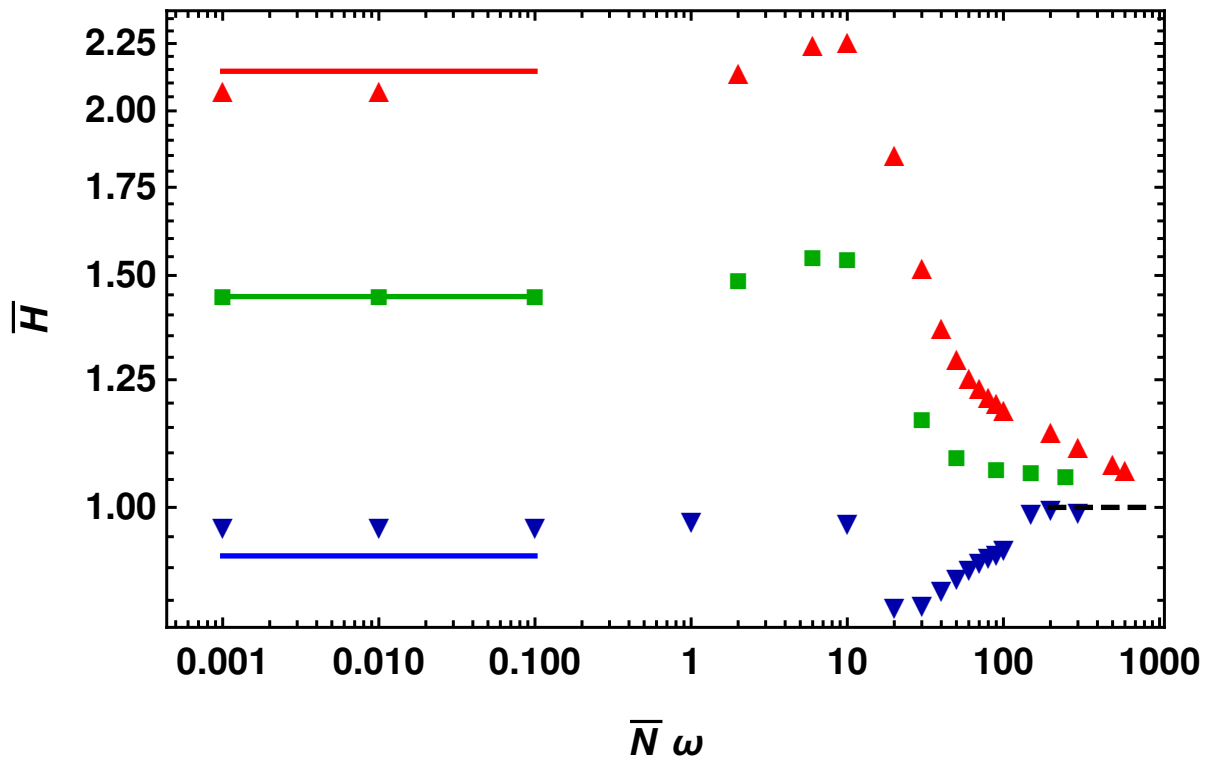


Fig. 3.7 Changing environment and on-average neutral selection: The time-averaged mean heterozygosity as a function of scaled environmental frequency for the dominance parameter $h = 0.3$ (\blacktriangledown), 0.5 (\blacksquare) and 0.7 (\blacktriangle) when both selection coefficient and population size change in time with equal frequency. The points are obtained by numerically solving (3.5) and (3.10), and the solid lines on the left show the analytical approximation (3.31) for slowly changing environment while the dashed line on the right is the equilibrium mean heterozygosity θ in the time-averaged environment. The other parameters are $\nu = 0.7$, $\theta = 1$, and $\bar{N}\sigma = 100$. For comparison, note that the mean heterozygosity in neutral constant environment is equal to $\theta = 1$.

3.5 Discussion

In this chapter, we are interested in understanding how a changing environment, be it due to demography and/or time-dependent selection, impacts genetic diversity within a population. The importance of demography in understanding various measures of genetic diversity has long been appreciated [12–14, 17, 18], but the effect of fluctuating selection on genetic variation has been relatively less studied [19, 20]. Here we have considered the

effect of periodic changes in selection coefficient and population size on the site frequency spectrum and mean heterozygosity.

When selection is weak, as expected, the time-averaged mean heterozygosity is essentially the same as for a neutral population with changing population size (see Figs. 3.4 and 3.5). We find that the effect of demography is mainly seen when the population size changes rapidly; in this case, the time-averaged statistical quantities behave the same way as in the constant environment, but with reduced population size N_e given by the harmonic mean of the changing population size over a period (see (3.34)). Although the effective population size has been derived when the population size has random fluctuations, and verified numerically for other models of demographic change [12–14], here we have analytically derived the effective population size for periodically changing population size in Appendices B.3 and B.4. Our explicit expressions for the time-averaged SFS are given by (B.3.17) and (B.3.24) for slowly and rapidly changing environments, respectively, and shown graphically in Fig. 3.3 when selection is absent.

For strong selection, if time-averaged selection is positive, a deterministic argument, given in Sec. 3.4.3, shows that the time-averaged SFS is the same as for a positively selected mutant in a constant environment for any cycling frequency. But the results depend on the cycling frequency when the selection is zero on-average. For large cycling frequencies, the time-averaged mean heterozygosity is found, in general, to behave as in constant neutral environment with effective population size N_e discussed above; however, in a slowly changing environment, the time-averaged SFS is the same as in the static environment with an effective population size \hat{N}_e given by the average population size over the time the selection is positive, see (3.24). For neutral populations, the census population size in changing environment can be replaced by an appropriately defined effective population size when the mutation rate is small relative to the frequency of change in the population size [12] and stochastic changes in the population size are

independent random variables [36, 38]. Here, in addition to the demographic N_e in fast changing environment discussed above, we find that the census population size in slowly changing, selective environment can be replaced by an effective population size, \hat{N}_e (refer to (3.24)).

Furthermore, in a slowly changing, on-average neutral environment, we find that the time-averaged SFS is a U-shaped function as one would expect for beneficial mutants in a constant environment. However, as shown in Fig. 3.6, relative to the SFS in constant environment, the time-averaged SFS is reduced by a factor equal to the fraction of the time the selection is negative in a seasonal cycle. Figure 3.6 also shows that there are more high-frequency variants in the changing environment as compared to in the constant, neutral environment; similar behavior has also been seen when there is fluctuating selection (but no demography) [19, 20]. Taken together, these studies suggest that varying selection, even if zero on-average, can produce an SFS similar to that in a constant, positively selective environment. Fig. 3.7 further shows that the time-averaged heterozygosity follows the same qualitative trend with dominance coefficient as for a positively selected mutant in a constant environment (see Fig. 3.1c).

Figure 3.5 shows that the time-averaged heterozygosity for on-average neutral selection and various environmental scenarios is smaller than for positively selected mutants in a constant environment. This is simply because a newborn mutant is more likely to be lost when selection is negative than when it is positive, and therefore the contribution to polymorphism comes only from the part of the cycle when the selection is strongly positive. In a recent study, we have also shown that while the neutral heterozygosity at a linked site due to beneficial and deleterious sweeps is identical in a constant selective environment, varying selection coefficient breaks the symmetry; in particular, it is found that the neutral heterozygosity is strongly affected by the deleterious sweeps even when selection is changing slowly while the footprint of beneficial sweeps is not much impacted

[25]. These results suggest that in addition to demography, varying selective environment (about zero mean) is a potentially important factor for explaining the lower levels of neutral genetic diversity than predicted on the basis of constant neutral population assumption (for a recent overview, see BUFFALO [39]).

Throughout this chapter, we have assumed completely unlinked sites; a generalization of the infinite sites model studied here to include finite recombination, and detailed investigations of the joint effect of changing selective environment and demography in more complex scenarios, such as when genetic hitchhiking occurs [40], remains a task for the future.

Copyright and Permission

The results of this chapter have been published in:

Jain, K., and S. Kaushik, 2022 Joint effect of changing selection and demography on the site frequency spectrum, *Theor Popul Biol.* **146**, 46–60.

©Elsevier B.V. Reproduced with permission. All rights reserved.

We have reproduced the materials here with kind permission from the Elsevier. See the publication link below:

<https://www.sciencedirect.com/science/article/pii/S0040580922000430?via>

Bibliography

- [1] ELLEGREN, H. and N. GALTIER, 2016 Determinants of genetic diversity. *Nature Reviews Genetics* **17**: 422 – 433.
- [2] MOUTINHO, A. F., T. BATAILLON, and J. Y. DUTHEIL, 2020 Variation of the adaptive substitution rate between species and within genomes. *Evolutionary Ecology* **34**: 315–338.
- [3] WRIGHT, S., 1938 The distribution of gene frequencies under irreversible mutations. *Proc. Natl. Acad. Sci USA* **24**: 253–259.
- [4] KIMURA, M., 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**: 247–269.
- [5] BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- [6] FAY, J. C. and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- [7] DESAI, M. M. and J. B. PLOTKIN, 2008 The polymorphism frequency spectrum of finitely many sites under selection. *Genetics* **180**: 2175–2191.

-
- [8] CHARLESWORTH, B. and K. JAIN, 2014 Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. *Genetics* **198**: 1587–1602.
- [9] BIRKNER, M., J. BLATH, and B. ELDON, 2013 Statistical properties of the site-frequency spectrum associated with Lambda-coalescents. *Genetics* **195**: 1037–1053.
- [10] KOOPMANN, B., J. MÜLLER, A. TELLIER, and D. ŽIVKOVIĆ, 2017 Fisher-Wright model with deterministic seed bank and selection. *Theoretical Population Biology* **114**: 29–39.
- [11] SCHRAIBER, J. G. and J. M. AKEY, 2015 Methods and models for unravelling human evolutionary history. *Nat Rev Genet* **16**: 727–740.
- [12] NEI, M., T. MARUYAMA, and R. CHAKRABORTY, 1975 The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.
- [13] MARUYAMA, T. and P. A. FUERST, 1985 Population bottlenecks and nonequilibrium models in population genetics. III. Genic homozygosity in populations which experience periodic bottlenecks. *Genetics* **111**: 691–703.
- [14] TAJIMA, F., 1989 The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- [15] BUSTAMANTE, C. D., J. WAKELEY, S. SAWYER, and D. L. HARTL, 2001 Directional selection and the site-frequency spectrum. *Genetics* **159**: 1779–1788.
- [16] WILLIAMSON, S. H., A. FLEDEL-ALON, and C. D. BUSTAMANTE, 2004 Population genetics of polymorphism and divergence for diploid selection models with arbitrary dominance. *Genetics* **168**: 463–475.
- [17] WILLIAMSON, S. H., R. HERNANDEZ, A. FLEDEL-ALON, L. ZHU, R. NIELSEN, and C. D. BUSTAMANTE, 2005 Simultaneous inference of selection and population

- growth from patterns of variation in the human genome. *Proc. Nat Acad. Sci. USA* **102**: 7882–7887.
- [18] ŽIVKOVIĆ, D., M. STEINRÜCKEN, Y. S. SONG, and W. STEPHAN, 2015 Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. *Genetics* **200**: 601–617.
- [19] HUERTA-SANCHEZ, E., R. DURRETT, and C. D. BUSTAMANTE, 2008 Population genetics of polymorphism and divergence under fluctuating selection. *Genetics* **178**: 325–337.
- [20] GOSSMANN, T. I., D. WAXMAN, and A. EYRE-WALKER, 2014 Fluctuating selection models and McDonald-Kreitman type analyses. *Plos ONE* -: e84550.
- [21] BAUDRY, E., B. VIGINIER, and M. VEUILLE, 2004 Non-African populations of *Drosophila melanogaster* have a unique origin. *Mol Biol Evol* **21**: 1482–1491.
- [22] MURRAY, J. D., 2002 *Mathematical Biology*. Springer-Verlag Berlin Heidelberg.
- [23] KIMURA, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**: 893–903.
- [24] DEVI, A. and K. JAIN, 2020 The impact of dominance on adaptation in changing environments. *Genetics* **216**: 227–240.
- [25] KAUSHIK, S. and K. JAIN, 2021 Time to fixation in changing environments. *Genetics* **219**: iyab 148.
- [26] EVANS, S. N., Y. SHVETS, and M. SLATKIN, 2007 Non-equilibrium theory of the allele frequency spectrum. *Theor Popul Biol.* **71**: 109–119.
- [27] GRAVEL, S., 2016 When is selection effective? *Genetics* **203**: 451–462.

-
- [28] KIMURA, M., 1957 Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**: 882–901.
- [29] JUNG, P., 1993 Periodically driven stochastic systems. *Phys. Rep.* **234**: 175–295.
- [30] GRIFFITHS, R., 2003 The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theo Pop Biol* **64**: 241–251.
- [31] WRIGHT, S., 1942 Statistical genetics and evolution. *Bull. Amer. Math. Soc.* **48**: 223–246.
- [32] EWENS, W., 2004 *Mathematical Population Genetics*. Springer, Berlin.
- [33] ABRAMOWITZ, M. and I. A. STEGUN, 1964 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover.
- [34] CORLESS, R., G. GONNET, D. HARE, D. JEFFREY, and D. KNUTH, 1996 On the Lambert W function. *Advances in Computational Mathematics* **5**: 329–359.
- [35] HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection. *V. Proc. Camb. Philos. Soc.* **23**: 838–844.
- [36] SJÖDIN, P., I. KAJ, S. KRONE, M. LASCoux, and M. NORDBORG, 2005 On the meaning and existence of effective population size. *Genetics* **169**: 1061–1070.
- [37] REIMANN, P., R. BARTUSSEK, R. HÄUSSLER, and P. HÄNGGI, 1996 Brownian motors driven by temperature oscillations. *Phys Lett A* **215**: 26–31.
- [38] IIZUKA, M., 2010 Effective population size of a population with stochastically varying size. *Journal of Mathematical Biology* **61**: 359–375.
- [39] BUFFALO, V., 2021 Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin’s Paradox. *eLife* **10**: e67509.

- [40] BARTON, N. H., 2000 Genetic hitchhiking. *Philosophical Transactions Royal Society B* **355**: 1553–1562.

Chapter 4

Effect of beneficial sweeps and background selection on genetic diversity in changing environments

In this Chapter, we relax the assumption of randomly mating population and discuss the reduction in the linked neutral genetic diversity due to fixation of mutant at the selected site when inbreeding occurs within population.

4.1 Introduction

There has been much interest in understanding how the neutral genetic diversity patterns are shaped for a long time. It has been observed that pairwise diversity in various natural species is several orders of magnitude lower than predicted by the neutral theory [1], and the neutral genetic diversity for various species varies in a narrow range for several orders of magnitude change in the population sizes—this is known as Lewontin’s paradox [2]. The indirect effects of selection on the linked neutral variation are one possible contributor to resolving Lewontin’s paradox [3, 4].

When a beneficial mutation spreads throughout an entire population, the linked neutral variants can hitchhike along with it; this is referred to as selective sweep [5–8]. Other variants not linked to the selected mutation are subsequently lost from the population, reducing genetic diversity at the linked neutral sites. The genetic diversity can also be reduced due to the elimination of the recurrent deleterious mutations from the population, known as background selection [9] because the deleterious mutations are selected against and subsequently eliminated from the population, which also eliminates the linked neutral variants. Several analytical and simulation studies have been conducted to investigate the effect of background selection [9–11] and selective sweeps [12, 8] on the linked neutral genetic diversity considering environment to be selectively constant.

Natural environments, on the other hand, are not static and change over time. It is therefore important to ask how the genetic diversity at the linked neutral sites is affected by selective sweeps and background selection in the changing environment. We primarily want to know how the levels of genetic variation change in the selectively changing environment. As selective sweeps and background selection effects are strong when recombination is low, we work in low recombination regimes throughout the Chapter.

In a recent study [13], the effect of a continually changing environment on the genetic diversity in a randomly mating population was investigated using a two-locus model with one neutral and one selected locus. It was discovered that the Maruyama-Kimura symmetry [14, 15] for conditional mean fixation time in static environment does not hold in a changing environment, and the deleterious sweeps are strongly affected due to varying selection. The beneficial sweeps, on the other hand, are robust for slow or moderate changes in the selective environment [13]. Here, we extend these results for selective sweeps to a more general case of inbreeding population. We use diffusion theory to calculate the conditional mean fixation time for the inbreeding population. When the inbreeding coefficient is high, the deleterious and beneficial sweeps are weakly affected in

the changing environment, and the changing environment effects are most significant in a randomly mating population.

Because the deleterious mutations are more likely to be lost than fix in population, we investigate the impact of background selection on the shape of the site frequency spectrum (SFS) in changing selective environment. In a finite asexual population, fittest class of individuals is eventually lost from the population [16, 17]. Previous research has shown that in slow Muller's ratchet regime and static environment, the reduction in the genetic diversity caused by strong selection against the recurrent deleterious mutations is equivalent to the neutral population with a reduced effective population size [9, 10] in the absence of recombination. However, the effect of background selection in the population where Muller's ratchet clicks fast is much less studied [11]. In this Chapter, we will investigate the effect of background selection for arbitrary selection strength in slowly changing environment. We give simple analytical expression for change in heterozygosity when Muller's ratchet clicks slowly and study the fast Muller's ratchet regime numerically.

We find that, even in a slowly changing environment, genetic diversity at linked neutral sites differs significantly from that in a static environment when the selected site contains deleterious mutations; in particular when the population fluctuates between zero and a negative selection coefficient, the shape of the time-averaged SFS is found to differ from the shape of the SFS with a time-averaged selection coefficient.

4.2 Selective sweeps in changing environments

The reduction in genetic variation caused by selective sweeps is significant in regions with low recombination and depends on the time it takes for the mutation to fix in the population [5, 18]. In this section, we discuss the conditional mean fixation time, which is calculated by averaging the time taken by the mutation to fix into the population only

in the trajectories destined to fix, when the selective environment is time-dependent and a single locus is under selection.

4.2.1 Diffusion theory

When the environment is selectively constant, the conditional mean fixation time for beneficial mutation with selective advantage $+s$ and dominance coefficient h is identical to that of deleterious mutation with selective disadvantage $-s$ and dominance $1 - h$ [14, 15]. In the static environment, this symmetry also holds for the inbreeding population [19]. It has recently been demonstrated that this symmetry breaks for randomly mating populations in a changing environment [13]. In this section, we consider the case when inbreeding occurs in the population.

Consider a diploid population of constant population size N . The biallelic locus experiences time-dependent selection coefficient, $s(t) = \bar{s} + \sigma \sin(\omega t + \theta_0)$ that varies periodically with time, where \bar{s} is the time-averaged selection coefficient, $0 < \theta_0 < 2\pi$ is the initial phase, $\sigma > 0$ is the amplitude of the oscillation, and ω is the rate of change of environment. The three genotypes, AA , Aa and aa have frequencies $p^2 + fpq$, $2pq(1 - f)$ and $q^2 + fpq$, and fitnesses 1 , $1 + hs(t)$, and $1 + s(t)$ respectively [20, 21] where h is the dominance coefficient ($0 < h < 1$), and f is the inbreeding coefficient ($0 < f < 1$). The p and q are the frequency of allele A and a , respectively.

The backward Fokker-Planck equation for the randomly mating population (Eqn. A4 in KAUSHIK and JAIN [13]) for mean fixation time for a mutation with initial frequency x at time t_0 , can be generalized for inbreeding population, which satisfies

$$-\frac{\partial \bar{T}(x, t_0)}{\partial t_0} - P_f(x, t_0) = s(t_0)g(x)\frac{\partial \bar{T}(x, t_0)}{\partial x} + \frac{x(1-x)}{2N_e}\frac{\partial^2 \bar{T}(x, t_0)}{\partial x^2} \quad (4.1)$$

where $g(x) = (x + h(1 - 2x) + f(1 - h - x(1 - 2h)))$, P_f is the eventual fixation probability, and N_e is the effective population size given by $N/(1+f)$ [23].

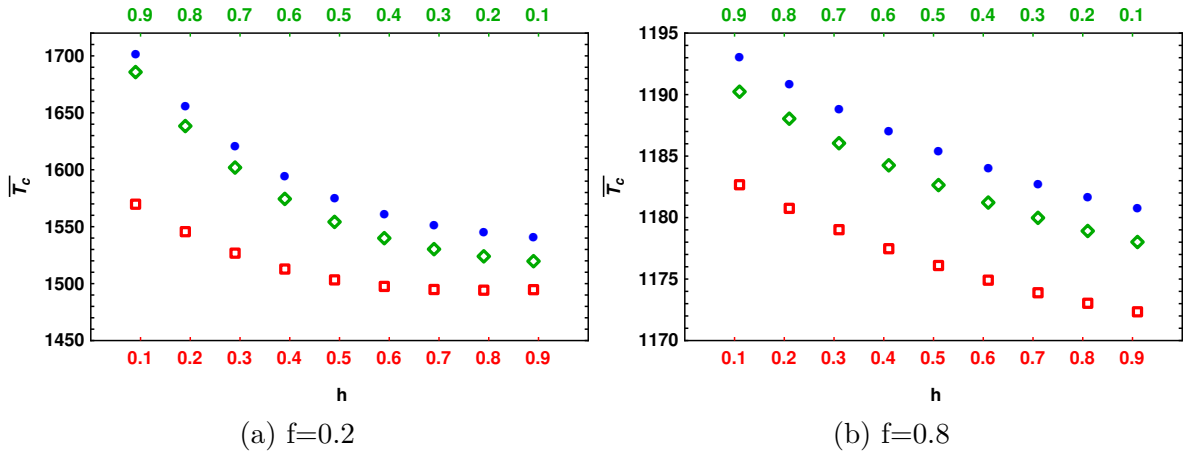


Fig. 4.1 The variation of conditional mean fixation time \bar{T}_c with dominance parameter (h) for moderate selection in static (\bullet) and slowly changing environment where selection coefficient is on-average neutral ($\bar{s} = 0$), and given by $s(t) = \sigma \sin(\omega t + \theta)$ (diamonds) and $-s(t)$ (squares) respectively. The other parameters are $N_e = 2 \times 10^3$, $N_e \omega = 0.05$, $\sigma = 0.01$, $\theta = \pi/4$, (a) $f = 0.2$ and (b) $f = 0.8$. When inbreeding is high, the conditional mean fixation time for both beneficial and deleterious is mildly affected in the slowly changing environment.

The above equation does not appear to be exactly solvable, so we use first-order perturbation theory for the approximate solution. Previous research [13] indicates that the effects of a changing environment are significant when the mutation is on-average neutral, so we focus on-average neutral mutation. In the slowly changing environment, $\omega \ll N_e^{-1} s(0)$, for on-average neutral mutation ($\bar{s} = 0$) the mean fixation time and eventual fixation probability are approximated as $\bar{T} \approx \bar{T}_0 + N_e \omega \bar{T}_1$, and $P_f \approx P_0 + N_e \omega P_1$. Here, \bar{T}_0 and P_0 are the mean fixation time and fixation probability in the static environment; and \bar{T}_1 and P_1 are the leading order deviation in ω in the mean fixation time and fixation probability, respectively. These are calculated in Appendix C.1.

The conditional mean fixation times for the slowly changing environment are shown in Fig. 4.1, where the data is obtained by numerically solving (C.1.3), (C.1.9), and (C.1.10). We find that the Maruyama-Kimura symmetry does not hold for the inbreeding population, and deleterious mutations are more affected than the beneficial ones by changing environment. Thus, in the case of an inbreeding population, the qualitative

behavior of conditional mean fixation time is similar to that of a randomly mating population in a slowly changing environment.

Whereas, unlike the randomly mating population where the recessive deleterious mutations are strongly impacted, deleterious mutations, whether recessive or dominant, are mildly affected due to the slowly changing environment when the inbreeding coefficient is high, as shown in Fig. 4.1b. This can be understood qualitatively using the following argument. In a randomly mating population, the recessive deleterious mutations are strongly affected because they have a higher fixation probability than dominant mutations [22] allowing recessive mutations to have a less strong tendency to fix quickly as compared to the dominant deleterious mutation, and former mutations can continue to segregate in the population for long periods of time, making them more sensitive to the changing environment [13]. However, when the inbreeding coefficient is high, the fixation probability for the deleterious mutations is weakly dependent on the dominance coefficient (see C.1.7, C.1.8). The trajectories for recessive, as well as dominant deleterious mutations, are then exposed to the changing environment for nearly the same amount of time and thus affected in the similar manner. Thus, these results suggest that in the case of a highly inbred population in a slowly changing environment, the deleterious and beneficial sweeps produce similar genetic diversity patterns.

4.2.2 Semi-deterministic theory: strongly beneficial mutations

The solution of diffusion equation does not have a closed form in the presence of selection and thus cannot be used to calculate the fixation time distribution. However, we can use semi-deterministic theory for the beneficial mutations when the population size is large ($2Ns \gg 1$), as recently used by MARTIN and LAMBERT [23], and KAUSHIK and JAIN [13]. In a static environment, the conditional mean fixation time, \bar{T}_c , for the mutation with selection coefficient $+s$ and dominance coefficient h is the same for the mutation

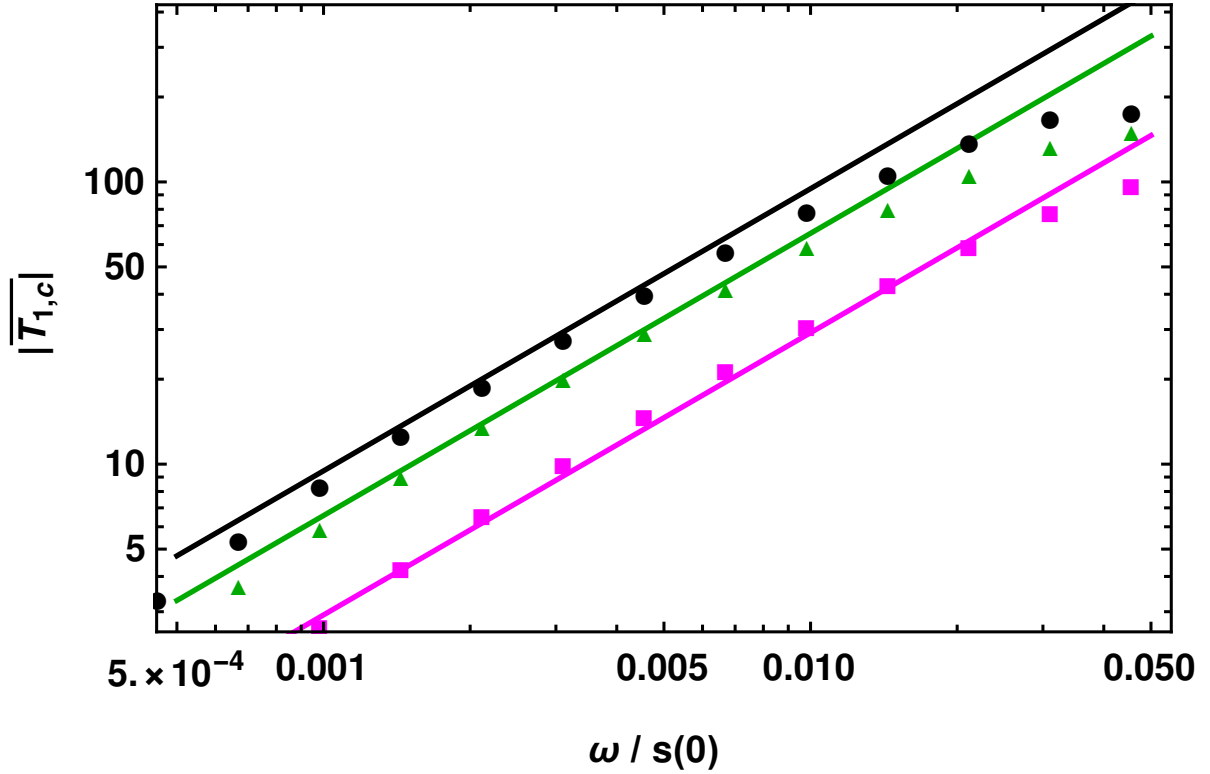


Fig. 4.2 Strongly beneficial mutations: The deviation in the conditional mean fixation time, $\bar{T}_{1,c}$, in the slowly changing environment with selection coefficient $s(t) = (\bar{s} + \sigma \sin(\omega t + \pi/4))$ when the mutation is beneficial at all times for $f = 0$ (black), $f = 0.2$ (green), and $f = 0.8$ (magenta) to show that the deviation is maximum for randomly mating population. The points are obtained by numerically solving (C.2.1). The other parameters are $N_e = 10^5$, $\bar{s} = 0.01$, $\sigma = 0.007$ and $h = 0.5$ respectively.

with selection coefficient $+s$, and dominance $1 - h$ for arbitrary inbreeding coefficient. In other words, $\bar{T}(+s, h, f) \approx \bar{T}(+s, 1 - h, f)$ [24, 19]. This symmetry is preserved in the changing environment for the panmictic population in the strong selection limit, as recently demonstrated by KAUSHIK and JAIN [13]. Here, we will explore the h to $1 - h$ symmetry when inbreeding occurs.

The allele frequency distribution, $\Phi_f(p, t|p_0, 0)$, that the mutation frequency is p at time t , given that the initial mutation frequency is p_0 at time t_0 obeys the following

forward Kolmogorov equation [25],

$$\frac{\partial \Phi_f(p, t|p_0, 0)}{\partial t} = -s(t) \frac{\partial}{\partial p} [g(p) \Phi_f(p, t|p_0, 0)] + \frac{\partial^2}{\partial p^2} \left[\frac{pq \Phi_f(p, t|p_0, 0)}{2N_e} \right] \quad (4.2)$$

where, $g(p) = s_a(1-p) + s_b p$, $s_a = h + (1-h)f$ and $s_b = 1 - (1-f)h$, respectively. For $p \rightarrow 0$, the (4.2) reduces to a Feller diffusion equation [26] which can be solved exactly. Conditional on fixation, $\Phi_f(p, t|p_0, 0)$ converges to a stationary distribution at large times, which gives the eventual fixation for the beneficial mutation having frequency p_0 at time $t = 0$ as

$$P_f(p_0, 0) = 1 - \exp \left[-\frac{2N_e p_0}{\int_0^\infty dt e^{-s_a t} \int_0^t dt' s(t')} \right] \quad (4.3)$$

Slowly changing environment

In a slowly changing environment, the fixation probability for an on-average neutral mutation ($\bar{s} = 0$), when the mutation occurs in the positive cycle of the selection coefficient with initial phase, $\theta_0 = \omega t_0$ ($0 < \theta_0 < \pi$), is obtained by a Taylor series expansion of (4.3) to the linear order in ω . The fixation probability is then given as

$$P_f(p_0) = 2N_e p_0 (s_a \sigma \sin \theta_0 + \omega \cot \theta_0) \quad (4.4)$$

The deviation in the fixation probability is independent of the dominance parameter and weakly dependent on the inbreeding coefficient. As previously discussed, the conditional mean fixation time is closely related to the eventual fixation probability; thus (4.4) suggests that a slowly changing environment is expected to affect conditional mean fixation time mildly.

The conditional mean fixation time is approximated as $\bar{T}_c \approx \bar{T}_{0,c} + \frac{\omega}{s(0)} \bar{T}_{1,c}$ when the environment changes slowly.

For a single mutation, and keeping terms up to the order $\frac{(\ln \alpha)^2}{\alpha}$, the first order correction to the conditional mean fixation time is given by (see Appendix C.2 for details),

$$\frac{\bar{T}_{1,c}}{2N_e} = -\frac{\sigma \cos \theta_0 (\ln \alpha)^2 (1+f)^2}{4s(0)\alpha s_a^2 s_b^2} \quad (4.5)$$

For a randomly mating population, $f = 0$, the above expression reduces to Eqn. (20) in KAUSHIK and JAIN [13]. In slowly changing environment, the leading order term in $\bar{T}_{1,c}$ has $h \rightarrow 1 - h$ symmetry. In other words, like in randomly mating population, $\bar{T}(+s, h, f) \approx \bar{T}(+s, 1 - h, f)$ holds in the slowly changing environment for the inbreeding population also. The deviation in the conditional mean fixation time, $\bar{T}_{1,c}$, due to changing environment is maximum when $f = 0$, as depicted in Fig. 4.2.

Fig. 4.2 shows that the conditional fixation times are affected very mildly in the wide range of environmental change. Thus, the results of the beneficial sweep in the changing environment for a randomly mating population are robust in the inbreeding population.

4.3 Background selection in changing environment

In the previous section, we have shown that the effect of changing environment on deleterious sweeps is strongest for the randomly mating population ($f = 0$). In this section, we focus on haploid asexual populations ($f = 0$) to explore the effect of background selection (BGS) on the linked neutral genetic variation in the changing environment. In particular, we will study the neutral site frequency spectrum (SFS) and neutral heterozygosity.

4.3.1 Multilocus Wright-Fisher model

We use the multilocus Wright-Fisher model to study the effect of recurrent deleterious mutations at the selected site in a haploid asexual population ($f = 0$) of size N . Consider a large genome in which a single selected site is linked to infinite number of neutral sites. At the linked neutral sites, the mutations occur according to an infinite-sites model, meaning that each mutation occurs at a new site that does not have any mutation before. The population evolves according to discrete-generation Wright-Fisher process [27, 28], described in a following manner. At $t = 0$, we start with a monomorphic population, and subsequent generations are obtained by first reproduction and then mutation. In reproduction, the offspring generation is generated by sampling individuals from the parent generation with equal probabilities. The neutral mutations occur in the population with a rate u_n at a monomorphic site. After a drift-neutral mutation equilibrium for the SFS has been reached, the deleterious mutations are introduced, and the chosen individual acquires deleterious mutations according to the Poisson distribution with mean Nu_d . The offspring generation is then generated by picking individuals from the parent's generation with probabilities proportional to their relative fitnesses. The absolute fitness for an individual with k deleterious mutations is given by $(1 - s(t))^k$. Unlike in the previous section, here we study the discrete-time model, and the selection changes periodically in the square waveform given by

$$s(t) = \begin{cases} \bar{s} + \sigma, & nT_p \leq t < \left(n + \frac{1}{2}\right)T_p \\ \bar{s} - \sigma, & \left(n + \frac{1}{2}\right)T_p \leq t < (n + 1)T_p \end{cases} \quad (4.6)$$

where $\sigma (< \bar{s})$ is the amplitude, T_p is the time period of the oscillation and n is a positive integer.

To quantify the effect of background selection on the linked neutral diversity, we mainly focus on the two summary statistics, time-averaged site frequency spectrum (SFS),

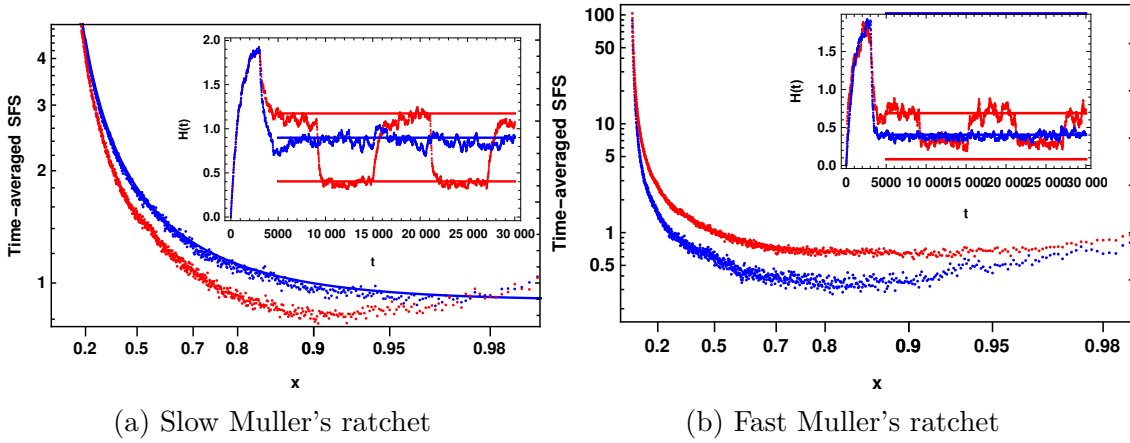


Fig. 4.3 Strictly deleterious mutations: The time-averaged site frequency spectrum for the static environment (blue) and for the slowly changing environment (red) when the selection coefficient is always negative and varies periodically with time period T_p between $\bar{s} + \sigma$ in first half of the cycle and $\bar{s} - \sigma$ in the second half of the cycle. The parameters are $N = 1000$, $Nu_n = 1$, $u_d = 0.08$, $\sigma = \bar{s}/2$, $T_p = 12000$, (a) $\bar{s} = 0.1$ and (b) $\bar{s} = 0.05$. The solid blue line is $(2Nu_n e^{-u_d/\bar{s}})/x$. The inset shows the time-dependent heterozygosity for the slowly changing environment (red) and static environment (blue), the solid lines represent heterozygosity for the neutral population with the corresponding effective population sizes. The population evolves under drift neutral mutations equilibrium for SFS till $t = 3000$, and after that, deleterious mutations are introduced.

and time-averaged heterozygosity, $\bar{H} = \frac{1}{T_p} \int_0^{T_p} H(t) dt$. It has previously been shown [9] that in static environment, for strongly deleterious mutations, the effect of the BGS on SFS is same as the neutral expectation with an effective population size, $N_e = N e^{-u_d/\bar{s}}$, where \bar{s} is the magnitude of the selection strength for the deleterious mutation. But for moderately strong selection, the background selection distorts the shape of the SFS from the neutral expectation of monotonically decaying function to a U-shaped function [29]. Here we investigate these results in changing environment.

4.3.2 Fluctuating environment between deleterious mutations

We consider the case when the selection coefficient changes between two negative selection coefficients so that the mutation always remains deleterious. We find that, in the slowly changing environment, the time-averaged SFS $\bar{f}(x)$ and time-averaged heterozygosity

\bar{H} are significantly different compared to the corresponding quantities in the static environment with the time-averaged selection coefficient as depicted in Fig. 4.3.

Figs. 4.3a, and 4.3b show respectively that when Muller's ratchet clicks slowly and rapidly, the time-averaged SFS is smaller and larger, respectively, in the slowly changing environment than the SFS with the time-averaged selection coefficient (\bar{s}) in the static environment. When the neutral allele frequency is very low, the effect of deleterious mutations is negligible, and the time-averaged SFS in changing environment is the same as the neutral SFS, which is analogous to the effect of background selection on neutral SFS in static environment at low frequency [29]. However, when the frequency of linked neutral mutations is intermediate, the deleterious mutations are exposed to a changing environment for a longer time, and the SFS at linked neutral sites changes more as a result (Figs. 4.3a, 4.3b). In a slowly changing environment, the time-dependent heterozygosity, $H(t)$, oscillates between the static heterozygosities, $H(\bar{s} + \sigma)$ and $H(\bar{s} - \sigma)$, reaching equilibrium in both positive and negative cycles of selection coefficient (inset of Fig. 4.3).

The effect of changing environment on mean heterozygosity for slow and fast Muller's ratchet is the same as that on SFS described above. For the slow Muller's ratchet, $\bar{H} < H(\bar{s})$, whereas when Muller's ratchet clicks fast, $H(\bar{s}) < \bar{H}$. The effect of changing environment is strong for the moderately strong selection ($N\bar{s} \lesssim 100$) as depicted by Fig. 4.4, where even in the slowly changing environment, the mean heterozygosity has deviated to 30 – 40% (inset of Fig. 4.4) from the corresponding heterozygosity in the static environment.

When the selection strength is strong ($N\bar{s} \gg 1$), and Muller's ratchet operates slowly, the time-averaged heterozygosity in the changing environment can be approximated by the mean of heterozygosity in the static environment with selection coefficient $\bar{s} + \sigma$, and

$\bar{s} - \sigma$,

$$\bar{H} \approx \frac{\theta}{2} \left(e^{-\frac{u_d}{\bar{s} + \sigma}} + e^{-\frac{u_d}{\bar{s} - \sigma}} \right) \quad (4.7)$$

where θ is the scaled mutation rate, $\theta = 2Nu_n$. In the strong selection limit ($\frac{u_d}{\bar{s}} \ll 1$), on expanding the *R.H.S* of above equation, we find

$$H(\bar{s}) - \bar{H} \approx \left(\frac{u_d}{\bar{s}} \right) \left(\frac{\sigma^2}{\bar{s}^2 - \sigma^2} \right) \quad (4.8)$$

which shows that mean heterozygosity (4.7) in changing environment is always smaller than that in static environment ($H(\bar{s}) = \theta e^{-u_d/\bar{s}}$), as shown in Fig. 4.4, and the deviation is given by,

Using the same argument, the time-averaged SFS in the slowly changing environment is smaller than that in the static environment for strong selection, as depicted in Fig. 4.3a.

When the selection strength is moderately strong, and σ is of the order \bar{s} (but smaller than \bar{s}), the selection coefficient in the second half of the cycle becomes weak ($\bar{s} - \sigma$), where Muller's ratchet operates fast, and the mutation classes other than the least loaded class can contribute to the SFS. The mean heterozygosity in fast ratchet regime when $\bar{s} - \sigma$ becomes very weak can be approximated by the neutral heterozygosity θ whereas the selection in the first half of the cycle, $\bar{s} + \sigma$, is strong and only the least loaded class contributes to the heterozygosity. The mean heterozygosity is then given by

$$\bar{H} \approx \frac{\theta}{2} \left(1 + e^{-\frac{u_d}{\bar{s} + \sigma}} \right) \quad (4.9)$$

which is greater than the static heterozygosity ($H(\bar{s}) = \theta e^{-u_d/\bar{s}}$). The time-averaged heterozygosity behaves non-monotonically with selection strength in the changing environment as observed in Fig. 4.4. The similar dependence is shown by GORDO *et al.*

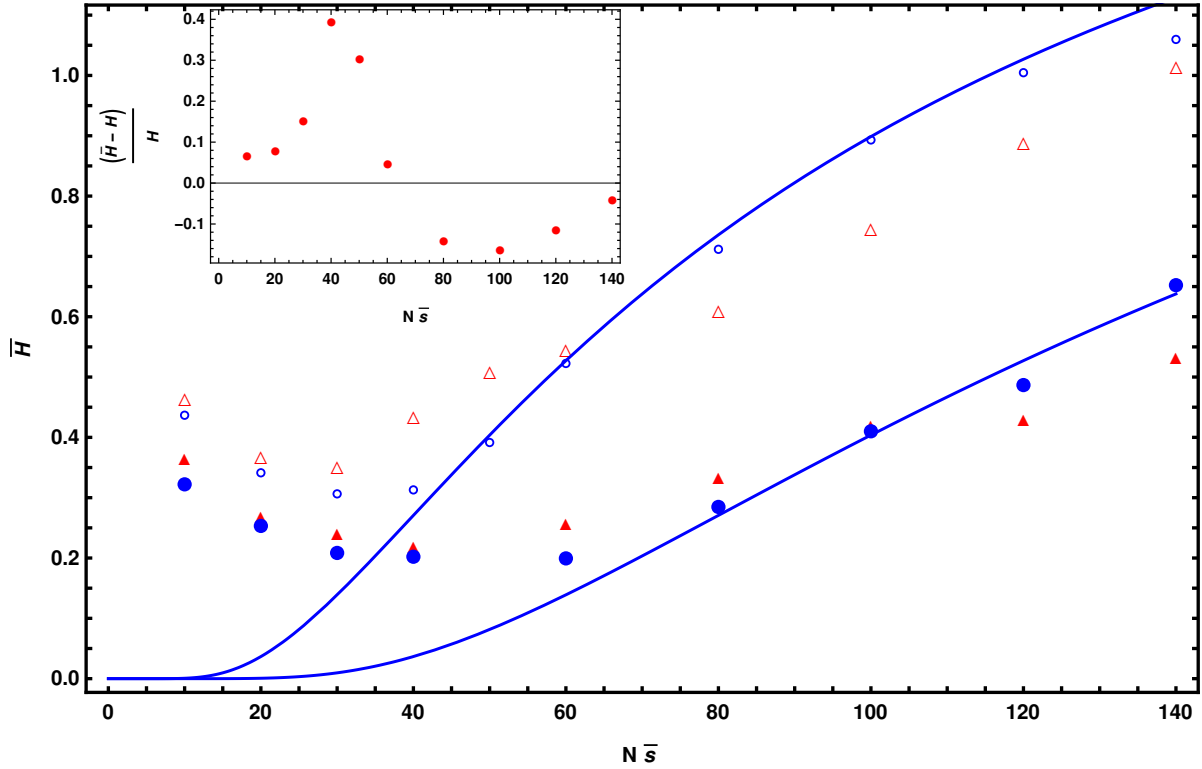


Fig. 4.4 Strictly deleterious mutations: The mean heterozygosity as a function of scaled selection coefficient for static environment (circles) and slowly changing environment with selection coefficient (triangles) varies between $\bar{s} + \sigma$ and $\bar{s} - \sigma$ with time period T_p . The other parameters are $N = 1000$ (open symbols), $N = 2000$ (closed symbols), $Nu_n = 1$, $u_d = 0.08$, $\sigma = \bar{s}/2$ and $T_p = 12000$ respectively. The solid lines represent mean heterozygosity for the neutral population with effective population size as $2Nu_n e^{-u_d/\bar{s}}$. The inset shows the relative change in the mean heterozygosity due to the changing environment as compared to the static environment for $N = 1000$.

[11], and GOOD *et al.* [30] for the mean number of pairwise differences in the static environment. Since the two regimes of Muller's ratchet are separated by $N\bar{s}e^{-u_d/\bar{s}} > 1$ [31, 32], the minimum of the mean heterozygosity scales linearly with $N\bar{s}$ for the fixed u_d/\bar{s} as shown in Fig. 4.4.

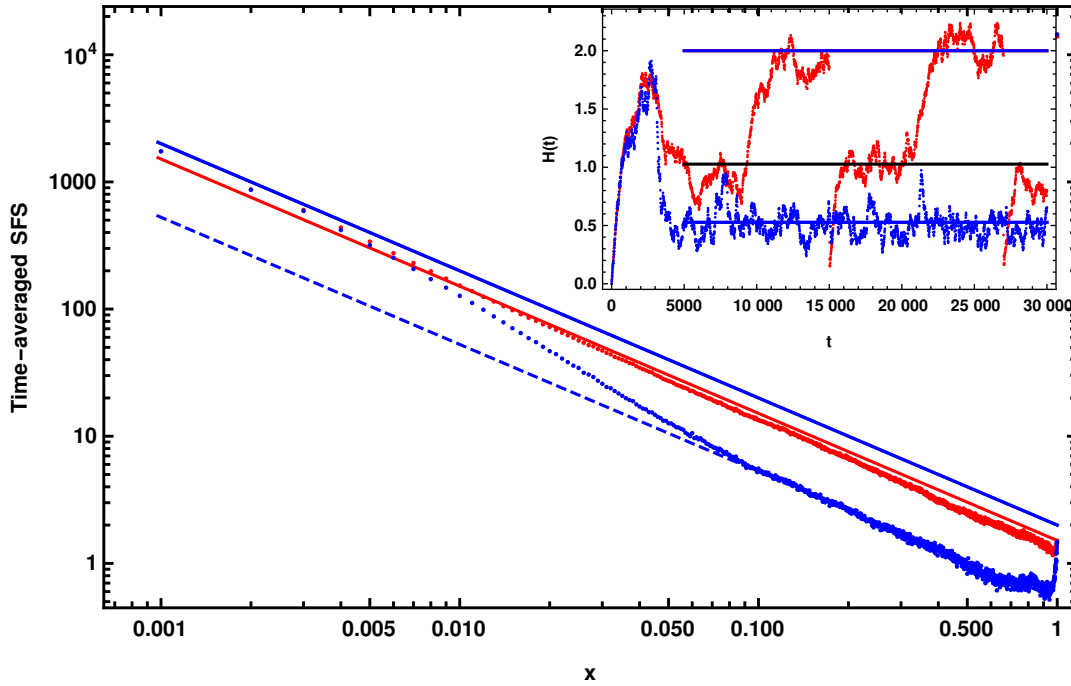


Fig. 4.5 Deleterious and neutral mutations: The time-averaged SFS, when the environment is static (blue) with selection coefficient (\bar{s}) and changes slowly (red) with selection coefficient varies between 0 (neutral) and $2\bar{s}$ (deleterious), and the Muller's ratchet clicks slowly in the deleterious cycle. The blue solid and blue dashed line represents SFS for the neutral mutation with population size N and $N_{eff} = Ne^{-u_d/\bar{s}}$ respectively. The red solid line represents the time-averaged SFS given as, $\frac{\theta}{2x} (1 + e^{-u_d/\bar{s}})$. The other parameters are $N = 1000$, $u_d = 0.08$, $u_n = 0.001$, $\bar{s} = 0.06$, and $\theta = 2Nu_n$. The inset shows the time-dependent heterozygosity in the slowly changing environment (red) which oscillates between the corresponding heterozygosity in the neutral population with size N and N_{eff} respectively. For the static environment (blue), the heterozygosity in the equilibrium population is given by $\theta e^{-u_d/\bar{s}}$. The solid lines from top to bottom represents θ , $\theta e^{-u_d/2\bar{s}}$ and $\theta e^{-u_d/\bar{s}}$ respectively. The population evolves under drift neutral mutations equilibrium for SFS till $t = 3000$, and after that, deleterious mutations are introduced.

4.3.3 Fluctuating environment between neutral and deleterious mutations

Here, we examine the scenario where the environment changes in such a way that the mutation is deleterious in the first half of the cycle where background selection operates, but neutral in the second half of the cycle where the neutral mutation can get eliminated or fixed due to genetic drift.

Figure 4.5 shows that, for the moderate selection, the time-averaged SFS in changing environment at the low frequencies (0.01 – 0.05) has a different shape than that in the static environment with a time-averaged selection coefficient, and it is more skewed in the latter case. The contribution of the neutral cycle (θ/x) to the SFS is larger than that of the deleterious cycle ($\theta e^{-u_d/2\bar{s}}/x$), resulting in the time-averaged SFS in the changing environment to have the similar shape to that in the neutral environment as depicted in Fig. 4.5.

Fig. 4.6 shows that \bar{H} in a slowly changing environment is significantly different from heterozygosity in a static environment ($H(\bar{s})$) when the selection is moderately strong,

$$\bar{H} \approx \frac{\theta}{2} \left(1 + e^{-\frac{u_d}{2\bar{s}}} \right) \quad (4.10)$$

whereas in the static environment, $H = \theta e^{-u_d/\bar{s}}$ where $\theta = 2Nu_n$. The mean heterozygosity in slowly changing environment, \bar{H} , is always greater than that in the static environment, $H(\bar{s})$, as depicted in Fig. 4.6. The deviation is approximately $(3/4)(u_d/\bar{s})$ when $(u_d/\bar{s}) < 1$. In the slowly changing environment, $\bar{H} = \theta$ in limits $N\bar{s} \rightarrow 0$ and $N\bar{s} \rightarrow \infty$ as shown in Fig. 4.4. The time-averaged heterozygosity follows the non-monotonic behavior with selection strength similar to the previous case where the mutation is always deleterious, and here also, the minimum of the time-averaged heterozygosity scales linearly with $N\bar{s}$ as shown in the inset of Fig. 4.6. Hence, these results suggest that the levels of genetic variation at the linked neutral sites are significantly different in the changing selective environment for the deleterious mutation than in the static environment.

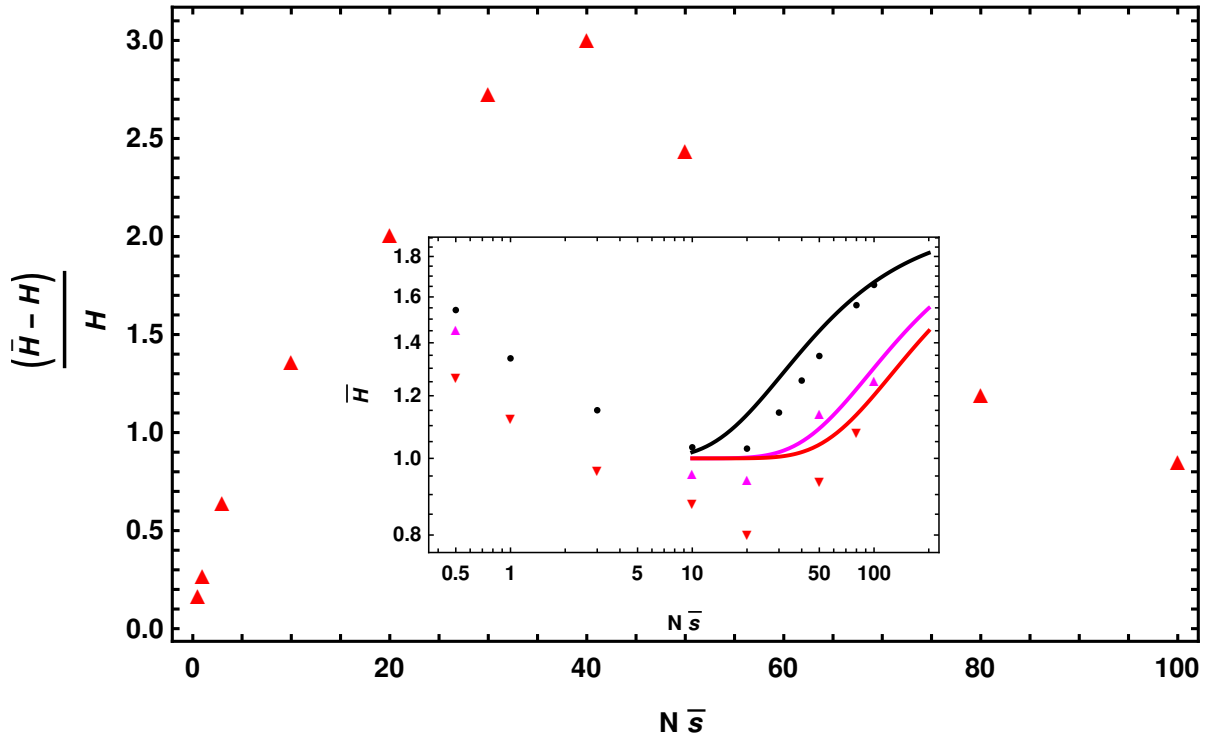


Fig. 4.6 Deleterious and neutral mutations: The relative change in the mean heterozygosity in the changing environment (\bar{H}) from the static environment (H) when selection coefficient varies periodically between zero (neutral) and $2\bar{s}$ (deleterious) with time period T_p . The other parameters are $N = 1000$, $Nu_n = 1$, $u_d = 0.8$ and $T_p = 12000$ respectively. The inset shows that the time-averaged heterozygosity behaves non-monotonically with the scaled selection coefficient and the minimum of the mean heterozygosity scales linearly with $N\bar{s}$. The other parameters are $N = 1000$ (\bullet), $N = 1500$ (\blacktriangle) and $N = 2000$ (\blacktriangledown) respectively. The solid lines represents $2Nu_n e^{-u_d/\bar{s}}$ for $N = 1000, 1500$, and 2000 from top to bottom.

4.4 Discussion

In a recent study by KAUSHIK and JAIN [13], it was found that the Maruyama-Kimura symmetry [14, 15] for the conditional mean fixation time in constant environment does not hold in changing environment. In this paper, we have explored this symmetry for the more general case when inbreeding occurs in the population. We find that the deviation in the conditional mean fixation time for both beneficial as well as deleterious sweeps is maximum in the case of randomly mating population. When the inbreeding coefficient is high, the symmetry between beneficial and deleterious conditional mean fixation time

breaks very mildly, which makes it difficult to distinguish the diversity patterns generated due to beneficial and deleterious sweeps.

Since the effects of changing environment are more significant in deleterious mutations in the randomly mating population, we investigated the effect of changing environment when recurrent deleterious mutations arise at the selected site and are eliminated from the population [9]. When the selection strength is moderate, the effect of background selection on time-averaged SFS is found to be qualitatively different from that in the static environment [29].

When the population fluctuates between zero and a negative selection coefficient in a slowly changing environment, the time-averaged SFS and mean heterozygosity are larger than in a static environment with a time-averaged selection coefficient. Whereas, when the selection coefficient fluctuates in such a way that it is always negative, the time-averaged SFS and mean heterozygosity are smaller (larger) than that in the static environment with a time-averaged selection coefficient when the Muller's ratchet clicks fast (slow). These findings imply that the recurrent deleterious mutations at the selected site result in qualitatively different levels of genetic variation in the changing environment.

Here, we have studied the effect of changing environment on the SFS due to the background selection using numerical simulations. However, there is a scope to develop the analytical theory to study the behavior of non-equilibrium SFS in the changing environment when Muller's ratchet operates fast. Here, we have the selected site being either deleterious or neutral; it will be interesting to look at the case where the selected site is beneficial where selective sweeps also occur. Throughout this Chapter, we have assumed that population size is constant, and it will be interesting to explore the effects of demography. Studying the background selection in the changing environment for the selfing population is also interesting because inbreeding is expected to reduce genetic diversity significantly.

This article has been submitted to bioRxiv. See the bioRxiv link below:

<https://doi.org/10.1101/2022.08.29.505661>

Bibliography

- [1] KIMURA, M., 1971 Theoretical foundations of population genetics at the molecular level. *Theo. Pop. Biol.* **2**: 174–208.
- [2] LEWONTIN R. C., 1974 *The Genetic Basis of Evolutionary Change..* New York: Columbia University Press.
- [3] BUFFALO, V., 2021 Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin’s paradox. *eLife* **10**: e67509.
- [4] CHARLESWORTH, B. and J. D. JENSEN, 2022 How can we resolve Lewontin’s paradox? *Gen. Biol. and Evol.* **14**: evac096.
- [5] MAYNARD SMITH, J. and J. HAIGH, 1974 Hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- [6] BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**: 1111–1117.
- [7] STEPHAN, W., T. H. E. WIEHE, and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Pop. Bio.* **41**: 237–254.
- [8] STEPHAN, W., 2019 Selective Sweeps. *Genetics* **211**: 5–13.

-
- [9] CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- [10] HUDSON, R. R. and N. L. KAPLAN, 1995 The coalescent process and background selection. *Phil. Trans.: R. Soc. Lond.B* **349**: 19–23.
- [11] GORDO, I., A. NAVARRO, and B. CHARLESWORTH, 2002 Muller’s ratchet and the pattern of variation at a neutral locus. *Genetics* **161**: 835–848.
- [12] BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- [13] KAUSHIK, S. and K. JAIN, 2021 Time to fixation in changing environments. *Genetics* **219**: iyab 148.
- [14] MARUYAMA, T., 1974 The age of an allele in a finite population. *Genetics. Res. Camb.* **23**: 137–143.
- [15] MARUYAMA, T. and M. KIMURA 1974 A note on the speed of gene frequency changes in reverse directions in a finite population. *Evolution* **28**: 161–163.
- [16] MULLER, H. J., 1964 The relation of recombination to mutational advance. *Mutat. Res.* **1**: 2–9.
- [17] FELSENSTEIN , J., 1974 The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- [18] BARTON, N. H., 2000 Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. B* **355** : 1553–62.
- [19] GLÉMIN, S., 2012 Extinction and fixation times with dominance and inbreeding. *Theo. Pop. Biol.* **81**: 310–316.

-
- [20] CHARLESWORTH, B. and D. CHARLESWORTH, 2010 *Elements of Evolutionary Genetics*. Roberts and Company, Colorado.
- [21] CROW, J. F. and M. Kimura, 2009 *An introduction to population genetics theory*. The Blackburn Press, New Jersey.
- [22] HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection. V. Proc. Camb. Philos. Soc. **23**: 838–844.
- [23] MARTIN, G. and A. LAMBERT, 2015 A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. Theo. Pop. Biol. **101**: 40–46.
- [24] EWING, G., J. HERMISSON, P. PFAFFELHUBER, and J. RUDOLF, 2011 Selective sweeps for recessive alleles and for other modes of dominance. J. Math. Biol. **63**: 399–431.
- [25] RISKEN, H., 1996 *The Fokker Planck equation. Methods of solution and applications*. Springer, Berlin.
- [26] FELLER, W., 1951a Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 227–246.
- [27] FISHER, R. A., 1930 *The genetical theory of natural selection*. Oxford: Clarendon Press.
- [28] WRIGHT, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.
- [29] CVIJOVIĆ, I., B. H. GOOD and M. M. DESAI, 2018 The effect of strong purifying selection on genetic diversity. Genetics **209**: 1235–1278.
- [30] GOOD, B. H., A. M. WALCZAK, R. A. NEHER, and M. M. DESAI, 2014 Genetic giversity in the interference selection limit. PloS Genetics **10**: e1004222.

-
- [31] GORDO, I. and B. CHARLESWORTH, 2000a The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics* **154**: 1379–1387.
- [32] JAIN, K., 2008 Loss of least-loaded class in asexual populations due to drift and epistasis. *Genetics* **179**: 2125–2134.

Chapter 5

Summary and open questions

The primary focus of this thesis is to understand how the various evolutionary forces shape the genetic diversity within a population. Since natural environments are not static, we have considered the situation when the environment is continually changing. In this thesis, we have studied the following three quantities in detail to quantify the effect of changing environments on genetic variation: conditional mean fixation time, site frequency spectrum (SFS), and heterozygosity. We have considered the environmental change due to changing selection coefficient and changing population size, and their effects on the genetic variations are summarised below.

Summary of results

Case 1: Changing selection coefficient and constant population size

In Chapters 2 and 4, we have studied the effect of selective sweeps and background selection on the linked neutral genetic variation when the selection coefficient is time-dependent but population size is constant in time.

Selective sweeps: The fixation of the mutant (selective sweep) at the selected site reduces the genetic variation at the linked neutral sites, and the reduction in the diversity is proportional to the conditional fixation time of the mutant. We find the following main results in the case of selective sweeps.

i) The Maruyama-Kimura symmetry for the conditional mean fixation time in static environment is not preserved in the changing environment, suggesting that the genetic diversity patterns generated by deleterious and beneficial sweeps are no longer identical.

ii) The constant environment assumption holds for the beneficial mutations, whereas the deleterious mutations are strongly impacted due to the changing environment. In particular, the recessive deleterious mutations are strongly affected.

iii) Even the slowly changing environment produces qualitatively different genetic diversity levels.

iv) The changing environment effect is most substantial for the randomly mating population, and the constant environment assumption holds for beneficial and deleterious sweeps when the inbreeding coefficient is high.

Background selection: The deleterious sweeps are very rare, and in most cases, the deleterious mutations arise in the population and are eliminated from the population due to their adverse effects on fitness (background selection). Hence, we have studied the effect of background selection on genetic diversity in the changing environment and found that:

i) When the population fluctuates between a neutral and negative selection coefficient, the time-averaged heterozygosity and SFS are always greater than that in the static environment with a time-averaged selection coefficient.

ii) When the selection coefficient is changing but it is always negative, the time-averaged heterozygosity and SFS are greater (smaller) than the corresponding heterozy-

gosity and SFS in the static environment with a time-averaged selection coefficient when the population experience slow (fast) Muller's ratchet.

Case 2: Combined effect of changing selection and changing population size

In Chapter 3, we have studied the site frequency spectrum in a randomly mating diploid population in which both population size and selection coefficient change periodically with time. We find that the time-averaged SFS for the changing environment is qualitatively and quantitatively different compared to that in the static environment. The main findings are:

- i) The time-averaged SFS for an on-average neutral mutations mimics the shape of the static SFS for the beneficial mutations.
- ii) The deleterious mutations are responsible for the deviation in the time-averaged SFS in the changing environment, and the extent of the deviation depends upon the phase difference between the population size and selection coefficient.

Open questions

Finally, we discuss some open questions related to our study.

- i) In Chapter 3, we have considered the infinite sites model where the sites are completely unlinked. It will be interesting to generalize the infinite sites model to include finite recombination and investigate the combined effect of changing population size and selection.
- ii) In Chapter 4, we have numerically studied the non-equilibrium SFS for the deleterious mutations. Still, there is a scope to develop the analytical theory which can potentially capture the effect of Muller's ratchet on non-equilibrium SFS in the changing

environment. The non-equilibrium SFS will also be interesting to study for neutral and beneficial mutations where selective sweeps also occurs.

Appendix A

A.1 Diffusion theory for time-inhomogeneous process

For a large finite population with small selection coefficient, the average fixation time can, in principle, be studied using the backward Fokker-Planck equation with time-dependent selection coefficient. The probability distribution $\Phi_b(x, t|p, t_0)$ obeys the following partial differential equation:

$$-\frac{\partial \Phi_b(x, t|p, t_0)}{\partial t_0} = s(t_0)g(p)\frac{\partial \Phi_b(x, t|p, t_0)}{\partial p} + \frac{pq}{2N}\frac{\partial^2 \Phi_b(x, t|p, t_0)}{\partial p^2} \quad (\text{A.1.1})$$

where $s(t_0) = \bar{s} + \sigma \sin(\omega t_0 + \theta)$ and $g(p) = pq(p + h(1 - 2p))$. In the above equation, the first term on the RHS is obtained on using that the deterministic rate of change of the mutant allele frequency is given by $dp/dt = s(t)g(p)$, and the second term is due to the sampling noise in a finite population. In [A.1.1](#), the probability distribution is assumed to be a function of the initial time t_0 and the final time t . But one can also consider the variables t_0 and the time interval $t - t_0$ which leads to (S5) in [UECKER and HERMISSON \[1\]](#). However, as the formulation ([A.1.1](#)) is much easier to handle, we work with it in the [Chapter 2](#).

The (unconditional) mean fixation time for a mutant arising at time t_0 can be written as

$$\bar{T}(p, t_0) = \int_{t_0}^{\infty} dt(t - t_0)\Phi_b(x \rightarrow 1, t|p, t_0) \quad (\text{A.1.2})$$

Using Leibniz integral rule,

$$\int_{t_0}^{\infty} dt(t - t_0) \frac{\partial \Phi_b(x, t|p, t_0)}{\partial t_0} = \frac{\partial}{\partial t_0} \int_{t_0}^{\infty} dt(t - t_0)\Phi_b(x, t|p, t_0) + \int_{t_0}^{\infty} dt \Phi_b(x, t|p, t_0), \quad (\text{A.1.3})$$

and (A.1.1), we then obtain

$$-\frac{\partial \bar{T}(p, t_0)}{\partial t_0} - u(p, t_0) = s(t_0)g(p) \frac{\partial \bar{T}(p, t_0)}{\partial p} + \frac{pq}{2N} \frac{\partial^2 \bar{T}(p, t_0)}{\partial p^2} \quad (\text{A.1.4})$$

where the eventual fixation probability $u(p, t_0) = \int_{t_0}^{\infty} dt \Phi_b(x \rightarrow 1, t|p, t_0)$ obeys [1, 2]

$$-\frac{\partial u(p, t_0)}{\partial t_0} = s(t_0)g(p) \frac{\partial u(p, t_0)}{\partial p} + \frac{pq}{2N} \frac{\partial^2 u(p, t_0)}{\partial p^2} \quad (\text{A.1.5})$$

We verify that (A.1.4) and (A.1.5) reduce to the corresponding equations for the time-homogenous process where the fixation probability and the fixation time are independent of the initial time [3]. The partial differential equations (A.1.4) and (A.1.5) along with the boundary conditions

$$u(0, t_0) = 0, u(1, t_0) = 1, \bar{T}(0, t_0) = \bar{T}(1, t_0) = 0 \quad (\text{A.1.6})$$

can, in principle, be used to find the mean fixation time for either sign of selection. However, these equations do not appear to be solvable, even for the dominance parameter $h = 1/2$, as the eigenfunction expansion method commonly employed for solving partial differential equations with time-dependent coefficients requires the eigenfunctions of the problem with constant selection that are, unfortunately, not known in a closed form [4].

A.2 Mean fixation time in slowly changing environments

In slowly changing environments where $\omega \ll N^{-1}$, $s(0)$ and $Ns(0)$ is arbitrary, the eventual fixation probability and the mean fixation time can be expanded in a power series in the small, dimensionless parameter $N\omega$, that is, $u = \sum_{i=0}^{\infty} (N\omega)^i u_i$, $\bar{T} = \sum_{i=0}^{\infty} (N\omega)^i \bar{T}_i$. But for small enough $N\omega$, it is a good approximation to terminate this series at $i = 1$.

To find the equation satisfied by u_i , it is useful to rewrite (A.1.5) for the fixation probability as

$$-N\omega \frac{\partial u(p, t_0)}{\partial \Theta} = Ns(\Theta)g(p) \frac{\partial u(p, t_0)}{\partial p} + \frac{pq}{2} \frac{\partial^2 u(p, t_0)}{\partial p^2} \quad (\text{A.2.1})$$

where $\Theta = \omega t_0 + \theta$. Substituting the power series expansion for u on both sides of the above equation, collecting terms with the same power of $N\omega$ and taking $\Theta \rightarrow \theta$ (that is, $\omega \rightarrow 0$), we obtain (2.4) and (2.6) in the main text for u_1 and u_0 , respectively. The fixation probability in a static environment obeys the boundary conditions, $u_0(0, t_0) = 0$, $u_0(1, t_0) = 1$. Therefore, due to (A.1.6), $u_1(0, t_0) = u_1(1, t_0) = 0$. In a similar fashion, the equations (2.5) and (2.7) for the mean fixation time and the corresponding boundary conditions can be derived.

The conditional mean fixation time in slowly changing environment can be written as

$$\bar{T}_c = \frac{\bar{T}}{u} \approx \frac{\bar{T}_0}{u_0} \left(\frac{1 + N\omega \frac{\bar{T}_1}{\bar{T}_0}}{1 + N\omega \frac{u_1}{u_0}} \right) \quad (\text{A.2.2})$$

from which (2.8) follows on using $(1 + \epsilon_1)/(1 + \epsilon_2) \approx 1 + \epsilon_1 - \epsilon_2$ for small ϵ_1, ϵ_2 .

A.3 Feller process with time-dependent coefficients

Taking the Laplace transform on both sides of (2.13), we find that $\tilde{\mathcal{F}}(\kappa, \tau) = \int_0^\infty dp e^{-\kappa p} \mathcal{F}(p, \tau)$ obeys a first order differential equation,

$$\frac{\partial \tilde{\mathcal{F}}}{\partial \tau} = (\kappa - \kappa^2 \ell(\tau)) \frac{\partial \tilde{\mathcal{F}}}{\partial \kappa} \quad (\text{A.3.1})$$

where $\tau(t) = h \int_0^t dt' s(t')$ and $\ell(\tau) = (2Nhs(t))^{-1}$. The above differential equation can be solved using the method of characteristics for the initial condition $\mathcal{F}(p, 0) = \delta(p - p_0)$, and we obtain [5, 6]

$$\tilde{\mathcal{F}}(\kappa, \tau) = \exp \left[- \frac{p_0 \kappa e^\tau}{1 + \kappa e^\tau \int_0^\tau d\tau' e^{-\tau'} \ell(\tau')} \right] \quad (\text{A.3.2})$$

$$= e^{-\frac{p_0}{\int_0^\tau d\tau' e^{-\tau'} \ell(\tau')}} \times \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{p_0}{\int_0^\tau d\tau' e^{-\tau'} \ell(\tau')} \right)^n \frac{1}{(1 + \kappa \int_0^\tau d\tau' e^{-\tau'} \ell(\tau'))^n} \quad (\text{A.3.3})$$

Taking the inverse Laplace transform of the summand in the last expression and then carrying out the sum over n , we get

$$\mathcal{F}(p, t) = \frac{1}{\langle p \rangle} \sqrt{\frac{\langle p \rangle}{p}} \frac{2Np_0 e^{-\frac{2Np_0(1+\frac{p}{\langle p \rangle})}{\int_0^\tau d\tau' e^{-\tau'} (hs(t'))^{-1}}}}{\int_0^\tau d\tau' e^{-\tau'} (hs(t'))^{-1}} I_1 \left(\frac{4Np_0 \sqrt{\frac{p}{\langle p \rangle}}}{\int_0^\tau d\tau' e^{-\tau'} (hs(t'))^{-1}} \right) \quad (\text{A.3.4})$$

where $I_n(z)$ is modified Bessel function of the first kind. In (A.3.4), $\langle p(t) \rangle = \int_0^\infty dp \mathcal{F}(p, t) p = p_0 e^\tau$ is the expected mutant allele frequency at time t , as can also be checked using (2.13); however, when conditioned on fixation, this frequency grows as $p_0 e^\tau / u$ which is faster than $\langle p \rangle$ [1]. We also note that the eventual fixation probability (2.14) can also be written as $u = \text{Lim}_{\kappa, \tau \rightarrow \infty} \tilde{\mathcal{F}}(\kappa, \tau)$.

A.4 Distribution of the fixation time of on-average beneficial mutant

In the stochastic phase A , although the expected mutant frequency grows exponentially with time, (A.3.4) and (2.14) show that at large times (where $\tau \rightarrow \infty$, since $s(t) > 0$ at all times), the random variable $y = p/\langle p \rangle$, conditioned on fixation, has a stationary distribution,

$$\mathcal{F}_c(y, t) = \frac{\mathcal{F}(y\langle p \rangle, t)\langle p \rangle}{1 - \mathcal{F}(0, t)} \xrightarrow{\tau \rightarrow \infty} ue^{-uy} \quad (\text{A.4.1})$$

(see also UECKER and HERMISSON [1]). Furthermore, in the vicinity of time t_1 , the mutant frequency in the stochastic phase A is given by $p = (p_0 y) e^{\tau(t)}$, $t \lesssim t_1$ and in the deterministic phase B , the average mutant frequency grows as $p(t) \approx p(t_1) e^{\tau(t) - \tau(t_1)}$, $t \gtrsim t_1$. Thus the initial mutant frequency in the deterministic phase is given by $p(t_1) = p_0 y e^{\tau(t_1)}$, and from (A.4.1), it follows that the random variable $p(t_1) e^{-\tau(t_1)}$ is exponentially distributed with mean $(2Nu)^{-1}$.

In the deterministic phase B that begins at time t_1 and ends at time t_2 , the average mutant frequency obeys $dp/dt = s(t)g(p)$; integrating this equation over time from t_1 to t_2 , we get $\tau(t_2) - \tau(t_1) = h[\mathcal{D}(p(t_2)) - \mathcal{D}(p(t_1))]$, where

$$\mathcal{D}(p) = \frac{\ln p}{h} - \frac{\ln q}{1-h} + \frac{2h-1}{h(1-h)} \ln(h + (1-2h)p). \quad (\text{A.4.2})$$

But as the frequency $p(t_1) \rightarrow 0$, $q(t_2) \rightarrow 0$, the initial frequency $q(t_2)$ in phase C is related to $p(t_1)$ as

$$\frac{h}{h-1} \ln[q(t_2) e^{\frac{1-h}{h}\tau(t_2)}] \approx \ln[p(t_1) e^{-\tau(t_1)}] - \frac{1-2h}{1-h} \ln\left(\frac{h}{1-h}\right) \quad (\text{A.4.3})$$

In the stochastic phase C , the wildtype population evolves stochastically from time t_2 until it goes extinct at time T_c . The wildtype frequency can be described by a Feller process

that obeys (2.12) for the distribution $\hat{\mathcal{F}}(q, t)$ when $p \rightarrow q, s(t) \rightarrow -s(t), h \rightarrow 1 - h$. For constant selection, it has been claimed that looking backward in time (that is, $t \rightarrow T_c - t$), the wildtype frequency obeys the same dynamics as the mutant frequency with $h \rightarrow 1 - h$ but s unchanged [7]; however, this prescription results in a forward Kolmogorov equation with negative population size which is clearly absurd. Therefore, we will work always looking forward in time.

Proceeding in a manner similar to that for stochastic phase A , the distribution $\hat{\mathcal{F}}(q, t)$ for the wildtype frequency subject to the initial condition $q(t_2)$ can be found for $t > t_2$. Then the probability that the wildtype goes extinct by time T_c is given by (refer (2.14) for a comparison)

$$1 - \int_0^\infty dq \hat{\mathcal{F}}(q, T_c) = \exp \left[- \frac{2Nq(t_2)}{\int_{t_2}^{T_c} dt e^{(1-h) \int_{t_2}^t dt' s(t')}} \right] \quad (\text{A.4.4})$$

$$\stackrel{T_c \gg t_2}{\approx} \exp \left[- \frac{2Nq(t_2) e^{\frac{1-h}{h} \tau(t_2)}}{\int_0^{T_c} dt e^{(1-h) \int_0^t dt' s(t')}} \right] \quad (\text{A.4.5})$$

On taking the derivative of the above cumulative distribution with respect to T_c and averaging over the distribution of $q(t_2)$ which can be found using (A.4.1) and (A.4.3), we finally arrive at (2.15) in the main text.

A.5 Mean fixation time of on-average beneficial mutant

To find the conditional mean fixation time given by (2.18), we need to express T_c as a function of Υ_C using (2.16) which is given by

$$\frac{2N}{\Upsilon_C} = \int_0^T dt' e^{(1-h)\bar{s}t'} e^{-\frac{(1-h)\sigma}{\omega} (\cos(\omega t' + \theta) - \cos(\theta))} \quad (\text{A.5.1})$$

For small ω , we first expand the exponent of the integrand on the RHS to linear order in cycling frequency to obtain

$$\frac{2N}{\Upsilon_C} \approx \int_0^T dt' e^{s_0 t'} e^{s_1 \omega \cos \theta \frac{t'^2}{2}} \approx \int_0^T dt' e^{s_0 t'} \left(1 + s_1 \omega \cos \theta \frac{t'^2}{2} \right) \quad (\text{A.5.2})$$

where $s_0 = (1 - h)s(0)$ and $s_1 = (1 - h)\sigma$. On carrying out the integrals and taking the logarithm on both sides, to order ω , we get

$$\ln \left(\frac{2N s_0}{\Upsilon_C} \right) = s_0 T + \frac{s_1 \omega \cos \theta}{s_0^2} \left(1 - s_0 T + \frac{s_0^2}{2} T^2 \right) \quad (\text{A.5.3})$$

which can be inverted to finally give

$$s_0 T = A + \frac{\omega s_1 \cos \theta}{s_0^2} \left(A - 1 - \frac{A^2}{2} \right) \quad (\text{A.5.4})$$

where $A = \ln \left(\frac{2N s_0}{\Upsilon_C} \right)$. Using this expression in the inner integral of (2.18) and carrying out the integrals, to leading and subleading orders in α , we obtain (2.19) and (2.20) in the main text.

A.6 Mean fixation time of initially deleterious mutant under strong selection

Below we consider the conditional mean fixation time of a co-dominant mutant in slowly changing environment ($\omega \ll N^{-1} \ll \sigma$) which is neutral on average ($\bar{s} = 0$) and can be written as $v_c = v_{0,c} + N\omega v_{1,c}$ with $v_{1,c} = \frac{v_1}{u_0} - v_{0,c} \frac{u_1}{u_0}$ (see (2.8)). In a static environment with selection coefficient $s(0) = \sigma \sin \theta$, the eventual fixation probability u_0 and the

unconditional mean fixation time v_0 are given by [3]

$$u_0(p, \alpha) = \frac{1 - e^{-\alpha p}}{1 - e^{-\alpha}} \quad (\text{A.6.1})$$

$$v_0(p, \alpha) = u_0(p, \alpha)G_0(1, \alpha) - G_0(p, \alpha) \quad (\text{A.6.2})$$

where

$$G_0(p, \alpha) = \frac{1}{\alpha} \int_0^p dy \frac{1 - e^{-\alpha(p-y)}}{y(1-y)} u_0(y, \alpha) \quad (\text{A.6.3})$$

and $\alpha = N\sigma \sin \theta$. In a slowly changing environment, due to (2.4) and (2.5) in the main text, the corresponding quantities are given by

$$\frac{u_1(p, \alpha)}{2 \cot \theta} = u_0(p, \alpha)H_1(1, \alpha) - H_1(p, \alpha) \quad (\text{A.6.4})$$

$$\frac{v_1(p, \alpha)}{u_0(p, \alpha)} = G_1(1, \alpha) - \frac{G_1(p, \alpha)}{u_0(p, \alpha)} \xrightarrow{p \rightarrow 0} G_1(1, \alpha) \quad (\text{A.6.5})$$

where

$$H_1(p, \alpha) = \int_0^p dy \frac{1 - e^{-\alpha(p-y)}}{y(1-y)} \frac{\partial u_0(y, \alpha)}{\partial \alpha} \quad (\text{A.6.6})$$

$$G_1(p, \alpha) = \int_0^p dy \frac{1 - e^{-\alpha(p-y)}}{y(1-y)} \left(\frac{u_1(y, \alpha)}{\alpha} + 2 \cot \theta \frac{\partial v_0(y, \alpha)}{\partial \alpha} \right) \quad (\text{A.6.7})$$

As discussed in the main text, for strong selection, the conditional mean fixation time for a mutant that remains beneficial until it fixes can be found within a semi-deterministic approximation and given by (2.20). Here, we therefore consider the difference $\Delta(p, \alpha) = v_{1,c}(p, \alpha) - v_{1,c}(p, -\alpha)$ which, for single initial mutant, is given by

$$\Delta \stackrel{p \rightarrow 0}{=} G_1(1, \alpha) - G_1(1, -\alpha) - v_{0,c}(p, \alpha) \left(\frac{u_1(p, \alpha)}{u_0(p, \alpha)} - \frac{u_1(p, -\alpha)}{u_0(p, -\alpha)} \right) \Big|_{p \rightarrow 0} \quad (\text{A.6.8})$$

To find Δ , we require the conditional mean fixation time $v_{0,c}$ and the fixation probability u_1 for arbitrary initial frequency. Although one can write exact expressions for them in

terms of the exponential integrals, here we are interested in the strong selection regime ($|\alpha| \gg 1$). Using (5.1.10) and (5.1.51) in ABRAMOWITZ and STEGUN [8], to leading order in large α , we obtain

$$v_{0,c}(y, \alpha) \approx \begin{cases} \frac{2(\ln |\alpha| + \gamma)}{|\alpha|} - \frac{y}{2} & , 0 < y < |\alpha|^{-1} & \text{(A.6.9a)} \\ \frac{\ln(1-y) - \ln y + \ln |\alpha| + \gamma}{|\alpha|} & , |\alpha|^{-1} \leq y < 1 & \text{(A.6.9b)} \end{cases}$$

which decreases with the increasing frequency of initial mutants, as expected. For an initially beneficial mutant (that is, $s(0) > 0$), the fixation probability can be approximated by

$$\frac{u_1(y, \alpha)}{2 \cot \theta} \approx -e^{-\alpha y} \ln(1-y) , \quad 0 < y < 1 , \quad \text{(A.6.10)}$$

while for a mutant with $s(0) < 0$, we have

$$\frac{u_1(y, \alpha)}{2 \cot \theta} \approx \begin{cases} -e^{-|\alpha|} \ln |\alpha| (e^{|\alpha|y} - 1) & , 0 < y < |\alpha|^{-1} & \text{(A.6.11a)} \\ e^{-|\alpha|} (e^{|\alpha|y} \ln y + \ln |\alpha|) & , |\alpha|^{-1} \leq y < 1 & \text{(A.6.11b)} \end{cases}$$

The above expressions (A.6.10) and (A.6.11a) reduce, respectively, to (11b) and (11c) of DEVI and JAIN [2] when a single mutant is initially present.

Using (A.6.9a), (A.6.10) and (A.6.11a), we find that

$$\frac{u_1(\alpha, p)}{u_0(\alpha, p)} \stackrel{p \rightarrow 0}{\approx} \frac{2 \cot \theta}{\alpha} , \quad \alpha \gg 1 \quad \text{(A.6.12)}$$

$$\frac{u_1(-|\alpha|, p)}{u_0(-|\alpha|, p)} \stackrel{p \rightarrow 0}{\approx} -2 \cot \theta (\ln |\alpha| + \gamma) , \quad \alpha \ll -1 \quad \text{(A.6.13)}$$

and therefore the last term on the RHS of (A.6.8) is given by $\frac{4 \cot \theta}{|\alpha|} (\ln |\alpha| + \gamma)^2$. To estimate the difference, $G_1(1, \alpha) - G_1(1, -\alpha)$, we first note that for a codominant mutant, the conditional fixation time, $v_0(p, \alpha)$ is symmetric about $\alpha = 0$ for *any* initial frequency

p (refer Sec. 5.4, [3]). From the definition (A.6.7), we then get

$$\begin{aligned} \frac{G_1(1, \alpha) - G_1(1, -\alpha)}{2|\alpha|^{-1} \cot \theta} &= \int_0^1 dy \frac{(1 - e^{-\alpha(1-y)})^{\frac{u_1(y, \alpha)}{2 \cot \theta}}}{y(1-y)} + \int_0^1 dy \frac{(1 - e^{\alpha(1-y)})^{\frac{u_1(y, -\alpha)}{2 \cot \theta}}}{y(1-y)} \\ &+ \int_0^1 dy \frac{2\alpha v_{0,c}(y, \alpha)}{y \sinh(\frac{\alpha}{2})} \sinh\left(\frac{\alpha y}{2}\right) \sinh\left(\frac{\alpha(1-y)}{2}\right) \quad (\text{A.6.14}) \end{aligned}$$

Using the results for $v_{0,c}$ and u_1 found above in the last equation and carrying out the integrals, we find that the first integral on the RHS is of order $1/|\alpha|$, the second integral, to order one, is equal to $\frac{1}{2}(\ln |\alpha|)^2 + \gamma \ln |\alpha|$ and the last integral is given by $\frac{3}{2}(\ln |\alpha|)^2 + 3\gamma \ln |\alpha| + \mathcal{O}(1)$. Putting all these results together shows that Δ decays as $1/|\alpha|$ or faster. As the order one terms in (A.6.14) seem rather hard to calculate, we also studied these integrals numerically and find our numerical analyses to be consistent with $\Delta \sim |\alpha|^{-1}$. Thus, while $v_{1,c} \sim \alpha^{-2}$ for beneficial mutant under strong selection (refer (2.20)), the change in the conditional mean fixation time for deleterious mutant decays slowly as $|\alpha|^{-1}$.

Bibliography

- [1] UECKER, H. and J. HERMISSON, 2011 On the fixation process of a beneficial mutation in a variable environment. *Genetics* **188**: 915–930.
- [2] DEVI, A. and K. JAIN, 2020 The impact of dominance on adaptation in changing environments. *Genetics* **216**: 227–240.
- [3] EWENS, W., 2004 *Mathematical Population Genetics*. Springer, Berlin.
- [4] JAIN, K. and A. DEVI, 2020 Evolutionary dynamics and eigenspectrum of confluent Heun equation. *J. Phys. A: Math Theor.* **53**: 395602.
- [5] FELLER, W., 1951b Two singular diffusion problems. *Annals of Mathematics* **54**: 173–182.
- [6] MASOLIVER, J., 2016 Nonstationary Feller process with time-varying coefficients. *Phys. Rev. E* **93**: 012122.
- [7] MARTIN, G. and A. LAMBERT, 2015 A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theo Pop Biol* **101**: 40–46.
- [8] ABRAMOWITZ, M. and I. A. STEGUN, 1964 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover.

Appendix B

B.1 Equilibrium site frequency spectrum for strong selection

In order to obtain an approximate expression for the SFS when the environment is static and selection is strong, we first note that

$$\int_x^1 dy e^{-Ay^2 - By} = \frac{\sqrt{\pi} e^{\frac{B^2}{4A}} \left(\operatorname{erf}\left(\frac{2A+B}{2\sqrt{A}}\right) - \operatorname{erf}\left(\frac{2Ax+B}{2\sqrt{A}}\right) \right)}{2\sqrt{A}} \quad (\text{B.1.1})$$

$$\underset{|A|, |B| \gg 1}{\sim} e^{\frac{B^2}{4A}} \left(\frac{e^{-A(x + \frac{B}{2A})^2}}{2Ax + B} - \frac{e^{-A(1 + \frac{B}{2A})^2}}{2A + B} \right) \quad (\text{B.1.2})$$

where the last expression is obtained using the asymptotic expansion of the error function for large argument (see (7.1.23) of ABRAMOWITZ and STEGUN [1]), and holds provided $2Ax + B$ and $2A + B$ are nonzero.

Using this result in (3.12), we find that in constant environment, when the population is under strong selection ($\bar{N}|\bar{s}| \gg 1$) and $h \neq 0, 1$, the stationary state SFS can be approximated as

$$f^*(x) = \frac{\theta e^{\bar{N}\bar{s}[(1-2h)x^2 + 2hx]} \int_x^1 dy e^{-\bar{N}\bar{s}[(1-2h)y^2 + 2hy]}}{x(1-x) \int_0^1 dy e^{-\bar{N}\bar{s}[(1-2h)y^2 + 2hy]}} \quad (\text{B.1.3})$$

$$\sim \frac{\theta}{x(1-x)} \frac{\frac{h}{h+x(1-2h)} - \frac{h}{1-h} e^{-\bar{N}\bar{s}l(x)}}{1 - \frac{h}{1-h} e^{-\bar{N}\bar{s}}} \quad (\text{B.1.4})$$

which further leads to

$$\frac{x(1-x)f^*(x)}{\theta} \approx \begin{cases} \frac{h}{h+x(1-2h)} - \frac{he^{-\bar{N}\bar{s}\ell(x)}}{1-h} & , \bar{N}\bar{s} \gg 1 & \text{(B.1.5a)} \\ e^{-\bar{N}|\bar{s}(1-\ell(x))} - \frac{(1-h)e^{-\bar{N}|\bar{s}|}}{h+x(1-2h)} & , \bar{N}\bar{s} \ll -1 & \text{(B.1.5b)} \end{cases}$$

where $\ell(x) = 1 - x^2 - 2h(x - x^2) = (1-x)[(1-x) + 2x(1-h)]$ lies between zero and one. If x is not close to one, the subleading (second) term in the above equations can be neglected, and we arrive at (3.14a) and (3.14b) in the main text.

Using (B.1.5a) and (B.1.5b) in (3.9) for the sample SFS, we obtain

$$f_{n,i}^* \approx \begin{cases} \frac{\theta n {}_2F_1(1, i; n; 2 - \frac{1}{h})}{i(n-i)} - \binom{n}{i} \frac{\theta h}{1-h} \frac{\Gamma(n-i)}{(2\bar{N}\bar{s}(1-h))^{n-i}} & , \bar{N}\bar{s} \gg 1 & \text{(B.1.6a)} \\ \binom{n}{i} \frac{\theta \Gamma(i)}{(2\bar{N}|\bar{s}|h)^i} & , \bar{N}\bar{s} \ll -1 & \text{(B.1.6b)} \end{cases}$$

where $\Gamma(i)$ is the gamma function and ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function [1]. To obtain the above result for negative selection, we needed to evaluate $\int_0^1 dx x^{i-1} (1-x)^{n-i-1} e^{-\bar{N}|\bar{s}|(x^2+2h(x-x^2))}$, which may be approximated by $\int_0^\infty dx x^{i-1} e^{-2h\bar{N}|\bar{s}|x}$ on noting that the main contribution to the integrand comes from small x when $\bar{N}|\bar{s}|$ is large. For positive selection, the second term on the RHS of (B.1.6a) is obtained in a similar fashion, and shows how the asymptotic result is approached.

We briefly discuss the equilibrium SFS for a completely recessive and dominant allele. For $h = 0$, we can use (B.1.2) for the integral in the numerator of the RHS of (B.1.3). But as $2Ax + B = 0$ for $x = 0$, we use the asymptotic expansion of the error function for the integral in the denominator, and finally obtain

$$\frac{x(1-x)f^*(x)}{\theta} \approx \begin{cases} \frac{1}{\sqrt{\bar{N}\bar{s}\pi x}} (1 - xe^{-\bar{N}\bar{s}(1-x^2)}) & , \bar{N}\bar{s} \gg 1 & \text{(B.1.7a)} \\ e^{-\bar{N}|\bar{s}|x^2} & , \bar{N}\bar{s} \ll -1 & \text{(B.1.7b)} \end{cases}$$

B.3 Nonequilibrium site frequency spectrum for neutral population

Although the moments of $g(x, t)$ have been studied when selection is absent and the population size is an arbitrary function of time [4, 5], an explicit expression for $g(x, t)$ for periodically changing population size has not been obtained. For $s(t) = 0$ at all times, from (B.2.2), we have

$$\frac{\partial v(x, t)}{\partial t} - \frac{x(1-x)}{2N(t)} \frac{\partial^2 v(x, t)}{\partial x^2} = r(x, t) \quad (\text{B.3.1})$$

where $r(x, t) = -(1-x) \frac{d\rho}{dt}$. We first write

$$v(x, t) = \sum_m a_m(t) \psi_m(x) \quad (\text{B.3.2})$$

$$r(x, t) = \sum_m b_m(t) \psi_m(x) \quad (\text{B.3.3})$$

where the eigenfunction ψ_n obeys the eigenvalue equation $x(1-x)\psi_n''(x) + 2\rho\lambda\psi_n(x) = 0$ with boundary conditions $\psi_n(0) = \psi_n(1) = 0$, and is orthonormal with respect to the weight function $w(x) = [x(1-x)]^{-1}$, that is, $\int_0^1 dx w(x) \psi_n(x) \psi_m(x) = \delta_{m,n}$. It is known that the eigenfunction $\psi_n(x) = C_n x(1-x) P_{n-1}^{(1,1)}(1-2x)$ with the eigenvalue $\lambda_n = n(n+1)/(2N)$, $n = 1, 2, \dots$ [6] where $P_n^{(a,b)}(x)$ is the Jacobi polynomial and the normalization constant C_n is given by

$$C_n = \sqrt{\frac{(2n+1)(n+1)}{n}}. \quad (\text{B.3.4})$$

on using (22.2.1) of ABRAMOWITZ and STEGUN [1].

Substituting (B.3.2) and (B.3.3) in (B.3.1) and using the orthonormality condition for the eigenfunctions, we find that the time-dependent coefficient $a_m(t)$ is the solution

of the following first order ordinary differential equation,

$$\rho(t) \frac{da_m(t)}{dt} + \frac{m(m+1)a_m(t)}{2\bar{N}} = -\frac{C_m}{m+1} \rho(t) \frac{d\rho(t)}{dt}. \quad (\text{B.3.5})$$

Substituting the normalized eigenfunction $\psi_n(x)$ obtained above in (B.3.2), and using the resulting expression for $v(x, t)$ in (B.2.1), we find that the site frequency spectrum for a large population is given by

$$f(x, t) = \frac{g(x, t)}{x(1-x)} = \theta \frac{\rho(t)}{x} + \theta \sum_{m=1}^{\infty} a_m(t) C_m P_{m-1}^{(1,1)}(1-2x). \quad (\text{B.3.6})$$

Furthermore, using the above expression and (7.391.2) of GRADSHTEYN and RYZHIK [7], the sample SFS is found to be

$$f_{n,i}(t) = \frac{\theta \rho(t)}{i} + \frac{\theta}{n+1} \sum_{m=1}^{\infty} a_m(t) C_m m {}_3F_2(1-m, 2+m, i+1; 2, n+2; 1) \quad (\text{B.3.7})$$

where ${}_3F_2(a_1, a_2, a_3; b_1, b_2; c)$ is the generalized hypergeometric function. Using the moments of $g(x, t)$, the above result has also been obtained by ŽIVKOVIĆ and STEPHAN [5].

For the periodically varying population size, the homogeneous solution of (B.3.5) decays to zero at large times and the inhomogeneous solution is a periodic function with frequency Ω . Therefore, at large times, the time-dependent part $a_m(t)$ can be expanded in a Fourier series as

$$a_m(t) = \sum_{k=-\infty}^{\infty} d_k^{(m)} e^{ik\Omega t}. \quad (\text{B.3.8})$$

Substituting the above equation in (B.3.5), we find that the coefficients $d_k^{(m)}$ obey the following three-term recursion equation,

$$d_k^{(m)} \left(\frac{m(m+1)}{2\bar{N}} + ik\Omega \right) + \frac{\nu\Omega}{2} \left((k-1)d_{k-1}^{(m)} - (k+1)d_{k+1}^{(m)} \right) = \mathcal{D}_k^{(m)} \quad (\text{B.3.9})$$

with $\mathcal{D}_1^{(m)} = \mathcal{D}_{-1}^{(m)} = -\frac{\nu\Omega C_m}{2(m+1)}$, $\mathcal{D}_2^{(m)} = \mathcal{D}_{-2}^{*(m)} = \frac{i\nu^2\Omega C_m}{4(m+1)}$ and $\mathcal{D}_k^{(m)} = 0$, $k \neq \pm 1, \pm 2$ where the subscript $*$ denotes the complex conjugate. It is clear from (B.3.9) that $d_{-k}^{(m)} = d_k^{*(m)}$ which ensures that $a_m(t)$ is real. The above recursion equation does not seem to be exactly solvable but we can obtain approximate expressions for the $d_k^{(m)}$'s when the cycling frequency is small or large relative to the inverse average population size.

Small cycling frequencies: We first consider the parameter regime $\Omega \ll 1/\bar{N}$. On adding (B.3.9) for $k = \pm 1$, we obtain

$$-\Omega \text{Im}(d_1^{(m)}) + \frac{m(m+1)}{2\bar{N}} \text{Re}(d_1^{(m)}) - \nu\Omega \text{Re}(d_2^{(m)}) = \text{Re}(\mathcal{D}_1^{(m)}) . \quad (\text{B.3.10})$$

Since $a_m(t)$ and hence $d_k^{(m)}$'s must vanish with Ω , it follows that the first and third term on the LHS of the above equation are of order Ω^2 or higher. But since the RHS is of order Ω , we have $\text{Re}(d_1^{(m)}) \approx \frac{2\bar{N}}{m(m+1)} \text{Re}(\mathcal{D}_1^{(m)})$. Using this result in (B.3.9) for $k = 0$, we find that

$$d_0^{(m)} \approx -\frac{2\nu^2\bar{N}^2\Omega^2 C_m}{m^2(m+1)^3} , \quad \Omega \ll 1/\bar{N} . \quad (\text{B.3.11})$$

The higher coefficients can also be obtained, and to order $\bar{N}^2\Omega^2$, we get

$$d_1^{(m)} \approx -\frac{C_m\nu\bar{N}\Omega}{m(1+m)^2} + i\frac{C_m\nu(2+\nu^2)\bar{N}^2\Omega^2}{m^2(1+m)^3} \quad (\text{B.3.12})$$

$$d_2^{(m)} \approx \frac{3C_m\nu^2\bar{N}^2\Omega^2}{m^2(1+m)^3} + i\frac{C_m\nu^2\bar{N}\Omega}{2m(1+m)^2} \quad (\text{B.3.13})$$

$$d_3^{(m)} \approx -i\frac{C_m\nu^3\bar{N}^2\Omega^2}{m^2(1+m)^3} \quad (\text{B.3.14})$$

$$d_k^{(m)} \approx 0, k > 3 . \quad (\text{B.3.15})$$

From (B.3.7), the time-averaged sample SFS can be written as

$$\bar{f}_{n,i} = \frac{\theta}{i} + \theta \sum_{m=1}^{\infty} \frac{d_0^{(m)} C_m m_3 F_2(1-m, 2+m, i+1; 2, n+2; 1)}{n+1} \quad (\text{B.3.16})$$

$$\approx \frac{\theta}{i} - \frac{\theta}{n+1} \sum_{m=1}^{\infty} \frac{2\nu^2\bar{N}^2\Omega^2(2m+1)}{m^2(m+1)^2} {}_3F_2(1-m, 2+m, i+1; 2, n+2; 1) \quad (\text{B.3.17})$$

where the last expression is valid for small cycling frequencies. The time-averaged mean heterozygosity can be obtained from the above equation using ${}_3F_2(1-m, 2+m, 2; 2, 4; 1) = \delta_{m,1}$, and is given by

$$\bar{H}^{(neu)} = \theta - \frac{\theta}{2}\nu^2\bar{N}^2\Omega^2. \quad (\text{B.3.18})$$

Large cycling frequencies: For $\Omega \gg 1/\bar{N}$, the recursion equation (B.3.9) for $k \neq 0, \pm 1, \pm 2$ can be approximated by $2ix_k + \nu(x_{k-1} - x_{k+1}) \approx 0$ where $x_k = kd_k^{(m)}$. The solution of this equation is given by $x_k = A_+x_+^k + A_-x_-^k$ where

$$x_{\pm} = i \frac{1 \pm \sqrt{1 - \nu^2}}{\nu}. \quad (\text{B.3.19})$$

In order that the Fourier series (B.3.8) converges, for positive k , we demand $d_k^{(m)} \xrightarrow{k \rightarrow \infty} 0$ which yields the coefficient $A_+ = 0$. Using the solution for $d_k^{(m)}, k \geq 2$ in (B.3.9) for $k = 1, 2$, we can find A_- and d_1 to leading order in $(\bar{N}\Omega)^{-1}$, and finally obtain

$$d_k^{(m)} \approx \frac{imC_m x_-^k}{2\bar{N}\Omega k}, \quad k \geq 2 \quad (\text{B.3.20})$$

$$d_1^{(m)} \approx -\frac{\nu m C_m}{2\bar{N}\Omega} \frac{1}{1 + \sqrt{1 - \nu^2}} + i \frac{\nu C_m}{2(m+1)}. \quad (\text{B.3.21})$$

Using the above result for $d_1^{(m)}$ in (B.3.9) for $k = 0$, we get

$$d_0^{(m)} \approx -\frac{C_m}{m+1}(1 - \sqrt{1 - \nu^2}), \quad \Omega \gg 1/\bar{N}. \quad (\text{B.3.22})$$

Thus, for large cycling frequencies, the time-averaged sample SFS is given by

$$\bar{f}_{n,i} \approx \frac{\theta}{i} - \frac{\theta(1 - \sqrt{1 - \nu^2})}{n+1} \sum_{m=1}^{\infty} (2m+1) {}_3F_2(1-m, 2+m, i+1; 2, n+2; \text{B.3.23})$$

$$= \frac{\theta}{i} \sqrt{1 - \nu^2} \quad (\text{B.3.24})$$

where the last expression follows on using that the sum on the RHS of (B.3.23) equals $(n+1)/i$, as can be verified numerically.

B.4 Nonequilibrium site frequency spectrum for weak selection

Here we develop a perturbation theory for weak selection when both the selection and population size oscillate with the same frequency. In the following discussion, we will write $s(t) = \sharp \zeta(t)$ where $\sharp = \bar{s}$, $\zeta(t) = 1 + \frac{\sigma}{\bar{s}} \sin(\omega t)$ and $\sharp = \sigma$, $\zeta(t) = \sin(\omega t)$ when on-average selection is nonzero and zero, respectively. We begin by writing $v(x, t) = v_0(x, t) + \bar{N} \sharp v_1(x, t)$ and substitute it in (B.2.2). Keeping terms up to linear order in $\bar{N} \sharp$, we find that v_0 obeys (B.3.1) and v_1 is a solution of the following equation,

$$\frac{\partial v_1(x, t)}{\partial t} = \frac{x(1-x)}{2N(t)} \frac{\partial^2 v_1(x, t)}{\partial x^2} + r_1(x, t) \quad (\text{B.4.1})$$

where,

$$\begin{aligned} r_1(x, t) &= -\zeta(t)x(1-x) \frac{\partial}{\partial x} [(x + h(1-2x))v_0(x, t)] \\ &\quad - \rho(t)\zeta(t)x(1-x)[1 - 2x + h(4x - 3)]. \end{aligned} \quad (\text{B.4.2})$$

As in Appendix B.3, we expand v_1 and r_1 as a linear combination of the eigenfunctions $\psi_m(x)$,

$$v_1(x, t) = \sum_{m=1}^{\infty} A_m(t) \psi_m(x) \quad (\text{B.4.3})$$

$$r_1(x, t) = \sum_{m=1}^{\infty} B_m(t) \psi_m(x) \quad (\text{B.4.4})$$

where $A_m(t)$ obeys the following ordinary differential equation,

$$\rho(t) \frac{dA_m(t)}{dt} + \frac{m(m+1)}{2\bar{N}} A_m(t) = \rho(t) \frac{B_m(t)}{\bar{N}} \quad (\text{B.4.5})$$

and its long time solution can be expanded in a Fourier series as $A_m(t) = \sum_{k=-\infty}^{\infty} D_k^{(m)} e^{ik\omega t}$. The coefficient $B_m(t) = \int_0^1 dx r_1(x, t) \psi_m(x) w(x)$ can be found using the orthonormality condition for the eigenfunctions (see Appendix B.3). But as the integrals involved in $B_m(t)$ do not seem to be known for arbitrary m , here we will focus on $B_1(t)$ which is related to mean heterozygosity and given by

$$\begin{aligned} B_1(t) &= -\frac{C_1^2(1-2h)\zeta(t)a_1(t)}{60} - \frac{C_1C_3(1-2h)\zeta(t)a_3(t)}{70} \\ &+ \frac{C_1C_2\zeta(t)a_2(t)}{30} + \frac{C_1h\zeta(t)\rho(t)}{6} \end{aligned} \quad (\text{B.4.6})$$

where $a_m(t)$ is defined in (B.3.8). The sample SFS averaged over a period can be written as

$$\bar{f}_{n,i} = \bar{f}_{n,i}^{(neu)} + \frac{\bar{N}\ddagger\theta}{n+1} \sum_{m=1}^{\infty} D_0^{(m)} C_m m_3 F_2(1-m, 2+m, i+1; 2, n+2; 1) \quad (\text{B.4.7})$$

where $\bar{f}_{n,i}^{(neu)}$ is given by (B.3.17). In particular, for the time-averaged heterozygosity, we have

$$\bar{H} = \bar{H}^{(neu)} + \frac{\theta\bar{N}\ddagger C_1}{3} D_0^{(1)} \quad (\text{B.4.8})$$

where $\bar{H}^{(neu)}$ is given by (B.3.18) for small cycling frequencies and $\bar{f}_{2,1}$ in (B.3.24) for large cycling frequencies.

To demonstrate the weak selection calculation, below we consider the case when $\ddagger = \bar{s}$, $\zeta(t) = 1$, that is, selection is constant. We find that the coefficients $D_k^{(1)}$ obey the

following recursion equation,

$$D_k^{(1)} \left(\frac{1}{2} + ik\bar{N}\Omega \right) + \frac{\nu\bar{N}\Omega}{2} \left((k-1)D_{k-1}^{(1)} - (k+1)D_{k+1}^{(1)} \right) = \mathcal{D}_k^{(1)} \quad (\text{B.4.9})$$

where

$$\begin{aligned} \mathcal{D}_k^{(1)} &= \frac{C_1 h}{6} \left(\left(\frac{\nu^2}{2} + 1 \right) \delta_{k,0} + i\nu\delta_{k,-1} - i\nu\delta_{k,1} - \frac{\nu^2}{4}\delta_{k,-2} - \frac{\nu^2}{4}\delta_{k,2} \right) \\ &- \frac{C_1^2}{60} (1-2h) \left(\frac{i\nu}{2}d_{k+1}^{(1)} + d_k^{(1)} - \frac{i\nu}{2}d_{k-1}^{(1)} \right) \\ &+ \frac{C_1 C_2}{30} \left(\frac{i\nu}{2}d_{k+1}^{(2)} + d_k^{(2)} - \frac{i\nu}{2}d_{k-1}^{(2)} \right) \\ &- \frac{C_1 C_3}{70} (1-2h) \left(\frac{i\nu}{2}d_{k+1}^{(3)} + d_k^{(3)} - \frac{i\nu}{2}d_{k-1}^{(3)} \right). \end{aligned} \quad (\text{B.4.10})$$

For small cycling frequencies, using an argument similar to that given in Appendix B.3 for $d_0^{(m)}$, we get $\text{Re}(D_k^{(1)}) = \text{Re}(\mathcal{D}_k^{(1)})$; on using $d_k^{(m)}$ given by (B.3.11)-(B.3.15), we find that $\text{Re}(D_1^{(1)}) \sim \mathcal{O}(\bar{N}\Omega)$. But as $\mathcal{D}_0^{(1)}$ is a constant in $\bar{N}\Omega$, we finally obtain

$$D_0^{(1)} = \mathcal{D}_0^{(1)} + \nu \text{Re}(D_1^{(1)})\bar{N}\Omega \approx \frac{C_1 h(2 + \nu^2)}{12}. \quad (\text{B.4.11})$$

For large cycling frequencies, as in Appendix B.3, $D_k^{(1)} = A_- x_-^k / k$, $k \geq 3$. Then using the properties of $d_k^{(m)}$ for $|k| < 3$ described above, we find

$$D_0^{(1)} = \frac{h(1 - \nu^2)}{\sqrt{6}}. \quad (\text{B.4.12})$$

B.5 Mean heterozygosity for co-dominant, on-average neutral mutant

For co-dominant allele in an on-average neutral environment with constant population size, the time-averaged mean heterozygosity is equal to θ for any cycling frequency and

strength of selection. For $h = 1/2$, $\bar{s} = \nu = 0$, (3.6) reduces to

$$\frac{\partial g(x, t)}{\partial t} = -\frac{\sigma \sin(\omega t) x(1-x)}{2} \frac{\partial g(x, t)}{\partial x} + \frac{x(1-x)}{2\bar{N}} \frac{\partial^2 g(x, t)}{\partial x^2}. \quad (\text{B.5.1})$$

We first note that the above equation has the following symmetry:

$$g(x, t) = \theta - g\left(1-x, t + \frac{\pi}{\omega}\right) \quad (\text{B.5.2})$$

where the boundary conditions, $g(0, t) = \theta$, $g(1, t) = 0$ have been used. Then using the definition $H(t) = 2 \int_0^1 dx g(x, t)$ and on integrating by parts, (B.5.1) yields

$$\frac{dH}{dt} = \sigma \sin(\omega t) \int_0^1 dx (1-2x) g(x, t) + \frac{\theta}{\bar{N}} - \frac{H(t)}{\bar{N}}. \quad (\text{B.5.3})$$

Integrating both sides over the period $2\pi/\omega$, we find that the LHS vanishes since $H(t)$ is a periodic function, and the first term on the RHS is also zero due to the symmetry property (B.5.2); we thus finally arrive at $\bar{H} = \theta$.

B.6 Nonequilibrium site frequency spectrum for strong selection

For on-average neutral selection and equal cycling frequencies for selection and population size, we write (3.6) as

$$\begin{aligned} \frac{\partial g(x, \psi)}{\partial \psi} &= -\frac{\bar{N}\sigma \sin \psi}{\bar{N}\omega} x(1-x) \frac{\partial}{\partial x} [(x + h(1-2x))g(x, \psi)] \\ &+ \frac{x(1-x)}{2\bar{N}\omega\rho(t)} \frac{\partial^2 g(x, \psi)}{\partial x^2} \end{aligned} \quad (\text{B.6.1})$$

where $\psi = \omega t$ and expand $g(x, t)$ in powers of $(\bar{N}\omega)^{-1}$. If $\bar{N}\sigma \ll 1$ and $\bar{N}\omega \gg 1$, the first term on the RHS must be neglected in comparison to the second term on the RHS, and therefore the leading order result for $g(x, t)$ is given by the neutral population already discussed in Appendix B.3 and the subleading correction due to selection in Appendix B.4. For strong selection, if the LHS is also ignored, we get $g(x, t)$ is a constant in allele frequency which is not true.

Here, we are therefore interested in the regime where $\bar{N}\sigma \gg 1, \bar{N}\omega \gg 1$ with σ/ω finite. Then for rapidly changing environment and strong selection, the second term on the RHS can be ignored resulting in a first order partial differential equation for $g(x, t)$. A numerical analysis of the resulting approximate equation with boundary conditions (3.7) is found to be in reasonable agreement with that for the exact equation. For simplicity, we consider the case when $h = 1/2$ and solve the first order partial differential equation by the method of characteristics.

For this purpose, we consider a change of variables as $\xi = \xi(x, \psi)$ and $\eta = \eta(x, \psi)$, and choose $\xi = \psi$ and $\eta = \text{constant}$ so that

$$\frac{\partial \eta}{\partial \psi} + \frac{\sigma \sin \psi}{2\omega} \frac{x(1-x)}{\partial x} \frac{\partial \eta}{\partial x} = 0 \quad (\text{B.6.2})$$

which leads to the characteristic equation,

$$\frac{dx}{d\psi} = \frac{\sigma \sin \psi}{2\omega} x(1-x) \quad (\text{B.6.3})$$

and the characteristic curve to be,

$$\eta = \ln \left(\frac{x}{1-x} \right) + \frac{\sigma}{2\omega} \cos \psi . \quad (\text{B.6.4})$$

We then obtain $\partial g/\partial \xi = 0$ which implies that $g(\xi, \eta) = \mathcal{G}(\eta)$ where \mathcal{G} should be found using the boundary conditions. We are, however, unable to implement the boundary conditions (3.7) in this solution. If instead we determine \mathcal{G} using the initial condition, $g(x, 0) = \theta(1 - x)$, we get

$$g(x, t) = \frac{\theta}{1 + \frac{x}{1-x} e^{\frac{\sigma(\cos(\omega t) - 1)}{2\omega}}} \quad (\text{B.6.5})$$

which shows that the $g(x, t)$ has phase difference of $\pi/2$ from selection, but this result is independent of ν and Ω which is inconsistent with the numerical results for $g(x, t)$.

Bibliography

- [1] ABRAMOWITZ, M. and I. A. STEGUN, 1964 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover.
- [2] MATHEWS, J. and R. L. WALKER, 1970 *Mathematical methods of physics*. Pearson Education Limited.
- [3] KIMURA, M., 1957 Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**: 882–901.
- [4] EVANS, S. N., Y. SHVETS, and M. SLATKIN, 2007 Non-equilibrium theory of the allele frequency spectrum. *Theor Popul Biol.* **71**: 109–119.
- [5] ŽIVKOVIĆ, D. and W. STEPHAN, 2011 Analytical results on the neutral non-equilibrium allele frequency spectrum based on diffusion theory. *Theor. Popul. Biol.* **79**: 184–191.
- [6] KIMURA, M., 1964 Diffusion models in population genetics. *J. Appl. Prob.* **1**: 177–232.
- [7] GRADSHTEYN, I. S. and I. M. RYZHIK, 2007 *Table of Integrals, Series, and Products*. Academic Press, New York.

Appendix C

C.1 Diffusion Theory

The eventual fixation probability, $P_f(x, t_0)$, that the mutation allele with initial frequency x at time t_0 fixes in the population obeys the following backward Kolmogorov equation, and depends upon on the arrival time of the mutation (t_0) [1–3] as

$$-\frac{\partial P_f(x, t_0)}{\partial t_0} = s(t_0)g(x)\frac{\partial P_f(x, t_0)}{\partial x} + \frac{x(1-x)}{2N_e}\frac{\partial^2 P_f(x, t_0)}{\partial x^2} \quad (\text{C.1.1})$$

When the mutation is on-average neutral, and environmental change is slow, fixation probability is approximated using perturbation theory as $P_f \approx P_0 + N_e\omega P_1$. The P_0 and P_1 are, respectively, the eventual fixation probability in the static environment and the leading order deviation in ω in the fixation probability, which obey the following partial differential equations [3],

$$0 = s(\theta_0)g(x)\frac{\partial P_0(x, \theta_0)}{\partial x} + \frac{x(1-x)}{2N_e}\frac{\partial^2 P_0(x, \theta_0)}{\partial x^2} \quad (\text{C.1.2})$$

$$-\frac{\partial P_0(x, \theta_0)}{\partial \theta_0} = N_e s(\theta_0)g(x)\frac{\partial P_1(x, \theta_0)}{\partial x} + \frac{x(1-x)}{2}\frac{\partial^2 P_1(x, \theta_0)}{\partial x^2} \quad (\text{C.1.3})$$

with the boundary conditions as $P_0(0, \theta_0) = 0$, $P_0(1, \theta_0) = 1$, $P_1(0, \theta_0) = 0$, and $P_1(1, \theta_0) = 0$. The $\theta_0 = \omega t_0$ is the initial phase at which mutation arises in the population. Equation (C.1.3) is similar to Eqn. (B.3) in [3] and can be solved using the

integrating factor method for the arbitrary inbreeding coefficient. Using (C.1.2), the fixation probability in static environment is given by

$$P_0 = \frac{1 - \frac{ae^{-A(x^2+2ax)}}{x+a}}{1 - \frac{ae^{-A(1+2a)}}{1+a}} \quad (\text{C.1.4})$$

where

$$A = |N_e\sigma \sin \theta_0| (1 - 2h - f(1 - 2h)) \quad (\text{C.1.5})$$

$$a = \frac{h + f(1 - h)}{1 - 2h - f(1 - 2h)} \quad (\text{C.1.6})$$

For strong selection $|N_e\sigma| \gg 1$, and the initial mutation frequency $x \rightarrow 0$, the P_0 , and P_1 are given as

$$P_0 = 2x|N_e\sigma \sin \theta_0| (1 - h(1 - f)) e^{-|N_e\sigma \sin \theta_0|(1+f)} \quad (\text{C.1.7})$$

$$P_1 = \frac{2x|N_e\sigma \cos \theta_0|(1 + f)(1 - h + fh)}{f + h(1 - f)} \times e^{-|N_e\sigma \sin \theta_0|(1+f)} \ln(|N_e\sigma \sin \theta_0| (h + f(1 - h))) \quad (\text{C.1.8})$$

Unlike in a randomly mating population, the deviation in fixation probability is weakly dependent on the dominance parameter when the inbreeding coefficient is high. Since conditional mean fixation time depends upon the fixation probability as discussed in the main text, the deleterious mutations, be it deleterious or recessive, are weakly affected in the slowly changing environment for highly inbred individuals, which is not the same in randomly mating population.

The time inhomogeneous Kolmogorov backward equation used for randomly mating populations in KAUSHIK and JAIN [4] can be extended easily for inbreeding populations. We use the same perturbation theory used in fixation probability for the conditional mean fixation time when mutation is on-average neutral, by writing $\bar{T}_c = \bar{T}_{0,c} + N_e\omega \bar{T}_{1,c}$

in (4.1). The $\bar{T}_{0,c}$, and $\bar{T}_{1,c}$ follows [4]

$$-N_e P_0 = N_e s(\theta_0) g(x) \frac{\partial T_{0,c}(x, \theta_0)}{\partial x} + \frac{x(1-x)}{2} \frac{\partial^2 T_{0,c}(x, \theta_0)}{\partial x^2} \quad (\text{C.1.9})$$

$$-\frac{\partial T_{0,c}(x, \theta_0)}{\partial \theta_0} - N_e P_1 = N_e s(\theta_0) g(x) \frac{\partial T_{1,c}(x, \theta_0)}{\partial x} + \frac{x(1-x)}{2} \frac{\partial^2 T_{1,c}(x, \theta_0)}{\partial x^2} \quad (\text{C.1.10})$$

C.2 Semi-deterministic approximation

We follow the same procedure as in MARTIN and LAMBERT [5], and KAUSHIK and JAIN [4] to calculate the fixation time for the beneficial mutations in an inbreeding population. Using (4.2), the fixation time distribution can be straightforwardly generalized for inbreeding population as

$$P(T_c) = \left(\frac{-s_a}{s_b}\right) \int_0^\infty \frac{dq B^2 e^{(s_b/s_a)\tau(T_c)} \exp(-Bq)}{2N_e \Gamma n} \left(\gamma_a q_2^{-s_a/s_b}\right)^n e^{-\gamma_a q^{-s_a/s_b}} \quad (\text{C.2.1})$$

where $\tau(T_c) = \int_0^{T_c} s(t') dt'$, and

$$B = \frac{2N_e}{\int_0^{T_c} dt e^{s_b \int_0^t s(t') dt'}} \quad (\text{C.2.2})$$

$$\gamma_a = 2N_e P_f e^{-A} \quad (\text{C.2.3})$$

where $P_f = s_a \sigma \sin \theta_0 / (1 + f)$ is the eventual fixation probability for a single mutation.

For the static environment the fixation time is given by [5]

$$\bar{T}_{0,c} = \frac{\left((c_1 + 1)^k - 1\right) \left(\log[\gamma_a^{s_b} (2N_e s_0)^{s_a}] + \gamma_{EM}(s_a + s_b)\right) - c_1 k s_b {}_2F_1(1 - k, 1, 2, -c_1)}{c_k s_0 s_a} \quad (\text{C.2.4})$$

where $s_0 = s_b s(0)$, k and n are the initial and the established copies of mutation allele respectively. The ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function [6]. The γ_a , c_1 , and

c_k are given by

$$c_1 = \frac{P_f}{1 - P_f} \quad (\text{C.2.5})$$

$$c_k = \frac{1 - (1 - P_f)^k}{(1 - P_f)^k} \quad (\text{C.2.6})$$

$$\gamma_a = 2N_e P_f e^{-A} \quad (\text{C.2.7})$$

$$A = \left(\frac{s_a - s_b}{s_b} \right) \ln \left(\frac{s_a}{s_b} \right) \quad (\text{C.2.8})$$

For a single mutation ($k = 1$), the (C.2.4) matches exactly with Eqn. (17) in GLÉMIN [7] where approximate expression of the conditional mean fixation time is obtained using diffusion theory. The deviation in the conditional mean fixation time in the slowly changing environment, $\bar{T}_{1,c}$ is then given by,

$$\bar{T}_{1,c} = -\frac{\sigma \cos \theta_0}{2s(0)^2 s_a^2} \sum_{n=1}^k \binom{k}{n} \frac{c_1}{c_k} \left[\left(\ln \left((2N_e s_b s(0))^{(s_a/s_b)} \gamma_a - \psi^0(n) \right) \right)^2 - \psi^1(n) \right] \quad (\text{C.2.9})$$

where, $\psi^n(x)$ is the Polygamma function, given by the $(n + 1)$ st derivative of the logarithm of the gamma function ($\Gamma(x)$) [6].

Bibliography

- [1] EWENS, W., 2004 *Mathematical Population Genetics*. Springer, Berlin.
- [2] UECKER, H. and J. HERMISSON, 2011 On the fixation process of a beneficial mutation in a variable environment. *Genetics* **188**: 915–930.
- [3] DEVI, A. and K. JAIN, 2020 The impact of dominance on adaptation in changing environments. *Genetics* **216**: 227–240.
- [4] KAUSHIK, S. and K. JAIN, 2021 Time to fixation in changing environments. *Genetics* **219**: iyab 148.
- [5] MARTIN, G. and A. LAMBERT, 2015 A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theo. Pop. Biol.* **101**: 40–46.
- [6] ABRAMOWITZ, M. and I. A. STEGUN, 1964 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover.
- [7] GLÉMIN, S., 2012 Extinction and fixation times with dominance and inbreeding. *Theo. Pop. Biol.* **81**: 310–316.

