UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

OSMAR FREITAS DA SILVA JÚNIOR

POPULATION PERSISTENCE AND DIVERGENCE DRIVEN BY ENVIRONMENTAL CHANGES

Recife

2025

OSMAR FREITAS DA SILVA JÚNIOR

POPULATION PERSISTENCE AND DIVERGENCE DRIVEN BY ENVIRONMENTAL CHANGES

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Física.

Área de Concentração: Física Teórica e Computacional

Orientador: Paulo Roberto de Araújo Campos

Coorientadora: Sabrina Borges Lino Araújo

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Silva Junior, Osmar Freitas da.
Population Persistence and Divergence Driven by Environmental Changes / Osmar Freitas da Silva Junior. - Recife, 2025.
143f.: il.

Tese (Doutorado) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Programa de Pós-Graduação Física, 2025.
Orientação: Paulo Roberto de Araújo Campos.
Coorientação: Sabrina Borges Lino Araújo.
Inclui referências e apêndices.

1. Evolutionary dynamics; 2. Environmental changes; 3. Persistence; 4. Speciation. I. Campos, Paulo Roberto de Araújo. II. Araújo, Sabrina Borges Lino. III. Título.

UFPE-Biblioteca Central

OSMAR FREITAS DA SILVA JÚNIOR

POPULATION PERSISTENCE AND DIVERGENCE DRIVEN BY ENVIRONMENTAL CHANGES

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Física.

Área de Concentração: Física Teórica e Computacional

Data de aprovação: 06/06/2025.

BANCA EXAMINADORA

_____

Prof. Dr. Paulo Roberto de Araujo Campos
Orientador
Universidade Federal de Pernambuco

_____

Prof. Dr. Pedro Valadão Carelli
Examinador Interno
Universidade Federal de Pernambuco

_____

Profa. Dra. Ana Carolina Oliveira de Queiroz Carnaval
Examinadora Externa
City University of New York

_____

Profa. Dra. Celia Beatriz Anteneodo de Porto
Examinadora Externa
Pontifícia Universidade Católica do Rio de Janeiro

_____

Prof. Dr. José Alexandre Felizola Diniz Filho
Examinador Externo
Universidade Federal de Goiás

ACKNOWLEDGEMENTS

Agradeço à meu pai por ter me ensinado que pra tudo na vida se dá um jeito, à minha mãe por me mostrar que quanto mais se vive mais a gente aprende, e aos meus irmãos, sem os quais me faltariam cúmplices nessa aventura que é fazer parte da nossa família.

Agradeço aos meu orientadores por me orientarem tanto na academia quanto na vida. Em particular, ao Prof. Paulo Campos pela oportunidade que me foi dada e pela paciência ao me orientar em uma área na qual sempre tive muito fascínio, porém poquíssima experiência. Também à Profa. Sabrina Araújo por ter topado fazer parte de minha formação, por me abrir tantas portas e por me aproximar da biologia e, especialmente, dos biólogos.

Agradeço também aos professores que me ensinaram muito além da sala de aula, à todo o grupo de Interações Biológicas, à todo o grupo do LTEEI e aos colegas do DF com quem partilhei essa caminhada na academia. Em particular, ao meu colega Diego por dividirmos as dores e o fascínio de ser cientista.

Agradeço também à FACEPE, sem a qual esta pesquisa não se realizaria; à estrutura fornecida pela UFPE, pelo Departamento de Física e pelo laboratório de Dinâmica Evolucionária.

E, por fim, agradeço à estocaticidade dos eventos históricos que fizeram com que hoje eu tenha, sempre ao meu lado, minhas companheiras de vida: Ravena e Suricat.

ABSTRACT

The evolution of life accompanies the dynamics of our planet Earth. From continental drift to glacial ages, geological events have shaped the Earth's surface and climate, influencing the evolutionary history of populations. Much more recently, the human species became capable of great modifications on the planet as well. Increasing global temperature, altering the course of rivers and an ever-expanding usage of the land are some examples of human activities that cause great pressure upon other species. In this work, we use stochastic evolutionary models to study two aspects of evolution driven by environmental changes: population persistence and divergence. Persistence is the ability of a population to survive and adapt to drastic environmental changes, while divergence involves the initial differentiation process leading to new species. Our first approach on persistence deals with evolutionary rescue, a process where adaptive evolution reverts the fate of a population doomed to extinction. With a combination of analytical and simulation methods, we show the relation between the probability of extinction and the intensity of stress level upon the population. We characterize the parallelism of the evolutionary response and its relation with demographic and genetic factors of density regulation and epistasis, respectively. In the sequel, we pass to study the divergence process at the speciation level in a population divided into a two-deme system where geographic environmental events determine the possibility of exchanging individuals. In this neutral selective model, only the migration rate among patches and the intermittent periods of connectance and isolation can promote population divergence leading to speciation. We register the dynamics through a phylogenetic speciation tree and show how the rate and mode of speciation are related to the time the population spent in isolation. At last, we study speciation in a metapopulation model under environmental fluctuations, where both selection and migration are present. We carry out a statistical analysis over different sets of parameters in order to examine the resulting diversity. We then compare how different scenarios of environmental variations - including the magnitude and frequency of these changes - affect the persistence and divergence of the metapopulation. We find that the pattern of speciation is only dependent on the net effect of the environmental disturbances, despite the rate at which the events occur. We hope that these models can give us insight into the processes affecting life's diversity and resilience.

Keywords: Evolutionary dynamics. Environmental changes. Persistence. Speciation.

RESUMO

A evolução da vida acompanha a dinâmica do nosso planeta Terra. Da deriva continental às eras glaciais, eventos geológicos moldaram a superfície e o clima da Terra, influenciando a história evolutiva das populações. Muito mais recentemente, a espécie humana também se tornou capaz de grandes modificações no planeta. O aumento da temperatura global, a alteração do curso dos rios e o uso cada vez maior da terra são alguns exemplos de atividades humanas que exercem grande pressão sobre outras espécies. Neste trabalho, utilizamos modelos evolutivos estocásticos para estudar dois aspectos da evolução impulsionados por mudanças ambientais: persistência e divergência populacional. Persistência é a capacidade de uma população sobreviver e se adaptar a mudanças ambientais drásticas, enquanto divergência envolve o processo inicial de diferenciação que leva ao surgimento de novas espécies. Nossa primeira abordagem sobre persistência trata do resgate evolutivo, um processo em que a adaptação reverte o destino de uma população fadada à extinção. Combinando métodos analíticos e simulação, mostramos a relação entre a probabilidade de extinção e o nível de estresse sobre a população. Investigamos a resposta evolutiva e sua relação com fatores demográficos e genéticos, envolvendo crescimento populacional e epistasia, respectivamente. Na sequência, passamos a estudar o processo de divergência à nível de espécies em uma população dividida em duas ilhas, onde eventos ambientais geográficos determinam a possibilidade de troca de indivíduos. Neste modelo neutro, apenas a taxa de migração entre as ilhas e os períodos intermitentes de conectância e isolamento podem promover a divergência populacional, levando à especiação. Registramos a dinâmica por meio de uma filogenia e mostramos como a taxa e o modo de especiação estão relacionados ao tempo de isolamento. Por fim, estudamos a especiação em um modelo de metapopulação sob flutuações ambientais, onde tanto a seleção quanto a migração estão presentes. Realizamos uma análise estatística sobre diferentes conjuntos de parâmetros para examinar a diversidade resultante. Em seguida, comparamos como diferentes cenários de variações ambientais – incluindo a magnitude e a frequência dessas mudanças – afetam a persistência e a divergência da metapopulação. Descobrimos que o padrão de especiação depende apenas do efeito líquido das perturbações ambientais, independentemente da taxa em que os eventos ocorrem. Esperamos que a discussão deles possa nos dar uma visão sobre os processos que afetam a diversidade e a resiliência da vida.

Palavras-chaves: Dinâmica evolucionária. Mudanças ambientais. Persitência. Especiação.

LIST OF FIGURES

Figure 1 – Simple distinction between independent and dependent density growth, represented by an exponential and logistic growth, respectively. 21

Figure 2 – Simple representation of fitness landscapes. (A) In continuum phenotypic landscapes, population can explore all posible combinations of traits. Red arrows represent a "trajectory" of a population. As selection increases the frequency of fitter individuals, adaptation is an walking towards the peak. (B) Discrete genotypic landscape for when only 4 mutations are responsible for the adaptation process. Arrows represent single mutation transitions, where the red ones are transitions to genotypes of higher fitness. 28

Figure 3 – Hypercube representation for L=3 and L=4. Only in the first case, fitness is assigned to each vertex in parenthesis. Arrows heads represent transitions to fitter states, while the circles outline a state that maximizes fitness - a fitness peak. 29

Figure 4 – Mutations arise at short time scales but are rapidly purged by genetic drift. With a probability proportional to s, mutations can reach a threshold frequency to survive drift and, once done, they can reach fixation at time t_{fix}. 31

Figure 5 – Representation of the discussed regimes in asexual and sexual populations. If *de novo* mutations establishes simultaneously, they compete for fixation. In contrast with clonal interference, we see how recombination can decreases the competing aspect by assimilating both genotypes in the same descendant. 32

Figure 6 – Signature of evolutionary rescue. Once a stressful enough environmental change happens, population declines and can only restore positive growth through adaptive evolution. Traced line shows the case for when the mutation is already present, and solid line for *de novo* mutations (note the failed attempts to survive genetic drift). 34

Figure 7 – Conceptual visualization of the Tree of Life, showing ancient life's differentiation into three great domains. The subdivision into further groups shows the variety of life forms in each domain. The evolutionary history within each of these groups (e.g. the animal kingdom) can also be represented by their corresponding tree. 35

Figure 8 – Visualization of the differentiation between individuals of a population. The arrows in the upper row indicate that reproduction is possible. Population diverges, acquiring distinct phenotypes and/or genotypes, leading to speciation. 37

Figure 9 – Representation of two distinct paths with the same initial and ending points. The Hamming distance h is measured for each point $\sigma^A \in \phi^A$ to every point $\sigma^B \in \phi^B$, and vice-versa. The shortest measure is stored. 47

Figure 10 – Three-dimensional Fisher's geometric model, $n = 3$. The wild-type population is at a distance $d = \Delta$ from the phenotypic optimum. The sphere of radius r delimits the domain of rescue, $W(u, \theta) = 1$. For a given angle θ, mutations of magnitude between u_- and u_+ are rescue mutations. Owing to the symmetry, all mutation vectors lying on the red-shaded conical surface are rescue mutations. 52

Figure 11 – Extinction probability as a function of the stress level δ for the phenotypic FGM. The plot compares the simulation results (data points) with analytical predictions (lines). In the left panel, the simulation data is contrasted with the analytical prediction in Eq. (2.14), whereas in the right panel, the simulation data is plotted alongside upper and lower bound estimates of the extinction probability for a population under density regulation, as provided by Eqs. (2.14) and (2.16). The parameter values are carrying capacity $K = 10000$, number of traits $n = 5$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. 53

Figure 12 – Extinction probability as a function of the stress level δ for the genotypic FGM. The plot compares the simulation results (data points) with the analytical predictions (lines), given by Eq. (2.27). The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, number of traits $n = 5$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. 55

Figure 13 – Extinction probability as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. The vertical line delimits non-extinction and rescue/extinction domains. The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points. 56

Figure 14 – Parallelism and mean path divergence as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. From the green vertical line onwards, the population is in a regime of rescue/extinction. The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points. 58

Figure 15 – Illustration of the modelling of evolutionary rescue. In this illustration, the wild-type population has phenotype f by randomly picking a sequence in the genotype space. The optimum is placed a distance Δ from f, such that the fitness of the wild-type population equals $1 - \delta$. The interval (u_-, u_+) delimits the phenotypic domain in which the population can restore a positive growth rate $(W > 1)$. Superimposed on the illustration is a drawing of the phenotype distribution, which is a Gaussian distribution. . . 63

Figure 16 – Incongruence and the fraction of potential rescue mutations as a function of the stress level. Curves denote different degrees of epistasis K. The thick solid line is the analytical result for the uncorrelated case, $K = 11$, obtained from Eq. (2.23). Error bars are smaller than points. Parameters for both cases are $L = 12$, $W_{max} = 1.5$, and $b = 6$. 64

Figure 17 – The distribution of phenotypic values over the entire genotype sequence. The epistatic parameter value is K = 0 (upper panels) and K = 11 (lower panels). In the left panel, the distribution is obtained from a single sample of the phenotypic landscape, whereas in the right panel, the distribution is taken from an ensemble of 1, 000 samples of the phenotypic landscape . . 67

Figure 18 – Probability of extinction for the uncorrelated landscape, $K = 11$, as a function of the stress level for the (left panel) phenotypic and (right panel) genotypic versions of the model. Curves represent distinct values of mutation probability μ. Solid lines are the analytical approximations, Eqs. (2.25) (phenotypic model) and (2.27) (genotypic model). Error bars are smaller than the points. Parameter values are $N_0 = 10^4$, $W_{max} = 1.5$, $b = 6$, and genome size $L = 12$ (genotypic model). 68

Figure 19 – Probability of extinction as a function of the stress level for distinct values of the epistatic parameter K. Error bars are smaller than points. Parameter values are $L = 12$, $\mu = 1 \times 10^{-4}$, $N_0 = 10^4$, $W_{max} = 1.5$, and $b = 6$. . . . 70

Figure 20 – Probability of extinction as a function of the mutation probability. Curves represent different stress levels δ. Error bars are smaller than points. Solid lines are analytical approximations for the full uncorrelated case ($K = 11$), Eq. (2.27). Parameter values are $L = 12$, $N_0 = 10^4$, $W_{max} = 1.5$, and $b = 6$. The epistatic parameter K is indicated in the subtitle of the panels. 71

Figure 21 – Example results from Derria-Higgs model on a two-deme population, showing how the total species richness of both islands N_T changes with migration rate ε and genome size G. 78

Figure 22 – Example of the model evolution under constant migration. There are three plots: the upper and middle panels correspond to the complete and extant phylogenies, respectively. The plot at the bottom illustrates species distribution abundance in each patch over time. In each speciation event, the most abundant species maintains its ancestral color. 79

Figure 23 – Exemplification of two topological features of phylogenies that carry information about the diversification process. A) The distribution of ramifications among branches - the phylogeny balance. B) The change in speciation rate over time. 80

Figure 24 – A sample tree exhibiting the necessary and sufficient parameters for the calculus of both topological metrics. The balance (J) requires the number of leaves ℓ_i and ramifications η_i for each node i. Acceleration of the speciation rate (α) requires the distance between each internal node g_i. 82

Figure 25 – A) Sea-level oscillation over the previous 800 thousand years obtained from Spratt2016. Horizontal dashed lines represent the thresholds h, in which the patches isolate when the sea level is above them (highlighted by the solid black lines). For example, despite seabed depth, the red dots mark the moments when the patches are fully connected and isolated, respectively. B) The relation between the time spent in isolation and seabed depth is approximately linear. 84

Figure 26 – Schematic overview of the study. (A) The model considers an initial population occupying two patches. At each model iteration, the individuals can migrate and reproduce. We evaluate a constant and an intermittent (as illustrated) migration. Reproduction promotes population variability due to recombination and mutation. (B) The model outputs allow us to identify species distribution as well as the complete (including extinctions) and extant (without extinction) phylogenies. 85

Figure 27 – Examples of the model evolution under constant (top) and intermittent (bottom) migration; as well as low (left) and high (right) migration. For each example, there are three plots: upper and middle panels correspond to the complete and extant phylogenies, respectively. The plot at the bottom illustrates species distribution abundance in each patch over time. In each speciation event, the most abundant species maintains its previous color. The horizontal black line in the lower plot indicates when patches are isolated. In all plots, the most abundant species keep the ancestral color. The values of migration probability and the percentage time in isolation ($t_h \approx |h|/100$) are indicated in the figure headings. 87

Figure 28 – Number of speciation events, species richness, β diversity, and species asymmetry distribution as a function of the mean migration rate ($\bar{\varepsilon}$) for different intermittent regimes (colors). The calculus of all metrics is made only by the end of the simulation ($T = 2000$). The solid curves represent the average over 50 replications, and the shadow area the confidence interval of 90%. 88

Figure 29 – Macroevolutionary patterns (phylogeny balance, speciation acceleration, and age) for complete phylogenies versus the mean migration rate($\bar{\varepsilon}$) for different intermittent regimes (colors). The solid curves represent the average value of the metrics and shadow areas indicate the confidence interval of 90%. Following the α definition, the dashed red lines indicate the interval in which the results do not differ from random branching models. 90

Figure 30 – Evolution of phylogeny balance and acceleration of speciation for combinations of isolation $t_h \approx \{80, 20\}$ and migration rate $\bar{\varepsilon} = 0.08$. Solid lines represent the average of 50 replications, while dotted ones delimit the confidence interval of 90%. Following the α definition, the dashed red lines indicate the interval in which the results do not differ from random branching models. Background gray bands indicate when patches are isolated for each t_h. 92

Figure 31 – Correlation tables between the analyzed metrics for the four scenarios of isolation (t_h indicated in the figure). The correlation was obtained comprising the whole interval of migration values $\bar{\varepsilon} \in (0.002, 0.08)$, with values reported at the end of simulations $(T = 2000)$ for all complete phylogenies. 93

Figure 32 – Representation of a metapopulation model, consisting of connected patches and the individuals that inhabit them. Colors allude to the differences in population composition. In this particular case of a homogeneous metapopulation, all the patches are mutually connected. 98

Figure 33 – Example of the one-dimensional Gaussian viability function for different widths α_r (colours). On the trait-axis, we have the individual phenotype at point r and the optimum phenotype of the patch it inhabits h. Selective pressure goes with $1/\alpha_r$, thus the individual has zero probability of survival for the blue curve, low probability for the green, and almost one for the red. In the limit $\alpha_r \to \infty$ one has $V \to 1$, thus individuals are always viable as there are effectively no selection mechanisms. 99

Figure 34 – Scheme of the model dynamics. Individuals are distributed among patches, illustrated by green, red, and yellow colors. At each generation, individuals go through Selection, Reproduction, and Migration. Selection: patches impose a survival probability (viability) on their inhabitants. The position of the peaks, on the phenotypic space $\vec{h}$ (illustrated by each colored vector) denotes the optimum phenotype, which changes from patch to patch. Reproduction: occurs sexually among individuals inhabiting the same patch and are not genetically too dissimilar, i.e., their genetic distance must be smaller than a critical value g. The offspring of a couple is a product of parental genomic recombination plus a possible mutation. Migration: occurs between patches with a probability m per individual. 100

Figure 35 – A mere illustration of moving optimum for a given patch. In this example, the displacement of the optimum phenotype is constant during each event, whereas in our simulations the displacement in each trait is taken from a Gaussian distribution. When we compare gradual (corresponding to $\tau = 1$) vs. abrupt environmental changes, the standard deviation for the gradual process is chosen such that over a time τ the distribution of the net displacement for the two processes are exactly the same. 102

Figure 36 – Species richness in terms of the width α_r and migration. The selective pressure is inverse with α_r, thus from right to left we go from the neutral regime where effectively there is no selection mechanism, to a domain of increased pressure where populations are at risk of local and global extinction. The values of the remaining parameters are those in bold in table 2. 105

Figure 37 – Average pairwise Hamming distance (left panel), average pairwise phenotypic distance (middle panel) and phenotypic distance to the optimum imposed by the inhabited patch (right panel) versus the width of the viability function, α_r. Migration probabilities are shown in the legends. The values of the remaining parameters are those in bold in table 2. 107

Figure 38 – The dependence of the species richness on the width of the viability function, α_r. The selection strength or pressure can be understood as $1/\alpha_r$. Different curves denote distinct values of the period of disturbance events, τ. From right to left $\tau = 200$, $\tau = 500$, $\tau = 1000$, and $\tau \to \infty$. The open symbols correspond to the gradual counterpart of the discontinuous environmental changes upon the condition that $\sigma^2_{gradual} = \sigma^2_h/\tau$. The values of the remaining parameters are those in bold in Table 2. 108

Figure 39 – Metapopulation extinction risk (left panel) and fraction of non-empty patches (right panel) versus the width of the viability function, α_r. The periods of disturbance events, τ, are indicated in the legend.The values of the remaining parameters are those in bold in Table 2. 109

Figure 40 – Average pairwise Hamming distance (left panel) and average pairwise phenotypic distance (right panel) as a function of the width α_r. Different values of the time to environmental changes are encoded by different colors, as indicated in the legends. Data are compared to their gradual counterparts, in which $\sigma_{gradual} = \sigma_h/\sqrt{\tau}$. The values of the remaining parameters are those in bold in table 2. 111

Figure 41 – Dependence on the number of traits. Species richness vs. the rescale quantity $\alpha_r/\sqrt{L}$, where L is the number of traits (values indicated in the legend). The values of the remaining parameters are those in bold in table 2. 112

Figure 42 – Dependence on the carrying capacity K. Species richness vs. the width α_r for different values of K, indicated in the legend. The values of the remaining parameters are those in bold in table 2. 113

Figure 43 – Dependence of the species richness on the width of the viability function, α_r. The standard deviations for environmental changes are indicated in the legend. The values of the remaining parameters are those in bold in Table 2. 114

Figure 44 – Dependence of the species richness on the width of the viability function, α_r under directional ($\mathcal{V} = 0.005$) and stabilizing selection. The two curves differ by abrupt or gradual environmental disturbances $\sigma^2_{gradual} = \sigma^2_h/\tau$, with $\sigma_h = 0.05$. The values of the remaining parameters are those in bold in table 2. 115

Figure 45 – Annual global average temperatures expressed as the difference from pre-industrial conditions based on the 1850-1900 average. Four different data sets are shown - HadCRUT5, NOAAGlobalTemp, GISTEMP, and Berkeley Earth - as well as two reanalyses - ERA5 and JRA-55. There is good agreement on the overall evolution of global temperatures and year-to-year variability. 116

Figure 46 – Species richness, average pairwise Hamming distance, average pairwise phenotypic distance and phenotypic distance to the optimum imposed by the inhabited patch versus the width of the viability function, α_r. Dependence on the magnitude of global trait variation σ_g is indicated in the legend. In the upper right panel, we distinguish the distance to local and global traits - full and empty points, respectively. The values of the remaining parameters are those in bold in table 2. 117

Figure 47 – One-dimensional representation of our modeling of phenotypic plasticity. The optimum phenotype is centered at a point h of the trait axis. Here, individuals whose phenotype is at a given distance from the optimum (purple interval), are considered sufficiently adapted to survive viability selection Eq. 4.1. In other words, if $|r - h| < \sigma_{pp}$ then $V = 1$. 118

Figure 48 – Species richness, average pairwise Hamming distance, phenotypic distance to the optimum and extinction risk for the effect of phenotypic plasticity σ_{pp} versus the width of the viability function, α_r. The values of the remaining parameters are those in bold in table 2. 119

Figure 49 – Depending on the slope at $\theta = 1$, the PGF $f(\theta)$ may have one additional solution at $\theta*$. 137

Figure 50 – Proportion of cases in which evolutionary rescue is attained by multiple mutational steps as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points. 144

Figure 51 – Extinction probability as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. From the green vertical line onwards, the population is in the regime of rescue/extinction. The plot compares the simulation results (data points) with the analytical predictions (lines), given by Eq.(19). The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points. 144

Figure 52 – Probability of extinction as a function of the stress level for the genotypic model. Curves represent different degrees of epistasis K. Error bars are smaller than points. Parameters for both cases are $L = 12$, $N_0 = 10^5$, $W_{max} = 1.5$, and $b = 6$. Mutation rate are $\mu = 10^{-5}$ for the left panel and $\mu = 10^{-3}$ for the right panel. 145

Figure 53 – Evolution of richness, acceleration of speciation and phylogeny balance for combinations of isolation $t_h \approx \{80, 20\}$ and migration rate $\bar{\varepsilon} = \{0.02, 0.08\}$. Solid lines represent the average of 50 replications, while dotted ones delimit the confidence interval of 90%. Following the α definition, the dashed red lines indicate the interval in which the results do not differ from random branching models. Background gray bands indicate when patches are isolated for each t_h. 146

CONTENTS

1 INTRODUCTION 19
1.1 AN INITIAL DISCLAIMER 20
1.2 POPULATION GROWTH 20
1.3 EVOLUTIONARY MECHANISMS 21
1.4 ADAPTATION ON FITNESS LANDSCAPES 27
1.5 POPULATION PERSISTENCE 33
1.6 POPULATION DIVERGENCE 35
1.7 OUTLINE 38
2 EVOLUTIONARY RESCUE 40
2.1 FISHER'S GEOMETRIC MODEL 41
2.2 POPULATION MODEL 43
2.3 ENVIRONMENTAL STRESS 44
2.4 DENSITY REGULATION AND PARALLELISM 44
2.4.1 Path Analysis 46
2.4.2 Simulation Protocol 48
2.4.3 Results 50
2.4.4 Discussion 59
2.5 EPISTASIS AND INCONGRUENCE 61
2.5.1 Kauffman's NK Model 61
2.5.2 Simulation Protocol 63
2.5.3 Results 64
2.5.4 Discussion 72
2.6 CONCLUSIONS 73
3 SPECIATION IN A NEUTRAL MODEL 75
3.1 DERRIDA-HIGGS MODEL OF SPECIATION 76
3.2 PHYLOGENETIC TREES 79
3.3 BARRIERS TO GENE FLOW 83
3.4 SIMULATION PROTOCOL 84
3.4.1 Parameter values and data treatment 86
3.5 RESULTS 87

3.6 CONCLUSION 94
4 SPECIATION IN A METAPOPULATION 96
4.1 MODEL 97
4.1.1 Metapopulation 97
4.1.2 Fitness landscape 98
4.1.3 Population growth and speciation 99
4.1.4 Enviromental perturbations 101
4.2 SIMULATION AND MEASUREMENTS 103
4.3 RESULTS 105
4.3.1 The role of migration in shaping diversity 105
4.3.2 Diversity and Environmental fluctuations 107
4.3.3 Further investigation of parameter space 111
4.4 CONCLUSIONS 119
5 OVERALL CONCLUSIONS 121
REFERENCES 123
APPENDIX A – BRANCHING PROCESS 135
APPENDIX B – DERRIDA-HIGGS DYNAMICS 140
APPENDIX C – CENTRAL LIMIT THEOREM 143
APPENDIX D – SUPPLEMENTARY RESULTS 144

1 INTRODUCTION

Nothing in biology makes sense
except in the light of evolution.
(Dobzhansky)

The evolution of life accompanies the dynamics of our planet Earth. From continental drift to glacial ages, geological events have shaped the Earth's surface and climate, influencing the evolutionary history of populations, with species regularly appearing, getting extinguished and modifying over larger time scales. Much more recently, the human species became capable of great modifications on the planet as well. Increasing global temperature, altering the course of rivers and an ever-expanding usage of the land are some examples of human activities that cause great pressure upon other species. Both natural and anthropomorphic causes can affect the stability of populations, leading to differential changes in species, a reorganization of species communities, and a risk of extinction. In an ever-changing planet, understanding the effect of environmental modifications on life's diversity and resilience is essential to grasp the fate of species, including our own.

Evolutionary Theory is one of the most beautiful and successful theories in science. Through the set of its essential mechanisms, we came to understand the complex collective behavior of living organisms, to track the origin of their common ancestors, and to account for the abundant variety of life forms. In this thesis, we focus on two aspects of evolution driven by environmental changes: population persistence and divergence. Persistence refers to a population's ability to survive drastic environmental alterations and their potential to adapt to the newly harsh conditions. Divergence refers to the initial process of differentiation that will ultimately give rise to new species, and encompasses the influence that variable environmental conditions have on the development and maintenance of diversity.

In this introductory chapter, we briefly present and discuss some relevant aspects that contextualize this work - the mechanisms shaping population evolution and evolutionary responses to environmental changes. The reader can jump to the last section of this chapter for the specific applications of these subjects in this thesis and their corresponding chapters.

1.1 An initial disclaimer

Evolution is inherently influenced by both random and deterministic factors. As such, a great body of evolutionary models is often stochastic, accounting for the inherent unpredictability of events like mutation, reproductive success, and environmental changes, which can lead to different evolutionary outcomes even with the same initial conditions (CROW; KIMURA, 1970).

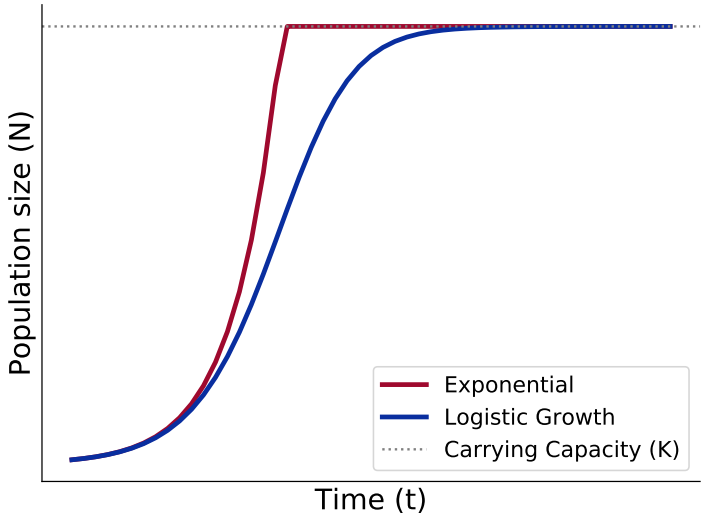
This thesis incorporates computational and probabilistic models to study population evolution driven by environmental changes. When possible, the biological concepts are presented together with simple probabilistic models of population genetics in the hope that they can complement each other to a better understanding.

Regarding the works resulting from this research, we utilize random branching processes for many of our analytical approximations, and the central limit theorem for a feasible comparison between abrupt and continuous environmental fluctuations (FELLER, 1991). On the computational aspect, all works are grounded in individual-based models with a genetic basis (DEANGELIS; GRIMM, 2014). This allows for simulation of complex ecological and evolutionary processes, including trait variation, demographic stochasticity, and the influence of individual-level mechanisms on population dynamics. In this microevolutionary approach, evolutionary patterns emerge from the collective interaction between individuals, their environment and the effects of genetic architecture.

1.2 Population Growth

In biology, a population is the summation of living organisms of the same group or species, inhabiting the same place. By reproducing, individuals of the same species pass on their hereditary traits to their descendants. The first and simplest life forms to arise on Earth replicate themselves, and at some point in life's history, some newly developed forms needed the combination of two parents to generate offspring. Even with a low life expectancy, a set of individuals will grow in number as long as food, space, and other resources are available. But of course, a population can not grow indefinitely. The limitations of space and resources impose an upper limit on the final population size, usually dubbed the system's *carrying capacity*. Additionally, the *rate* at which a population grows can change based on the current population density (Fig. 1).

Figure 1 – Simple distinction between independent and dependent density growth, represented by an exponential and logistic growth, respectively.



Source: The author (2024).

Density-dependent growth may reflect intrinsic ecological properties such as competition for resources, prey-predator dynamics, or simply crowding. Density dependence is important to conservation because it can lead to either population regulation (i.e., stabilization of population size) or population destabilization (thus increasing the probability of population going extinct) (GOMULKIEWICZ; HOLT, 1995). In particular, extinction risk increases significantly if population size or density falls below a certain threshold value (OVASKAINEN; HANSKI, 2003).

As long as environmental conditions do not interfere with the population structure or its carrying capacity, an established population may maintain its numbers invariant over long time scales. However, more commonly than not, intense or abrupt environmental changes can disturb stable populations and lead to their decline (WILLI; HOFFMANN, 2009).

1.3 Evolutionary mechanisms

The individuals comprising a population are different, and this information is stored in their genotypes (genes) and expressed in their phenotypes (traits). Over generations, evolutionary mechanisms shape the population composition, increasing or decreasing its variability. In this section, we present a short overview of the genotype-phenotype mapping, the inheritance of traits, and the main drivers of population evolution, namely selection, mutation, migration, and genetic drift.

Genotype-Phenotype mapping

The genome is a collection of "sites" (loci) that store the genetic information of an organism. In the molecular context of DNA polymerase, each locus is filled with one of the four possible nucleotides $\{A, T, G, C\}$. For a given genome with a length of L loci, there are 4^L unique combinations of this four-letter alphabet. Only some regions of the genome carry the genetic information that codes for proteins [1]. When read by sequence-reading proteins, this information is biologically expressed through phenotypic traits (RIDLEY, 2003): continuous or discrete observable characteristics including its morphology, immune defenses, reproduction rates, behavior, biochemical properties, etc.

Importantly, the mapping genotype-phenotype ($G \rightarrow P$), is not necessarily a one-to-one map (DROSSEL, 2001). Gene expression is regulated by a variety of mechanisms that may depend on stimuli from the external environment. Organisms may exhibit heritable phenotypical changes that do not involve changes in the genetic sequence of a cell at all - for example, through epigenetics and phenotypic plasticity. Furthermore, some genes may be active at different stages in the developmental process. In sum, the phenotype is the result of an extremely complex interplay of the different parts of the genome, and of its environment.

Drivers of evolutionary change

Mutation

The most common way to pass genetic material is through vertical gene transfer [2], most commonly known as reproduction. Asexual organisms make clones of themselves, while sexual organisms mix the genomes of both parents following a specific recombination rule. Both copying and mixing are molecular processes prone to *errors*, resulting in a change of the discrete hereditary information not present in the previous generation - a mutation.

At the molecular level, mutations are copying errors in the form of *base nucleotide substitutions*. Although other forms of modification of genetic information are well known (such

[1] Not all sections of the genome result in expression: some are responsible for the regulatory activity of reading-and-coding, catalytic tasks in the cells, and some others are regarded as inactive. A complete gene has several loci, typically of the order of $\sim 10^4$ in simple prokaryotes, reaching about $\sim 10^9$ in humans - where, in the latter, only ~ 3 percent code for proteins (NOWAK, 2006).

[2] Horizontal gene transfer is the transmission of DNA between organisms during their lifetime, rather than from parent to offspring. Most common in prokaryotic cells.

as insertion and deletion), the heuristic arguments for their negligible occurrence rely on the mechanism of self-correction present in the code transcription process (DROSSEL, 2001). Although some regions of the genome have a higher mutation rate than others, which mutation will arise and in which generation is completely unpredictable (LYNCH et al., 2016). Mutations are, therefore, a random source of genetic variation.

Genetic drift

The reproduction process inherently bears the mechanism of genetic drift - stochastic fluctuations in population gene pool. Genetic drift carries the probabilistic nature of a sampling process, since chance has a role in determining which individual survives, how many offspring they produce, and which genes they pass on to the next generation. Random drift does not have a preferential direction, it can increase or decrease a gene frequency, leading a genotype to be established[3] or be lost. As is the case with all stochastic processes, its effects manifest more rapidly and are more pronounced in smaller populations, which is even more relevant for the fate of novel mutations.

Let us use a simple stochastic reproductive model of asexual populations to illustrate genetic drift. In the Wright-Fisher model (GILLESPIE, 1998), the descendants of generation $t+1$ are randomly sampled with replacement from the parents' generation t with some probability $p \in [0, 1]$. The generations are discrete and do not overlap, all individuals reproduce and die simultaneously, i.e., once generation $t + 1$ is obtained, it replaces the previous one. With a constant population size N, let $X = 0, 1, 2, ..., N$ be a random variable assigning the number of copies of one of the lineages present. The probability to sample a number $X = j$ in N independent trials follows a binomial distribution $X \sim Bin(N, p)$, where

$$Prob\left(X_{t+1} = j | X_t = i\right) = \frac{N!}{j!(N-j)!} p^j (1-p)^{N-j},$$

which is known to have mean $E[X] = Np$ and variance $Var[X] = Np(1 - p)$. Here, p is simply the fraction of the individuals present of a given lineage,

$$p = \frac{p_k}{\sum_k p_k} = \frac{i}{N}, \tag{1.1}$$

which makes the sample process equivalent to randomly picking balls from a box with replacement [4]. Here we have a simple mathematical description of random genetic drift. At

[3] We say a genotype has established when its gene frequency or density becomes large enough to overcome the random fluctuations of genetic drift.

[4] This case represents the absence of selection or a neutral selective model.

any given generation, individuals lineage frequencies change at a rate inversely proportional to population size. Thus, the Wright-Fisher model indicates that random drift is stronger in small populations. Furthermore, its dependency on the current frequency implies that rare lineages (very low frequency) have a higher chance of being extinct in subsequent steps.

Selection

Selection is a mechanism responsible by the adaptation of the organisms to their environment(DARWIN, 1859). In a given environment, some traits may be particularly advantageous when compared to others enhancing the chance of an organism's survival and reproduction. Individuals with selectively favored traits are said to be *well-adapted* to the conditions they live in, because they have a *higher chance* of passing their genes, increasing trait frequency in future generations. Since evolution is a stochastic process, sometimes it is not the most adapted individual that thrives, especially when populations are small or their relative advantage is low.

The relative advantage is quantified as fitness. Assuming that the advantage is purely additive and does not change regarding individuals' frequency, in an isogenic population with fitness $F = 1$, a distinct genotype σ [5], is said to have fitness $F(\sigma) = 1 + s$. If $s > 1$, this genotype is favored by selection to increase its frequency. However, since reproduction is a stochastic sample process, all genotypes are susceptible to random genetic drift. Even the fittest among them can be purged in subsequent generations - especially in small populations. So it is natural to ask the probability of a genotype with relative fitness $F(\sigma) = 1 + s$ to escape random drift.

To understand how this advantageous effect acts on the probability of genotype survival let us use a classical branching process approach of following the fate of a single genotype (CROW; KIMURA, 1970). A population grows in discrete non-overlapping generations. Let $p_0, p_1, p_2, ..$ be the probabilities that the lineage will leave $0, 1, 2, ...$ descendants in the next generation, i.e. $P(X = k) = p_k$ and $\sum_k p_k = 1$. In particular, p_0 is the probability that given genotype will go extinct. Such assumption have a probability generating function:

$$P_X(Z) = p_0 + p_1\, Z + p_2\, Z^2 + ... = \sum_{Z=0}^{\infty} p_k\, Z^k, \tag{1.2}$$

where the probability of leaving k mutant genes for the next generation is given as the coefficient of Z^k.

[5] Therefore, we are assuming that such genotype contributes to a phenotype which ultimately affects fitness.

Now, let the number of offspring per birth event be drawn from a Poisson distribution with mean $\lambda = 1 + s$. Thus, if $X \sim Poisson(\lambda)$, so $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ with offspring number $k = 0, 1, 2....$ Therefore, its generating function can be obtained as

$$\begin{aligned} P_X(Z) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} Z^k \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda Z)^k}{k!} \\ &= e^{-\lambda} e^{\lambda Z} \\ &= e^{(1+s)(Z-1)}. \end{aligned}$$

The probability of extinction of a branching process starting with a single individual is the smallest root of the equation $P_X(Z) = Z$ for $Z \in (0, 1)$ (see Appendix A), which admits a fixed point solution given by $Z = e^{(1+s)(Z-1)}$. Rewriting $Z = 1 - \pi$, where π is the probability that such genotype is fixed or established (i.e. does not go extinct), one have $P_X(1 - \pi) = e^{(1+s)(1-\pi-1)}$. Solving the fixed point for π:

$$\pi = 1 - e^{-(1+s)\pi}, \tag{1.3}$$

which is a transcendental equation for π. We then solve for the selective advantage s:

$$\begin{aligned} -(1 + s)\pi &= ln(1 - \pi) \\ s &= -1 - \frac{ln(1 - \pi)}{\pi}. \end{aligned}$$

Assuming that π is small, and hence s, we expand the natural log and neglect the terms of order $O(\pi^3)$ and higher, such that

$$\begin{aligned} s &= -1 - \frac{1}{\pi}\left[-\pi - \frac{\pi^2}{2} + O(\pi^3) \right] \\ \Rightarrow \pi &\approx 2s. \end{aligned} \tag{1.4}$$

Thus, under a random branching process, survival probability is linear with the fitness difference. An extensive revision made by Wahl and Patwa (PATWA; WAHL, 2008), summarizes many different approaches to determine the probability of survival of a genotype (or to escape drift) as a function of its selective advantage. Strikingly, all the models' predictions point to a linear dependence on $\pi \approx cs$, differing only by a scalar magnitude $c \in \mathbb{R}$. This semi-quantitative result implies that, despite the distinction in the approaches, the analyses are valid for a large class of exchangeable models, as they converge to expected general results.

Lastly, we would like to address the interplay between selection and drift. If there is no selective advantage $s = 0$, the probability that a single genotype survive is equal to $\pi = \frac{1}{N}$. So, there should be a threshold for fitness values, small enough such that

$$\pi = 2s \approx \frac{1}{N}. \tag{1.5}$$

Therefore, for a fixed s, drift dominates in very small populations, where $Ns \ll 1$, despite selection. On the other hand, in the $Ns \gg 1$ limit, corresponding to large populations sizes, selection plays a role by increasing the frequency of the fitter genotypes despite drift (DESAI; FISHER, 2007).

In sum, adaptation is the evolution of trait frequencies through the process of selection. While mutation and drift are inherently random forces, selection has a *direction*, guiding evolution towards heritable adaptations to the current environment.

Migration and Gene flow

Last, but no less important, is the mechanism of migration or gene flow: the exchange of individuals and, consequently, their genetic material within or between divided subpopulations. When an individual migrates to a new population and breeds, it introduces genes from its ancestral population into the new one. Depending on migration intensity, gene flow might be enough to homogenize the populations, reducing genetic variability of the total population (GAVRILETS, 2004).

A simple model illustrates gene flow homogenizing effect (RIDLEY, 2003): let us consider the case of two asexual populations, 1 and 2, that can exchange migrants with probability m per generation. The frequency of a specific genotype in population 1 in generation t is written p_t. Supposing population 2 is sufficiently large, we can assume that the frequency of the same genotype in population 2 is not changing between generations and write it as q. In the absence of mutation and selection, genotype frequency at the next generation p_{t+1} depends on the number of migrants received from population 2 with probability $m\,q$ and of natives that do not migrate with probability $(1-m)\,p_t$,

$$p_{t+1} = (1-m)\,p_t + m\,q.$$

Solving for $t = 0$, we have the frequency $p_1 = (1-m)\,p_0 + q\,m$. If we keep applying this recursive relation, we found that the genotype frequency in population 1 at any generation t

is given by

$$p_t = q + (p_0 - q)\,(1 - m)^t.$$

The equation says that the difference between the gene frequency in population 1 and population 2 decreases by a factor $(1 - m)$ per generation. Given enough time, the small population will have the same gene frequency as the large population $p = q$. Similar arguments apply if, instead of there being one source and one recipient population, the source is a set of many subpopulations (RIDLEY, 2003; GAVRILETS, 2004).

High migration rates can even restrict genetic variation promoted by other mechanisms, such as mutation and recombination. Seeing it the other way around, restrictions to gene flow will allow genetic variation to increase, resulting in population divergence at the genetic level, which, as we will discuss below, is an essential step for the emergence of new species. Migration highlights how a population's organizational structure and habitat are also important for its genetic diversity. In particular, regarding adaptation, migration can also increase the population's potential to persist local environmental changes, as it can bring genetic novelty not present in the receiving subpopulation.

These mechanisms all have the potential to alter the frequencies of genes and traits within populations, thereby acting as drivers of evolutionary change. Nonetheless, natural selection and genetic drift modify the prevalence of existing genes by increasing or decreasing the occurrence of traits. On the other hand, mutation and migration can introduce entirely novel characteristics into a population.

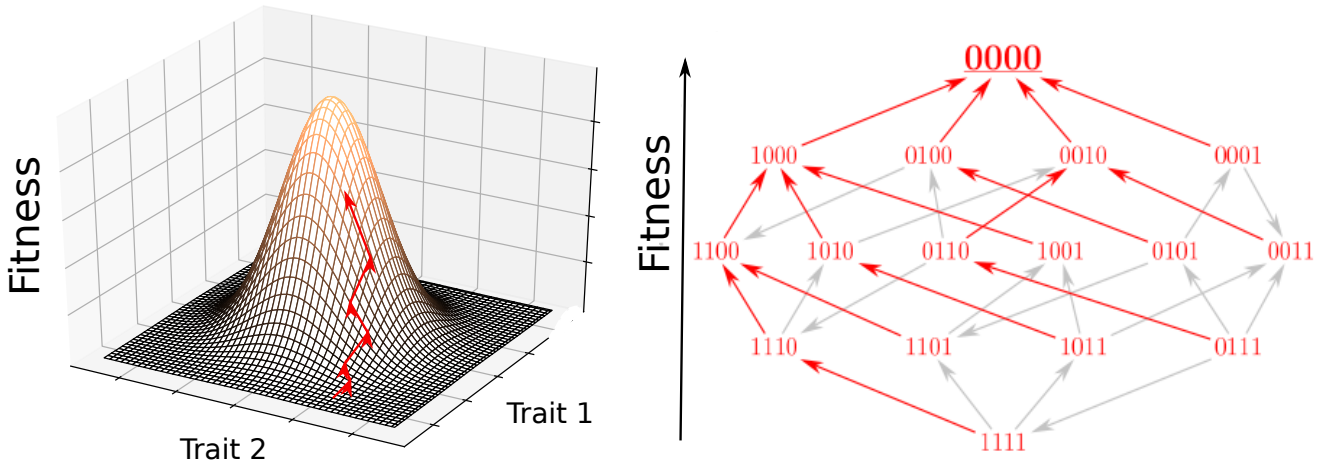
1.4 Adaptation on Fitness Landscapes

Fitness encompasses the organism's ability to survive, mate, and produce offspring. Through experimental and observational methodologies, researchers can approximate a genotype or phenotype's fitness by attributing it to a quantitative value (for instance, the number of offspring it produces) (BARRICK; LENSKI, 2013). However, it is important to recognize that fitness is inherently a relative measure, dependent on the environment in which the organism lives.

Historically, the field of evolutionary biology has benefited from mathematical models of fitness landscapes (FRAGATA et al., 2019). Fitness or adaptive landscapes offer insights into evolutionary dynamics by metaphorically illustrating how traits or genes are related to an organism's reproductive success and the dynamics of adaptation. In its simplest form, most

models assume the existence of an optimum combination of traits that maximizes fitness in a given environment (ORR, 2005b). This idea is better visualized as a landscape-like figure (Figure 2A), where each "point" in the base grid represents a specific combination between the axis traits, and the height corresponds to their respective fitness.

Figure 2 – Simple representation of fitness landscapes. (A) In continuum phenotypic landscapes, population can explore all posible combinations of traits. Red arrows represent a "trajectory" of a population. As selection increases the frequency of fitter individuals, adaptation is an walking towards the peak. (B) Discrete genotypic landscape for when only 4 mutations are responsible for the adaptation process. Arrows represent single mutation transitions, where the red ones are transitions to genotypes of higher fitness.



Source: The author (2024).

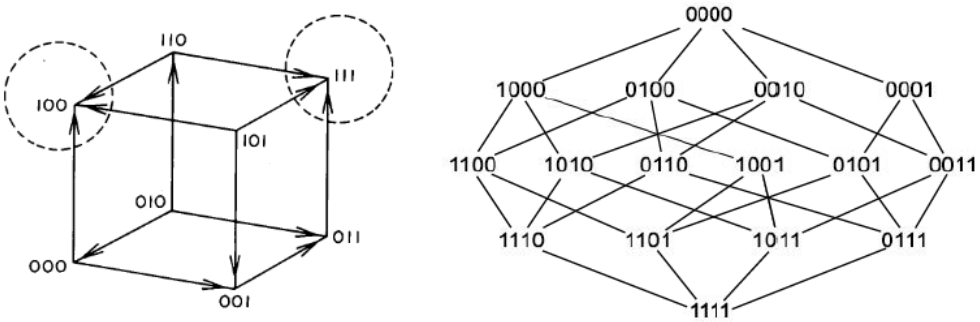
Using this heuristic geometric representation, each individual of a population can be represented by a point in space, whose position matches their respective gene or trait. Over generations, evolutionary mechanisms shape the population composition. Mutation and recombination "move" the offspring through the fitness landscape as vectors of random length and directions. Newly arising mutations (*de novo* mutations) are regarded as beneficial or deleterious depending on their contribution to the population's overall fitness (DESAI; FISHER, 2007), i.e., if they bring individuals closer or far from the optimum phenotype, respectively. Since selection can increase or decrease traits' frequency depending on their fitness, under this framework, adaptation is seen as population trajectories over a fitness landscape towards fitness peaks.

Genome size and Epistasis

Choosing between phenotypic or genotypic landscapes is a matter of the biological system of interest. A phenotypic model presumes that the traits under selection can assume continuous values (such as weight, body part lengths, or metabolic rate), in other words, the population has the potential to "explore" all trait space. This reasoning underlies the common assumption in population genetics that adaptation consists of fine-tuning the phenotype with *very large numbers of genes* carrying mutations of relatively small effect (ORR, 2005b). Effectively, this approximation is equivalent to a genome of infinite sites or alleles, where every mutation is a *de novo* mutation.

However, adaptation to specific environmental conditions might involve very few loci. For example, the adaptation of bacterial β-lactamase to a novel antibiotic is related to only 5 mutations that control the phenotype conferring resistance. In these cases, a common approach for most studies is to model a finite biallelic genome of length L, with alleles $\{0, 1\}$ merely to indicate if a mutation is present or absent in the original sequence (KAUFFMAN; LEVIN, 1987). This leads to a discretization of the fitness landscape (Figure 2B), where each of the 2^L possible combinations is mapped into its respective fitness value. Evolution on genotypic space carries a notion of neighborhood and distance distinct from their phenotypic counterpart. Since mutations are modeled as 1-single letter transitions, this gives rise to a binary space connected through single-digit flip - or an L-dimensional hypercube $\mathbb{H}_L^2$ (Figure 3). This space is characterized by short distances and high dimensionality (NOWAK, 2006).

Figure 3 – Hypercube representation for L=3 and L=4. Only in the first case, fitness is assigned to each vertex in parenthesis. Arrows heads represent transitions to fitter states, while the circles outline a state that maximizes fitness - a fitness peak.



Source: Modified from (KAUFFMAN; WEINBERGER, 1989)

Furthermore, the discretization of genotypic space may exhibit epistasis (KAUFFMAN; WEINBERGER, 1989). Epistasis refers to a phenomenon where the fitness effect of a mutation at a given locus may depend on the mutations at other loci - in other words, it may depend on the genetic background. Mathematically, if the fitness of a genotype $\sigma = (s_1, s_2, .., s_L)$ can be decomposed into a sum of independent contributions from each site, thus

$$F(\sigma) = \sum_{i=1}^{L} F(s_i),$$

then the sites will be uncorrelated, and the landscape exhibits a single fitness peak like in Figure 2. Otherwise, correlated sites can be thought of as a coupling or interaction between the sites, similar to a Hamiltonian for a system of interacting particles (MANHART; MOROZOV, 2014).

Unlike the continuum space of large genomes, where each mutation contributes independently to fitness, epistasis introduces non-additive effects: a double mutant (with mutations in two loci) may have a phenotype that differs significantly from the sum of the effects of each single mutation (De Visser; KRUG, 2014). As we shall discuss in Chapter 2, epistasis is responsible for determining landscape topography and exhibits multiple fitness peaks, which, ultimately, affects the accessibility of evolutionary trajectories.

Population size and mutation fate

Population size mediates the counteracting forces between selection and genetic drift. At large population sizes $N >> 1$, selection prevails, purging deleterious mutations and rapidly increasing the size of beneficial ones. Over fitness landscapes, population trajectories resemble monotonic walks transitioning over states of ever-increasing fitness. At finite sizes, however, the population is better described as a cloud around a given trait moving on the landscape, but in some cases can divide itself into subpopulations, each of them describing their particular trajectories, which can culminate in the same or even different fitness peaks. Population evolutionary trajectories are subject to stochastic fluctuations, and new or rare mutations (which have low frequency) can be lost due to random genetic drift (see Figure 4).

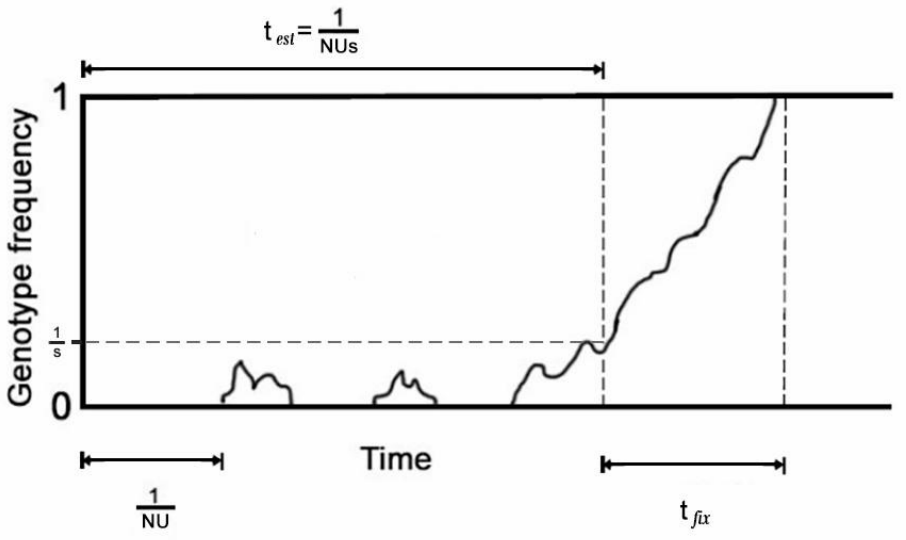
Let us assume that, during reproduction, there exists a chance U for the descendant being born with a different genetic code from their parents. Hence, considering a population of size N, NU is the influx of new mutants in the evolving population at each generation, providing a genotypic diversity upon which selection can act. Additionally, let s (from Eq. 1.4) be the

probability they survive drift. If, over generations, mutations arise at rate NU on average, a mutation takes a time $t = \frac{1}{NU}$ to arise, and a time

$$t_{est} \sim \frac{1}{NUs}, \tag{1.6}$$

to get established within the population. The establishment refers to a threshold frequency needed for a genotype to no longer be extinct by drift. It does not have a priori definition, is an *ad hoc* threshold imposed by an arbitrary confidence interval (DESAI; FISHER, 2007). Upon our assumptions from previous sections, the mutant lineage must reach a size $N \sim 1/s$ before it becomes "safe" from extinction and begins to grow mostly deterministically (DESAI; FISHER, 2007).

Figure 4 – Mutations arise at short time scales but are rapidly purged by genetic drift. With a probability proportional to s, mutations can reach a threshold frequency to survive drift and, once done, they can reach fixation at time t_{fix}.

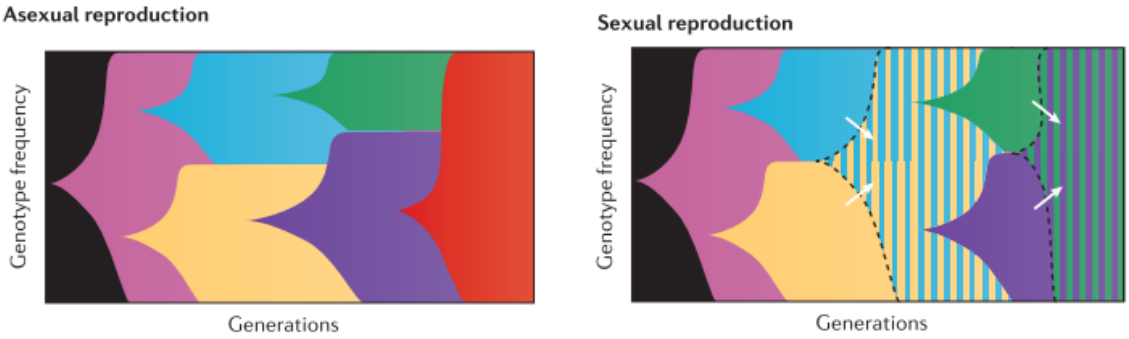


Source: The author (2024).

Typically, before a mutation can sweep to fixation, *de novo* beneficial mutations arise in a different lineage and become established, leading to competition between each other (Figure 5). This phenomenon is called clonal interference (BARRICK; LENSKI, 2013), common in asexual populations where there is no recombination. On the other hand, through sexual reproduction, previously competing mutations can coexist in the offspring, leading to an increase in fitness.[6]

[6] This is one of the factors that make sex an advantageous process.

Figure 5 – Representation of the discussed regimes in asexual and sexual populations. If *de novo* mutations establishes simultaneously, they compete for fixation. In contrast with clonal interference, we see how recombination can decreases the competing aspect by assimilating both genotypes in the same descendant.



Source: Modified from (BARRICK; LENSKI, 2013).

An important disclaimer

As a metaphor, fitness landscapes are universally used to describe adaptation, emphasizing the potential of species to adapt and explore new habitats and resources, and more recently have been used to study speciation (GAVRILETS, 1997). However, some considerations must be made regarding their limitations and applications.

A tentative picture emerging from the fitness landscape metaphor is the possibility of continuous optimization by selection. This view might lead to a common misconception that survival is guaranteed only for the fittest, better-adapted individuals. One can not stress enough that the evolutionary process is inherently probabilistic: genetic variants that aid survival and reproduction are much more likely to become common than variants that don't. Both in natural and lab populations, maladapted individuals still coexist in low frequency.

The construction of empirical fitness landscapes is only feasible for simple lab-cultivated microorganisms (De Visser; KRUG, 2014), for which the variety of expressed traits is less abundant, the mapping genotype-fitness is straightforward, and genomic variation can be carefully registered by whole-genome sequencing (BARRICK; LENSKI, 2013).

Nevertheless, phenotypic and genotypic fitness landscapes are a semi-quantitative approximation that "highlights" the microscopic details of the evolutionary process. Properties of the landscape - such as peak fraction, deviation from additivity, and accessibility of paths (LOBKOVSKY; WOLF; KOONIN, 2011) - are investigated to better understand the restrictions that the space itself imposes on evolutionary dynamics, integrating a path-dependent analysis

to evolutionary theory. For our particular interest, it can give us insight about the processes involving population persistence and divergence.

1.5 Population Persistence

The environment sets the conditions by which populations adapt. Presumably, the set of phenotypic traits of natural populations is responsible for maintaining stable population sizes in a constant environment (BARRETT; HENDRY, 2012). When the environment changes, existing phenotypes are expected to be less adapted. Abrupt environmental changes induce levels of stress that may cause population decline, hamper its adaptation, and, in the last instance, lead to its extinction. As a response to escape such a fate, populations can either disperse or adapt via phenotypic plasticity and genetic evolution (LALAND et al., 2015; BELL, 2017).

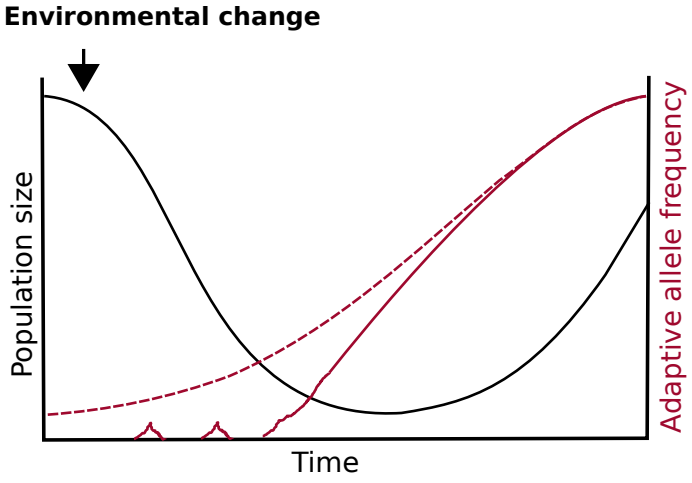
If your home becomes unfavorable or even threatening, the first and most intuitive response of any individual is to disperse in search of better, safer places. But to organisms lacking mobility, such as plants, or large-scale environmental changes like global temperature oscillations, all that is left is to adapt or to perish (DAVIS; SHAW; ETTERSON, 2005; BANK, 2022).

Phenotypic plasticity is an adaptation mechanism present in some organisms that allows the variation of traits during their lifespan without any genetic change. In analogy to the previous section, the map $G \rightarrow P$ is no longer unique because the same genotype no longer expresses a trait value but encompasses a trait range [7]. Plants are everyday evidence of a plastic response: put a plant in a water vessel or a very different soil and it will change its root lengths to allocate its resources better; or note the change in its leaves' shape and size when growing in altered light levels.

Populations lacking a plastic response can only generate the necessary novelty to cope with environmental changes through genetic evolution: a fitter genotype must be established - either by new mutations or standing genetic variation. Therefore, adaptation must happen sufficiently fast to ensure population survival. When a declining population, fated to go extinct, restores a positive growth rate by genetic adaptation, we say the population has undergone an *Evolutionary Rescue*.

[7] By definition, plasticity has an energetic cost to maintain the biochemical mechanisms used to sense the environment and produce the appropriate response. Otherwise, one may wonder, if plasticity is so good and fast to cope with environmental changes, would plastic organisms out-compete non-plastic ones?

Figure 6 – Signature of evolutionary rescue. Once a stressful enough environmental change happens, population declines and can only restore positive growth through adaptive evolution. Traced line shows the case for when the mutation is already present, and solid line for *de novo* mutations (note the failed attempts to survive genetic drift).



Source: The author (2024).

The signature of evolutionary rescue is the U-shaped curve for population decline and recovery (Fig 6). Populations have a higher chance of going extinct the smaller their size and the longer the time spent in small sizes (HANSKI, 1998).

Evolutionary rescue has been frequently documented in laboratory-controlled experiments involving microorganisms, providing a means to estimate the frequency of rescue under various conditions. Observations in natural populations are somehow cumbersome. The lack of demographic data, the correlation between multiple traits, and density-dependent regulation can all obscure the evolutionary response (CARLSON; CUNNINGHAM; WESTLEY, 2014). Nevertheless, the development of antibiotics, the agricultural application of pesticides, and the invasion of new species are accepted examples involving rescue (BELL, 2017). Comprehending the mechanisms involved in evolutionary rescue is essential for conservation biology and population management in the face of environmental instability.

Evolutionary rescue integrates a non-trivial relation between demography and evolution (ORR; UNCKLESS, 2008; WILLI; HOFFMANN, 2009). On one hand, a population's adaptive evolutionary response depends on the availability of beneficial genotypes that can restore a positive growth rate. On the other hand, at small population sizes, genetic drift is enhanced, reducing genetic variation and the establishment chance of fitter genotypes, further reduces evolutionary potential (LYNCH et al., 2016).

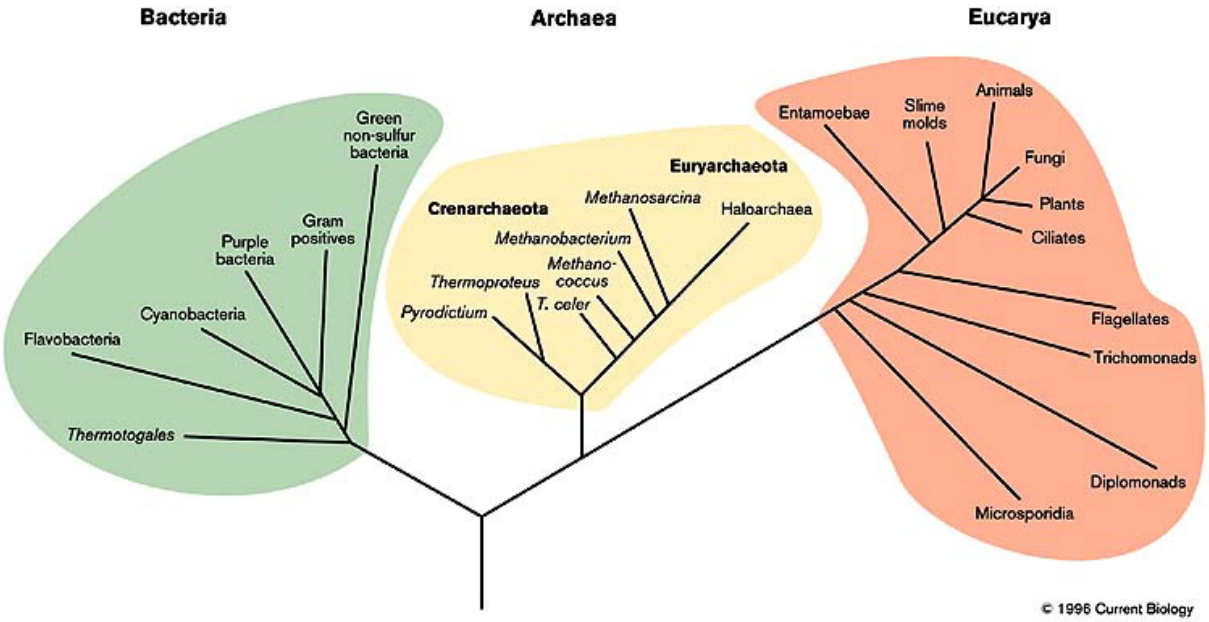
Mathematically, both the fitness contributions of genotypes to population growth and the probability of surviving drift can be analytically derived from classical assumptions of population

genetic theory and random branching processes. The probability of evolutionary rescue can then be correlated with the availability of rescue mutations. We develop and compare these derivations with numerical simulations in Chapter 2.

1.6 Population Divergence

The evolution of life on Earth shows a sustained diversification process with species regularly appearing, getting extinguished and modifying over larger time scales. Modern life is believed to have descended from a single ancestor and then split into different groups. Based on molecular data (WOESE; KANDLER; WHEELIS, 1990), the classification of life divides the early life forms into three domains: it has branched first in Bacteria and Archaea; and the Eukarya (the domain we belong to) branched from Archaea. This diversification is famously illustrated in the Tree of Life (Fig.7), which connects our common ancestors' evolutionary history to today's living forms.

Figure 7 – Conceptual visualization of the Tree of Life, showing ancient life's differentiation into three great domains. The subdivision into further groups shows the variety of life forms in each domain. The evolutionary history within each of these groups (e.g. the animal kingdom) can also be represented by their corresponding tree.



Source: From (WOESE, 1996).

The emergence and maintenance of diversity is one of life's inherent features and can be observed and classified across different biological scales - from domains to species.

Species is a fundamental "unit" used to classify diversity, and speciation is the process by which an ancestral population splits into two or more distinct descendant populations. How much "distinct" is required to identify a species varies, because several closely related species concepts exist (RIDLEY, 2003). The biological species concept (MAYR, 1963) defines species in terms of interbreeding (or gene flow) [8]: individuals belong to the same species as long as reproduction of fertile offspring is possible among them. In other words, the rise of new species requires *reproductive isolation*. The emergence of reproductive isolation involves the interruption of gene flow between populations, leading populations to harbor distinct gene pools and evolutionary trajectories, which in turn promotes speciation (DOBZHANSKY, 1970). Under this genetic perspective, population genetic divergence will always accompany speciation.

In all the works in this thesis regarding speciation, we utilize the biological species concept described above, with focus on the mechanisms that promote genetic differentiation or population divergence. Below we present how geographic isolation and selection (in the form of local adaptation or *divergent selection*) are the two most common mechanisms that enhance reproductive isolation, and thus the evolution of population divergence (FITZPATRICK; FORDYCE; GAVRILETS, 2009).

Geographic Isolation

Geographic isolation is the most accepted and simplest mechanism to explain speciation events. As populations are separated from each other, random change in gene frequencies through mutation, recombination and genetic drift, together with the impossibility of gene flow, will eventually lead to different genetic compositions. Early studies on the gene flow effect on speciation usually adopted the framework of a metapopulation - a population divided among regions or demes and connected by migration. These early models of biogeography (MACARTHUR; WILSON, 1967) could provide valuable insights into the intensity of gene flow required to prevent or allow speciation to occur, and the effects of habitat structure on diversity patterns.

In natural populations, barriers can be permeable or simply do not last forever. For instance, sea-level oscillations driven by Earth's glacial periods induce periods of connection and isolation between oceanic islands (HE et al., 2019) and rivers close to continental shelf (BAGGIO

[8] Other species definitions are more suitable for asexual organisms, for example, their similarity in morphological, genetic, or ecological characteristics.

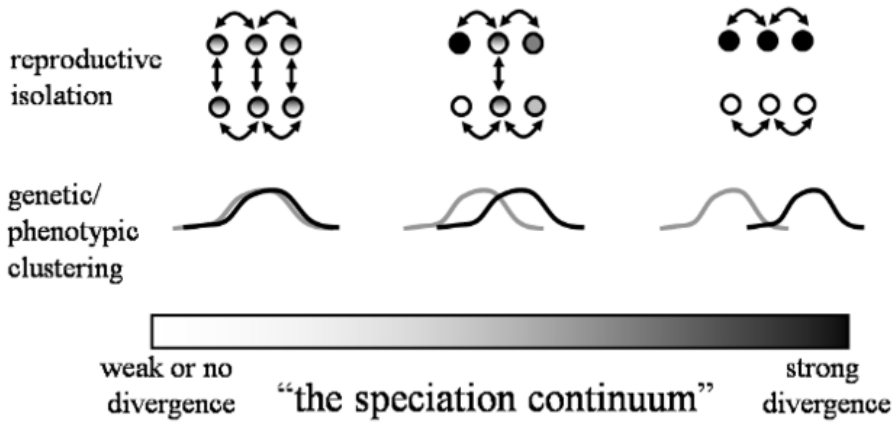
et al., 2017). The secondary contact between previously isolated populations can affect the speciation dynamics in a non-trivial way (AGUILéE; LAMBERT; CLAESSEN, 2011). In chapter 3, we investigate the speciation dynamics of a two-deme system when migration is intermittent, in other words, the population evolves in altered periods of connection and isolation.

Selection

Genetic divergence may also be enhanced by natural selection. Populations adapting to different environments or exploiting different resources, generate phenotypic and genetic differences among them, leading to reproductive isolation (RUNDLE; NOSIL, 2005). Replicate laboratory experiments with *Drosophilia* fruit flies have shown that lineages selected to different environments have a lower chance of mating success than lineages evolving to the same laboratory conditions (NOSIL; HARMON; SEEHAUSEN, 2009). Furthermore, great genetic divergence followed by a rapid emergence of reproductive isolation was found for populations adapting in high-temperature environments (FRY, 2009).

An important final consideration is the strength of selection or its selective pressure. Low genetic variance is expected in scenarios where selection strongly favors a specific combination of traits, in opposition to a neutral scenario where selection is absent. In metapopulations, with local adaptation in each region, strong selection reduces the effective number of migrants, since many of them are fated to extinction when moving between environments. In this sense, strong selective pressure might overcome moderate gene flow to result in population divergence.

Figure 8 – Visualization of the differentiation between individuals of a population. The arrows in the upper row indicate that reproduction is possible. Population diverges, acquiring distinct phenotypes and/or genotypes, leading to speciation.



Source: Modified from (NOSIL; HARMON; SEEHAUSEN, 2009).

It is important to highlight that selection and migration may act together or separately. Furthermore, both mechanisms can be altered by abiotic and biotic environmental changing factors such as climate, habitat structure, food resources and, of course, geographic barriers. In this sense, environmental changes can alter the speciation process, e.g. resulting in large or lower diversity. In chapter 4, we model a metapopulation connected by migration in which each deme favors distinct traits. We report the degree of genetic differentiation among the subpopulations at which the homogenizing effect of gene flow and the differentiating effect of local adaptation interplay. We then ask how perturbations in the local environmental conditions might affect the diversity patterns.

1.7 Outline

Stochastic evolutionary models are developed to mathematically describe the processes that change the distribution of heritable variants in a population over time. The environment defines both the conditions necessary for population adaptation and its organizational structure. In a changing world, alterations in the environmental conditions can profoundly influence a population's evolutionary history.

This thesis includes the works I have done over the course of my Doctorate studies and the different subjects approached here are connected by environmental changes' role in the evolution of population persistence and divergence. The chapters' ordering is not chronological but tailored so that the structure of this thesis is divided into two parts. The first half (chapter 2) addresses the persistence of populations through evolutionary rescue. In the latter half (chapters 3 and 4), we study population divergence and the patterns of speciation driven by environmental changes.

More specifically, in Chapter 2, a well-adapted asexual population is submitted to an abrupt variation in the conditions necessary for survival. Under this level of selective stress, population size is intensely reduced and can lead to its extinction. Population extinction can be prevented by density regulation, but if environmental change is too stressful, only the rise and establishment of a new mutation that can cope with the new environment guarantees population persistence. We assess the population extinction risk, the degree of parallelism of the evolutionary response, and how epistasis may alter the outcome due to the incongruence of the fitness landscape map. Its results were published in Refs. (FREITAS; CAMPOS, 2024b; FREITAS;

CAMPOS, 2024a).

In Chapter 3, geographic environmental events affect the connection between a population divided into a two-deme system. In this neutral selective model, only the migration rate between the patches and the intermittent periods of connectance and isolation can promote population divergence leading to speciation. We report the dynamics of population divergence through its phylogenetic speciation tree, characterize their structure, and ask if the intermittent periods of connectance and isolation leave a signature in the phylogeny's macroevolutionary patterns. Its results were published in Ref. (FREITAS; CAMPOS; ARAUJO, 2024).

In Chapter 4, both selection and migration are present in a metapopulation model. We measure genomic divergence and speciation patterns for varying intensities of gene flow and selective pressure. We then compare how different scenarios of environmental variations - including the magnitude and frequency of these changes - affect the maintenance of diversity. Its results were published in Ref. (FREITAS; ARAUJO; CAMPOS, 2022).

2 EVOLUTIONARY RESCUE

Mal escapo à fome, mal escapo aos tiros
Mal escapo aos homens, mal escapo ao vírus
Passam raspando, tirando até meu verniz.
(Gilberto Gil)

Evolutionary rescue (ER) is the process in which a population declining due to environmental changes, fated to go extinct, averts its fate by increasing in number through adaptive evolution alone (BELL, 2017; LALAND et al., 2015). Under ER, adaptation must happen sufficiently fast to ensure population survival. In turn, the probability of converting a positive growth rate over short time scales depends on the availability of genetic variation of significant effects and their chance to survive drift when rare (GOMULKIEWICZ; HOLT, 1995; ORR; UNCKLESS, 2008; ANCIAUX et al., 2018; MCDONOUGH; CONNALLON, 2023).

In natural populations, ER has been observed in populations persisting herbicide resistance (KREINER; STINCHCOMBE; WRIGHT, 2018), the introduction of novel-pathogens (SEARLE; CHRISTIE, 2021), and climate change (SCHIFFERS et al., 2013) as common causes of population decline. ER is also observed in the experimental evolution of microbial systems, studying population stress response to antibiotics (WEINREICH; WATSON; CHAO, 2005), increased salinity concentration (BELL; GONZALEZ, 2011), and temperature variation (HUANG et al., 2018).

Theoretical research has made significant contributions to identifying the key factors that influence the probability of evolutionary rescue. Initial population sizes (FLATHER et al., 2011; BELL, 2013) and the pace of population decline, which is based on how stressed out or how maladapted the population is to the unfavorable environment (HARMAND et al., 2017; MARREC; BITBOL, 2020; WAHL; CAMPOS, 2023), are known demographic factors. Standing genetic variation (genotypes that exist before the environmental change) and the rate of de novo mutations from the dwindling wild-type population (ORR; UNCKLESS, 2014; UECKER; OTTO; HERMISSON, 2014; TANAKA; WAHL, 2022) are relevant genetic factors.

Although treated separately in most studies, in reality, genetic and demographic factors are intertwined. Population size and its variation rate mediate the evolutionary mechanisms that shape the genetic diversity within populations, the availability of beneficial mutations, and the

constraints on adaptive pathways (ORR, 2005a; De Visser; KRUG, 2014). These demographic effects on genetic variation show the importance of integrating evolution and ecology for studying population persistence in the face of environmental stress (GOMULKIEWICZ; HOLT, 1995; ORR; UNCKLESS, 2008).

In this chapter, we investigate the influence of demographic and genetic factors on evolutionary rescue (WILLI; HOFFMANN, 2009). In the first work, we study the fate of a density-regulated population under stress. We show that stress level divides the dynamics into two distinct regimes, one where density regulation guarantees population persistence, and the other where survival depends on whether evolutionary rescue is achievable. We characterize the probability of evolutionary rescue and the degree of genetic parallelism of the evolutionary response. In the latter work, we implement a three-layer genotype $\rightarrow$ phenotype $\rightarrow$ fitness map to address the influence of genetic background on ER. We show that the likelihood of extinction is smaller when the degree of epistasis is higher. We discuss the results in light of the emerging non-linearity/incongruence of the fitness landscape.

In the following sections we first present the common design for both studies: the implementation of Fisher's Geometric Model as a fitness landscape; the population model of demographic variation that may or may not contain a density dependence; and our procedure for environmental change. Next, the sections of each study, namely 2.4 Density Regulation and Parallelism and 2.5 Epistasis and Incongruence, contain their respective discussions, analytical and simulation results.

2.1 Fisher's Geometric Model

Historically, the field of evolutionary biology has benefited from mathematical models of fitness landscapes (FRAGATA et al., 2019), including the study on ER (TENAILLON, 2014; ANCIAUX et al., 2018; WAHL; CAMPOS, 2023). In Fisher's geometric model (FGM), one ascribes a phenotype vector to each individual whose components correspond to trait values (FISHER, 1930; ORR, 2005a). Thus, each phenotype is seen as a point in a phenotypic space of dimensionality n, where n denotes the number of traits that characterize the individual. Therefore, a phenotype is represented as a vector $\vec{z} = (z_1, z_2, \ldots, z_n)$, in which z_i is the value of trait i.

The FGM assumes the existence of an optimum value for each trait, o_i, such that the optimum phenotype $\vec{o} = (o_1, o_2, \ldots, o_n)$ also corresponds to a point in the $n-$dimensional phenotypic space (WAXMAN, 2006). We assume that all traits are under Gaussian stabilizing selection, such that the fitness of a given individual with phenotype $\vec{z}$ is proportional to

$$S(\vec{z}) = \exp\left[-\frac{d^2}{2\alpha_z^2}\right], \quad \text{with} \quad d^2 = \sum_{i=1}^{n} |z_i - o_i|^2, \tag{2.1}$$

where $1/\alpha_z$ is the selection strength or the width of the Gaussian curve, and d is the phenotypic or Euclidean distance from the individual to the optimum phenotype. Throughout this work, we assume $\alpha_z = 1$ [1].

In the case of classical FGM, which relies on an infinite number of loci, the phenotype of the mutant offspring will amount to $\vec{z}+\vec{u}$, where the size of mutation effects, r, on the phenotype, is taken from an exponential distribution of mean value λ. Following the same procedure as in Ref. (CONNALLON; CLARK, 2014; MCDONOUGH; CONNALLON, 2023), the mutation vector with size r is

$$\vec{u}|r = \left(\frac{ry_1}{\sqrt{\sum_k y_k^2}}, \frac{ry_2}{\sqrt{\sum_k y_k^2}}, \ldots, \frac{ry_n}{\sqrt{\sum_k y_k^2}}\right), \tag{2.2}$$

where y_k with $k = 1, \ldots, n$ are independent random variables taken from a standard normal distribution. Thus, the current model assumes that mutations affects all traits (isotropic mutation).

Additionally, we also consider a version of the FGM in which the individuals are equipped with an explicit genetic ground: a finite-locus biallelic genome. Henceforth, this FGM version is referred to as the genotypic FGM (HWANG; PARK; KRUG, 2017; WAHL; CAMPOS, 2023). In the genotypic FGM, the genome size is L, and so, to each individual, one associates its phenotype $\vec{z}$ together with its genotype $S = (s_1, s_2, \ldots, s_L)$. The genotypic space consists of 2^L possible sequences, as $s_j = 0, 1$. The state $s_j = 0$ means the absence of mutation in locus j. Once a mutation occurs, a single locus is randomly chosen to mutate.

To build the phenotypic space underlying the genotypic space, L mutation vectors are required, $\vec{\eta_1}, \vec{\eta_2}, \ldots, \vec{\eta_L}$. Taken as reference the sequence $S = (0, 0, \ldots, 0)$ to which one ascribes, without loss of generality, the phenotype $\vec{0} = (0, 0, \ldots, 0)$, the phenotype associated to any of the 2^L sequences is done by making use of the additive assumption (TENAILLON,

[1] In Chapter 4, the effect of selective strength in population evolution is an important varying parameter.

2014). For example, the corresponding phenotype of genotype $S = (0, 1, 1, 0, \ldots, 0)$ is simply obtained by additively combining mutation vectors $\vec{\eta_2}$ and $\vec{\eta_3}$, such that $\vec{z_s} = \vec{0} + \vec{\eta_2} + \vec{\eta_3}$. A library of L mutations is prepared in advance, and when a back mutation occurs, the phenotypic value is adjusted by subtracting the effect of that mutation. Each mutation vector $\vec{\eta_k}$ is determined in the same way as in classic FGM [2], and follows Eq. (2.2).

2.2 Population model

We assume an evolving asexual population with non-overlapping generations of haploid individuals whose size N is variable. As aforesaid, the fitness landscape assumes an optimum value for the trait, $\vec{o}$ (WAXMAN, 2006). In particular, we assume that the population is under density regulation. In a density-regulated population of size N, and with selective pressure upon the individual given by Eq. 2.1, the absolute fitness of an individual of phenotype $\vec{z}$ is then calculated as (CHEVIN; LANDE, 2010)

$$W(\vec{z}, N) = W_{max}^{1-N/K} S(\vec{z}), \tag{2.3}$$

where K is the carrying capacity and W_{max} is the maximum reproductive rate. At each reproduction event, the number of offspring for each individual is drawn from a Poisson distribution of parameter $W(\vec{z}, N)$. Each offspring inherits the phenotype of its parent, but mutations can occur for each birth event with probability μ per individual per generation.

Our choice of density-dependence holds when population growth is hindered by crowding, predators, or competition (OSMOND; MAZANCOURT, 2013). In natural populations, this kind of dependence is observed in the early stages of invading species and pathogen host-switching (IWASA; MICHOR; NOWAK, 2004), where the abundance of resources or the lack of predators allows a rapid increase in population size despite the struggle to adapt to unexplored environments or new-host conditions, which is then followed by a density-dependent growth as crowding or competition increases (SEARLE; CHRISTIE, 2021).

[2] Importantly, the genotypic FGM does NOT assume a direct relationship from genotype to fitness. Its main contribution is the discretization of phenotypic space. In the section of our next study, we define a proper model that accounts for a genotype-fitness-map.

2.3 Environmental Stress

Without loss of generality, we assume that the optimum trait is at position $\vec{0}$. As an initial condition, a isogenic population is fully adapted to its environment (thus, $\vec{z} = \vec{0}$ and $S = 1$ for all individuals), and the initial population size equals K (thus, $W = 1$). At time $t = 0$, an abrupt environmental change takes place, and now the phenotypic optimum is placed at a distance $d = \Delta$ from $\vec{0}$, but in a random direction. Distance Δ is chosen such that initial population fitness equals $W_0 = 1 - \delta$, where the parameter δ refers to the initial fitness drop or stress level. Substituting all this in Eq. 2.3

$$1 - \delta = e^{-\Delta^2/2\alpha_z^2},$$

we find that the initial phenotypic distance to the optimum at fixed δ is

$$\Delta = \sqrt{-2\alpha_z^2 \ln(1-\delta)}. \tag{2.4}$$

Because of the density regulation, at this point, we may distinguish the existence of two regimes: the non-extinction regime and the rescue/extinction regime. Without density regulation and in the absence of mutations, the population will face a geometric decline in its size and will ultimately face extinction unless a rescue mutation occurs. Upon density regulation, we easily notice that, even in the absence of mutation, when

$$\delta < 1 - \frac{1}{W_{max}}, \tag{2.5}$$

the population can restore a positive growth rate by adjusting its population size. When the condition (2.5) holds, thus $W > 1 = W_{max}^{1-\frac{N^+}{K}}(1-\delta)$, and in the absence of mutations, positive growth rates are restored at population size N^+, given by

$$N^+ = K\left[1 + \frac{\ln(1-\delta)}{\ln(W_{max})}\right]. \tag{2.6}$$

When the environmental stress is sufficiently high, such that $\delta > 1 - \frac{1}{W_{max}}$, a positive growth rate can only be restored through the occurrence of adaptive mutations.

2.4 Density regulation and Parallelism

Diversified populations embarking on seemingly comparable habitats are expected to display similarity at both phenotypic (BOLNICK et al., 2018) and genotypic levels (SCHLUTER et al.,

2004; ORR, 2005b). However, despite similar environmental conditions, populations undergoing evolution often exhibit divergent evolutionary trajectories. This divergence underscores the complex interplay of genetic, ecological, and stochastic factors contributing to the variability observed in the evolutionary outcomes of seemingly analogous populations (STUART et al., 2017). To quantify how recurrent or similar the process leading to adaptation is, theoretical models and replicate experiments (WICHMAN et al., 1999; BAILEY; RODRIGUE; KASSEN, 2015; LENSKI, 2017; BLOUNT; LENSKI; LOSOS, 2018) evolve independent populations to equal or different environmental conditions to assess their degree of parallel evolution.

Parallelism measures repeated evolutionary changes across various biological scales, including phenotypic, functional, biochemical, or genetic traits (BOLNICK et al., 2018). Its analysis can focus either on the similarity of the outcomes or of the evolutionary trajectories. Parallel evolution does not need to be complete, and evaluation of replicate populations often resides within a spectrum between total distinction and parallelism (THOMPSON; OSMOND; SCHLUTER, 2019; YAMAGUCHI; WILEY; OTTO, 2022). Given its broader application, one must adequately define parallelism for a particular problem. Assuming specific genes conduct the adaptation process differently and their recurrence might be essential to ER, it sounds natural to inquire if the same set of mutations are responsible for rescuing populations that face similar challenges.

In particular, the demographic dynamics typical of an evolutionary rescue scenario might affect parallelism in non-trivial ways. The probability of parallel evolution is known to be affected by population size: a higher degree of parallelism is expected with increased population size because this reduces the effect of drift and favors the establishment of large-effect mutations (BAILEY et al., 2017). Therefore, one expects that different regimes of population variation influence parallelism.

Following this reasoning, in this work, we investigate how the intensity of environmental change affects population persistence and the parallelism of its evolutionary trajectories. We study the evolution of density-regulated populations and characterize the regimes in which evolutionary rescue is attained. We also evaluate the degree of parallelism of mutational pathways between replicated populations and how this evolutionary process is affected by the different demographic regimes.

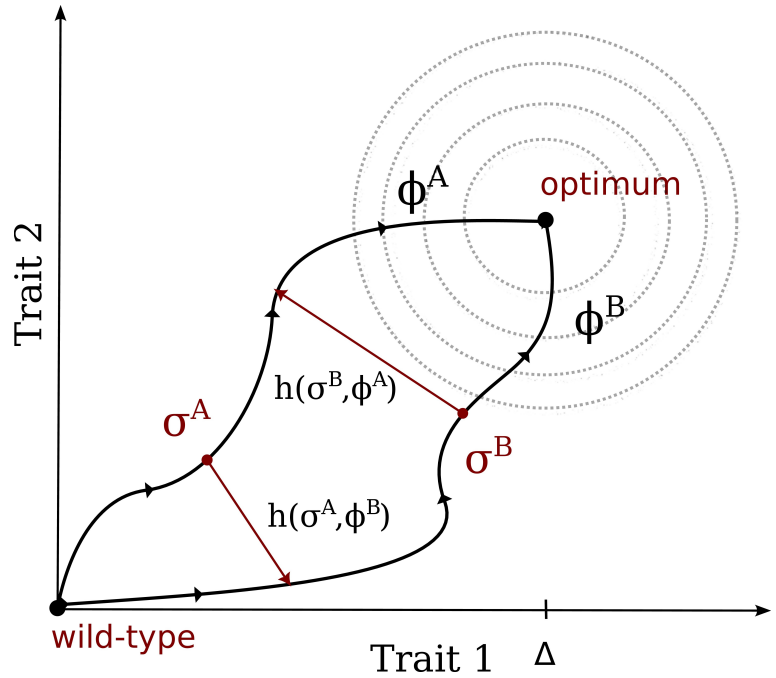
2.4.1 Path Analysis

Over time, the population evolves subject to selection, genetic drift and mutations. From a genetic viewpoint, adaptation at the DNA sequence level proceeds via mutational pathways: one can construct an evolutionary trajectory through a succession of genotypes by tracking a set of mutations arising at a given generation. By starting with a population of clonal individuals of genotype σ_0, the succession $\sigma_0 \to \sigma_1 \to \ \ldots \ \to \sigma_n$ is dubbed an evolutionary path ϕ. By registering genotypic changes along the dynamics, one can study evolution from the perspective of adaptive trajectories over fitness landscapes (LOBKOVSKY; KOONIN, 2012; SZENDRO et al., 2013; De Visser; KRUG, 2014). Evolutionary constraints determine the likelihood of evolution taking any particular path among the many possible competitive trajectories (WEINREICH; WATSON; CHAO, 2005). By comparing the trajectories of independent populations, this framework provides a statistical tool to characterize parallelism at a microevolutionary scale, where the order in which mutations arise and are established will be relevant to the population's rescue.

One must establish a criterion to determine which genotypes represent each step of the evolutionary path. Following a previous study (SZENDRO et al., 2013), we store each *de novo* mutation that arises for the first time in each generation, along with the genotype from which it originated. After the simulations have ended and the dominant genotype has been identified, we trace the line of descent that links it to the ancestral genotype. Alternatively, a second method that relies on the fittest genotype present in each generation is considered (FREITAS; WAHL; CAMPOS, 2021). Accordingly, the evolutionary trajectory consists of a collection of genotypes that represent an ascending route, from a fitness perspective, toward the ending point of the evolutionary path. Both methods yield similar qualitative results.

To quantify parallelism, we use two distinct measures: the degree of parallelism and the mean path divergence. The degree of parallelism is defined here as the probability that the evolutionary paths of the two replicates of the population, here denoted as replicates A and B, subject to the same environmental conditions, are the same. Under this perspective, paths that diverge by at least a single genotype in the sequence are treated as distinct. Thus, the degree of parallelism is quantified by measuring the fraction of trajectories in a given ensemble of size E in which ϕ_k^A equals ϕ_k^B with $k = 1, \ldots, E$, where the superscript $\ell = A, B$ denote the replicate index. On the other hand, the mean path divergence is a more refined measurement

Figure 9 – Representation of two distinct paths with the same initial and ending points. The Hamming distance h is measured for each point $\sigma^A \in \phi^A$ to every point $\sigma^B \in \phi^B$, and vice-versa. The shortest measure is stored.



Source: The author (2024).

of parallelism since it quantifies the degree of similarity between the paths - or, analogously, of dissimilarity.

We measure the divergence among accessible paths to account for the inner-path distance between all genotypes of a pair of paths (LOBKOVSKY; WOLF; KOONIN, 2011; MANHART; MOROZOV, 2014). One defines $\Omega(\phi_k^A, \phi_k^B)$ as the distance, or divergence, between two paths ϕ_k^A and ϕ_k^B. For the sake of clearness, from now on we will omit the index k. The divergence between two path is then calculated as:

$$\Omega(\phi^A, \phi^B) = \frac{1}{\Gamma(\phi^A) + \Gamma(\phi^B)} \left(\sum_{\sigma^A \in \phi^A} h(\sigma^A, \phi^B) + \sum_{\sigma^B \in \phi^B} h(\sigma^B, \phi^A) \right), \tag{2.7}$$

where $\Gamma(\phi)$ is the length (number of steps) of path ϕ, and $h(\sigma^A, \phi^B)$ is the shortest Hamming distance between a genotype σ^A and all genotypes $\sigma^B \in \phi^B$. To be clearer, *for each* genotype σ^A comprising pathway ϕ^A, one estimate its Hamming distance to *every* genotype $\sigma^B \in \phi^B$. The lowest distance is then stored and repeated until all genotypes in ϕ^A are rated. The process is then repeated in reverse, from ϕ^B to ϕ^A (see Figure 9). The divergence $\Omega(\phi^A, \phi^B)$ is the mean value of those shortest Hamming distances. The mean path divergence is calculated as the average of $\Omega(\phi_k^A, \phi_k^B)$ with $k = 1, \ldots, E$, in the ensemble of E trajectories.

Upon the definition of the quantities above, the degree of parallelism accounts for the situations in which $\Gamma(\phi^A) = \Gamma(\phi^B)$, and also $\Omega(\phi^A, \phi^B) = 0$.

2.4.2 Simulation Protocol

In this section, we describe how we implement our simulations. We distinguish how we estimate the probability of rescue from that of the calculations of the properties related to the evolutionary trajectories, as the latter ones are more costly from a computational perspective. Table 2 lists the parameters used and their values in our study.

parameter	definition	value
N_0	initial population size	10000
K	carrying capacity	10000
W_{max}	reproductive factor	1.5
L	number of loci	12
n	number of traits	3; **5**; 10; 20; 30
U	mutation probability	0.005
λ	mean phenotypic effect	0.4
$1/\alpha_z$	strength of selection	1.0
δ	stress level	0.0-0.8

Table 1 – Parameters of the model. The third column either defines the value or range of values of each parameter explored along the work. In bold, the most prevalent used value of each parameter.

At the beginning of each run, one has a clonal population whose genotype is the same as the reference sequence $S = (0, 0, \ldots, 0)$ with corresponding phenotype $\vec{0} = (0, 0, \ldots, 0)$, as explained above. For each run, we draw L mutations vectors $\vec{\eta_1}, \vec{\eta_2}, \ldots, \vec{\eta_L}$, in which the magnitude r of each of those vectors is a random variable taken from an exponential distribution of mean value λ. Given r, the vector $\vec{\eta_i}$ has its components determined by Eq. (2.2), and so n random draws from a standard normal distribution are required.

After setting the genotype-phenotype mapping, the phenotypic optimum $\vec{o} = (o_1, o_2, \ldots, o_n)$ is placed at a phenotypic distance Δ from $\vec{0}$, i.e., it lies in a random position of a hypersurface of radius Δ centered at $\vec{0}$. Once the position of the phenotypic optimum is defined, one can now easily calculate the fitness value associated with any sequence $S = (s_1, s_2, \ldots, s_L)$, as the fitness depends on the phenotypic distance to the optimum.

The clonal population with initial fitness $1-\delta$ is then let to evolve subject to selection and mutation. The offspring number of each individual is given by a Poisson distribution of parameter $W(\vec{z},N)$. The offspring is an exact copy of its parent, except when a mutation occurs with probability U for each birth event. In the case of mutation, a random genome locus is chosen and changed, $0 \rightarrow 1$ or $1 \rightarrow 0$. According to our description, in the first generation, there will be $N_0(1-\delta)U$ mutants, and this number of new genetic variants will shrink as time advances unless a rescue mutation occurs.

The population is let to evolve up to extinction, i.e., population size shrinking to zero, or when the number of individuals of fitness value larger than one, meaning positive growth rate, exceeds a critical size, which we choose as 100. The outcome is independent of this particular choice, provided that it is not too small but large enough to warrant that the mutant has been established. Simulation results are precisely the same as those in which the population is left to evolve for a long time. This criterion is the same as in Ref. (OSMOND; OTTO; MARTIN, 2020; WAHL; CAMPOS, 2023). We then store the outcome of the run, extinction or rescued, and the probability of extinction is just a fraction of the runs in which the population was extinct. The likelihood of rescue is just its complement.

To estimate the degree of parallelism and mean path divergence, the initial population and fitness landscape are set up precisely in the same way as described above. The path analysis is carried out among those trajectories where the population did not go extinct. However, instead of stopping the simulation when the number of individuals with a positive growth rate attains 100, the population evolves up to either reaching the genotype providing the highest fitness, i.e., the global optimum of the genotypic fitness landscape, or for $5,000$ generations, whatever occurs first. Note that, in the genotypic model, because of the discretization of the projection of the fitness landscape onto the genotypic space, the phenotype corresponding to the genotype that provides the highest fitness will almost certainly differ from $\vec{o}$. Usually, the accessibility of the global optimum is pretty high, though the population can sometimes evolve to another local optimum of the fitness landscape. It is important to highlight at this point that contrary to the phenotypic FGM in which a single optimum exists, the discreteness of the genotype space gives rise to the existence of multiple local optima of the fitness landscape (HWANG; PARK; KRUG, 2017).

2.4.3 Results

Probability of evolutionary rescue

Here we quantify how the severity of the environmental variation, here quantified by the parameter δ, affects the probability of evolutionary rescue.

An exact expression for the probability of rescue upon density regulation is quite troublesome, as the fitness varies with the population size. However, we can get essential insights from a more straightforward scenario. Therefore, in the formulation that follows, the fitness is assumed to be equal to $W_{geom} = W_{max}\exp(-d^2/2\alpha_z^2)$, i.e., there is no mechanism of density regulation. Under this assumption, a wild population with initial fitness $W_0 = 1-\delta$ will face a geometric decline in size unless a rescue mutation occurs. In this former approach, we will consider the phenotypic FGM. Suppose the wild-type population is at a distance $\vec{d}$ from the optimum phenotype. Given a mutation $\vec{u}$, the fitness of the mutant is then calculated as

$$\begin{aligned} W(\vec{z}) &= W_{max}\exp\left[-\frac{1}{2}(\vec{d}-\vec{u})^2\right] \\ &= W_{max}\exp\left[-\frac{d^2}{2}-\frac{u^2}{2}+du\cos\theta\right] \\ &= W(u,\theta), \end{aligned} \tag{2.8}$$

where θ is the angle between vectors $\vec{u}$ and $\vec{d}$. We want to find the probability $\bar{\pi}$ that a randomly chosen mutation is a rescue mutation and survives drift. Mutations' size and orientation are drawn from the distribution in Eq. 2.2, and we have:

$$\bar{\pi} = \iint\limits_{W(u,\theta)\geq 1} P_u(u)\,P_\theta(\theta)\,\pi(u,\theta)\,d\theta\,du. \tag{2.9}$$

Here, $\pi(u,\theta)$ is the probability of establishment of a mutation of fitness $W(u,\theta)$. We use the branching process formulation (see Chapter 1) to determine $\pi(u,\theta)$, and so $\pi(u,\theta)$ is numerically obtained as the fixed point of the probability generating function for a Poisson-distributed offspring number, $\pi = 1-e^{-\pi W(u,\theta)}$.

The double integral is bounded by the intervals of u and θ in which population has a positive growth rate $W(u,\theta)\geq 1$ (i.e. rescue is attained). In principle, we are free to choose the distribution $P_u(u)$. On the other hand, the distribution $P_\theta(\theta)$ must account for the dot product $(\vec{u}.\vec{d})$ and the dimensionality of phenotype space n. Without loss of generality, one

may assume that $\vec{d}$ lies along the x-axis. In this case, $\cos\theta$ is just the projection of the unit vector along $\vec{u}$-direction on the x-axis (see Fig. 10 for a schematic representation). Importantly, $\cos\theta$ is invariant to changes in the length of the mutation vector $\vec{u}$. Eq. (2.9) is better written as

$$\bar{\pi} = \iint\limits_{W(u,\cos\theta)\geq 1} P_u(u)\, P_\theta(\cos\theta)\, \pi(u,\cos\theta)\, d(\cos\theta)\, du. \tag{2.10}$$

where $P_\theta(\cos\theta)$ is the probability density of $\cos\theta \in (-1,1)$, given by

$$P_\theta(\cos\theta) = C_n \left[1-\cos^2\theta\right]^{\frac{n-3}{2}}, \quad \text{with} \quad C_n = \left[\sqrt{\pi}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}\right]^{-1}, \tag{2.11}$$

as demonstrated by Wahl and Campos (2023).

Let us now define the limits of integration in Eq. (2.9). As aforementioned, a rescue mutation must restore a positive growth rate. This means that the mutant must lie within a hypersphere of radius r in the phenotypic space such that

$$W(r) = W_{max}\exp\left[-\frac{1}{2}r^2\right] = 1.$$

This hypersphere centered at the phenotypic optimum and of radius $r = \sqrt{2\ln W_{max}}$ delimits the domain of rescue. From Eq. (2.8), one sees that the condition $W(u,\theta) = 1$ holds when $-\frac{1}{2}(\vec{d}-\vec{u})^2 = r^2$. Solving for u, one finds that the upper and lower limits for u are

$$u_\pm = d\cos\theta \pm \sqrt{d^2(\cos^2\theta - 1) + r^2}. \tag{2.12}$$

Thus, for a given θ, the interval $u \in (u_-, u_+)$ delimits the subset of rescue mutations (as in Fig. 10). When $\theta = 0$, one has $u \in (d-r, d+r)$, while θ_{max} is such that $\vec{u}$ tangents the hypersphere, wielding a single solution $u = d\cos\theta_{max}$, and so

$$\cos\theta_{max} = \sqrt{1-\left(\frac{r}{d}\right)^2}.$$

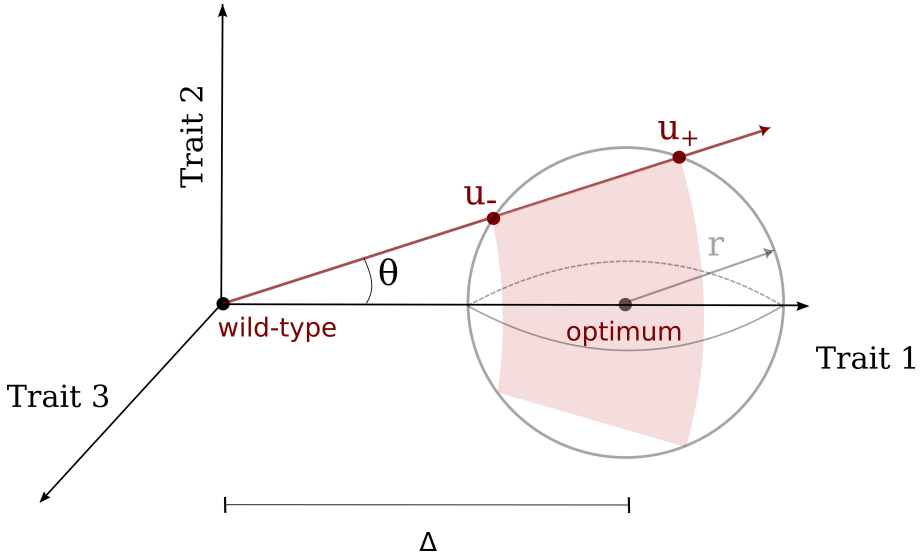
Note that this construction holds only when $d > r$, where $d = |\vec{d}|$. Therefore the probability of rescue is finally provided by

$$\bar{\pi} = \int_{u_-}^{u_+}\int_{\cos\theta_{max}}^{1} P_u(u)\, P_\theta(\cos\theta)\, \pi(u,\cos\theta)\, d(\cos\theta)\, du. \tag{2.13}$$

Because $\bar{\pi}$ (Eq. 2.13) corresponds to the probability of rescue of a single mutation; the probability of extinction of the population is simply

$$P_{ext} = (1-\bar{\pi})^A, \tag{2.14}$$

Figure 10 – Three-dimensional Fisher's geometric model, $n = 3$. The wild-type population is at a distance $d = \Delta$ from the phenotypic optimum. The sphere of radius r delimits the domain of rescue, $W(u, \theta) = 1$. For a given angle θ, mutations of magnitude between u_- and u_+ are rescue mutations. Owing to the symmetry, all mutation vectors lying on the red-shaded conical surface are rescue mutations.



Source: The author (2024).

where A is the total number of mutants arising before extinction (ORR; UNCKLESS, 2008). For a population declining geometrically

$$A = UN_0 \sum_{i=1}^{\infty} W_0^i = UN_0 \frac{1-\delta}{\delta},$$

where $W_0 = 1 - \delta$ is the wildtype fitness.

The left panel of Figure 11 presents the probability of extinction as a function of the stress level δ for the case of density-independence, i.e., the population faces a geometric decline in the absence of rescue mutations. In this simpler scenario, we obtain a perfect agreement between simulation results and our analytical prediction (Eq. 2.14).

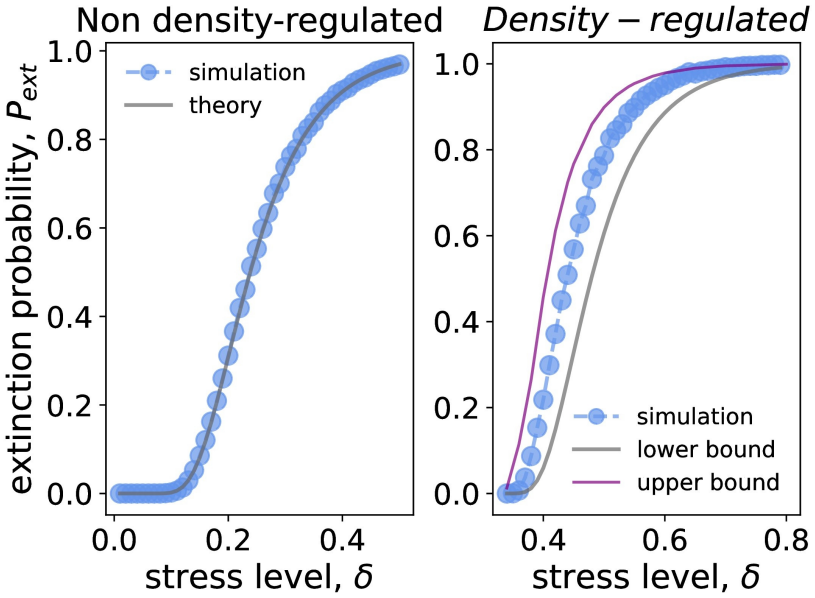
An approximation for the case with density regulation

Upon density regulation, the fitness of both wild-type and mutants are no longer constant, as they depend on the population size (Eq. 2.3). However, the previous development for the density-independent case can provide an upper and lower bound for the probability of rescue/extinction when density regulation exists.

For the lower-bound estimate of the probability of extinction, we assume that fitness remains constant and equal to $W(\vec{z}) = W_{max} S(\vec{z})$. Therefore, we assume a scenario in which fitness is over-evaluated, enhancing (reducing) the chance of rescue (extinction).

As the stress level δ is taken at the reference of the initial population fitness, in which $N_0 = N$, one has that the fitness of the wildtype population is simply $W(\vec{0}) = W_{max}(1-\delta)$. As said before, when $\delta > 1 - \frac{1}{W_{max}}$, the growth rate becomes negative, and so the population will be doomed to extinction unless a rescue mutation happens. In the right panel of Figure 11, we compare the simulation for a population upon density regulation and the theoretical lower bound for this probability that assumes density-independence.

Figure 11 – Extinction probability as a function of the stress level δ for the phenotypic FGM. The plot compares the simulation results (data points) with analytical predictions (lines). In the left panel, the simulation data is contrasted with the analytical prediction in Eq. (2.14), whereas in the right panel, the simulation data is plotted alongside upper and lower bound estimates of the extinction probability for a population under density regulation, as provided by Eqs. (2.14) and (2.16). The parameter values are carrying capacity $K = 10000$, number of traits $n = 5$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$.



Source: The author (2024).

In its turn, an upper bound can also be found by considering the formulation of constant fitness, but now including the dependence on the time the mutation occurs. In this approach, the fitness value to be considered is the fitness when the mutation arises, which now carries a dependence on the population size. The reasoning is that the fitness at the moment mutant appears will be central to its establishment. The population size at time t can be determined by the recursive equation

$$N_{t+1} = N_t W_{max}^{1-\frac{N_t}{K}}(1-\delta), \tag{2.15}$$

and so the fitness of the mutation at time t will be estimated as given by Eq. (2.3). The

likelihood of extinction of the population can be obtained as

$$P_{ext} = \prod_t \left[1 - \bar{\pi}(t)\right]^{UN_t}, \tag{2.16}$$

where now $\bar{\pi}(t)$ depends on the time t the mutant arises, and UN_t is the number of mutants generated at generation t. In the expression above, Eq (2.16), the probability of extinction is just the product of extinction probabilities of mutants arising in all times t.

The reason Eq.(2.16) overestimates the actual extinction can be understood as follows: When a rescue mutation occurs at a time t with corresponding population size $N_t > 1$, one does not observe an immediate reversion of the tendency of decreasing the total population size as the rescue mutation is rare at the moment it occurs. It will take a while until it gets established and increases in number to compensate for the decline of the wild-type population. The upper bound is also plotted in the right panel of Figure 11.

The genotypic FGM

From now on, we focus on the genotypic version of the FGM. As defined earlier, we consider a finite biallelic genome of size L. In contrast to the previous scenario, the pool of mutation vectors is now restrained to L draws, and so *de novo* mutations at a given time t are Poisson distributed with mean $A_t/L = N_t U/L$. We will use the same reasoning employed to the upper limit of the extinction probability developed above, but now under the assumption of a finite genome size L.

Importantly, not all mutations can lead the population to the rescue domain. One can define the proportion or probability that a given mutation can be a rescue mutation as

$$f = \iint\limits_{W(u,\cos\theta)\geq 1} P_u(u)\, P_\theta(\cos\theta)\, d(\cos\theta)\, du.$$

The probability that a single *de novo* mutation occurs i times at time t and fails to rescue the population is $[1 - \pi(t)]^i$. Summing over all independent lineages i,

$$\sum_{i=0}^{\infty} \frac{e^{-\mu}\mu^i}{i!}[1 - \pi(t)]^i = \exp\left[\frac{-\pi(t)A_t}{L}\right]$$

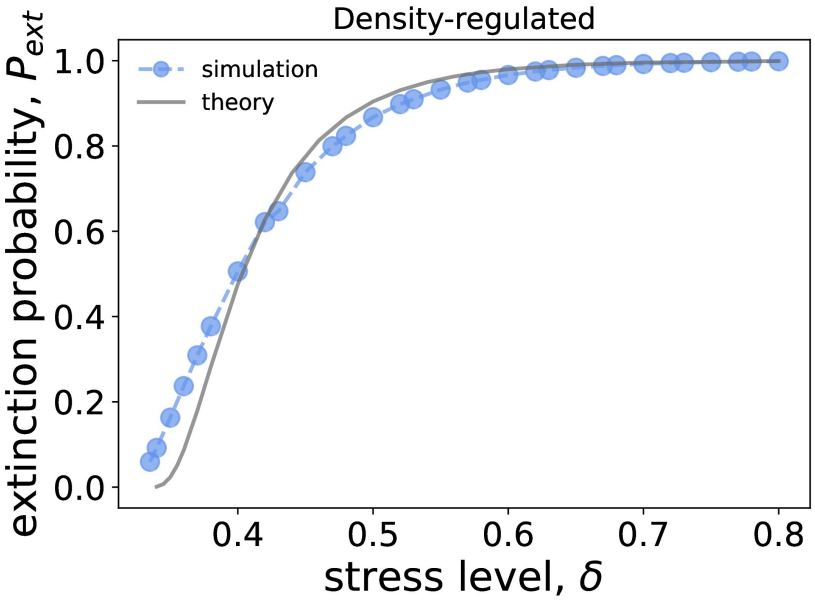
in which $\mu = A_t/L$ (WAHL; CAMPOS, 2023). Now integrating over the rescue domain, we find the probability of rescue owing to mutations at time t:

$$R_t = 1 - \frac{1}{f} \iint\limits_{W(u,\theta)\geq 1} P_u(u)\, P_\theta(\theta)\, \exp\left[\frac{-\pi(t)A_t}{L}\right] d\theta\, du. \tag{2.17}$$

Where, once again, the fixation probability depends on the magnitude of mutation effects u and $\cos\theta$. Since there are L possible mutation vectors, the probability that a mutation rescues the population given L is $P_{rescue|L}(t) = 1 - (1 - R_t f)^L$ (WAHL; CAMPOS, 2023). In this way, our upper estimate for the probability of extinction for a genotypic FGM is simply

$$P_{ext} = \prod_t (1 - R_t f)^L. \tag{2.18}$$

Figure 12 – Extinction probability as a function of the stress level δ for the genotypic FGM. The plot compares the simulation results (data points) with the analytical predictions (lines), given by Eq. (2.27). The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, number of traits $n = 5$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$.



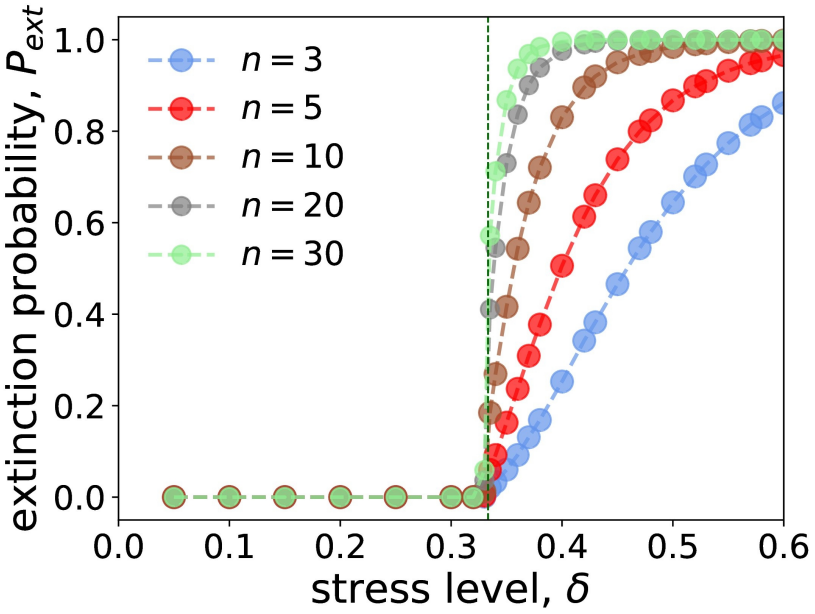
Source: The author (2024).

Figure 12 shows the extinction probability versus the stress level δ for the genotypic FGM. In this instance, we have considered genome size $L = 12$. The approximation provided by Eqs.(2.17)-(2.27), which works as an upper limit, fits reasonably well our simulation data in a broad range of δ-values. However, we see that at small δ, instead of providing an upper limit, the theoretical curve underestimates P_{ext}. Our theoretical development considers that the evolutionary rescue owes to the occurrence of one-step rescue mutations (OSMOND; OTTO; MARTIN, 2020), but the assumption is not so critical. The occurrence of two or multiple-step rescue mutations remains at low probability despite its increase with the stress level, as shown in Figure S3. We also observe that the likelihood of multiple-step rescue mutations grows with the number of traits n but still remains as non-dominant events. The larger contribution

of double or multiple mutations to the rescue with n occurs because the probability that a mutation is beneficial and its fitness gain decreases with the number of traits n (ORR, 2000).

The occurrence of double or multiple mutants may also have a prominent role when δ is small. However, because in the range of small δ, rescue mutations might confer small phenotypic effects, the existence of double mutations is more likely to have a harmful effect on fitness. As already pointed out, the fraction of beneficial mutations decreases as one approaches the phenotypic optimum (WAXMAN, 2006; RAM; HADANY, 2015). Therefore, once the population resides close to the phenotypic optimum and a beneficial mutation is carried by an individual, i.e., and so approaching even further the phenotypic optimum, the probability that a second mutation has a detrimental effect will be enhanced.

Figure 13 – Extinction probability as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. The vertical line delimits non-extinction and rescue/extinction domains. The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points.



Source: The author (2024).

In Fig. 13, we explore the dependence of the extinction probability on the stress level δ and the dimensionality of the phenotypic space n. We observe an evident qualitative and quantitative behavior change at the critical $\delta = \delta_c$. By increasing δ, we move from the domain of non-extinction to the domain in which the probability of extinction is non-null. For large δ, the extinction probability will eventually level off to one. The width of this transition shrinks with the number of traits n. The theoretical prediction, Eq. (2.27), works well within

the limitations of the framework for small and intermediate n but breaks down for large dimensionalities. Figure S1 shows results up to $n = 10$, and the prediction of the approximation, which is supposed to work as an upper bound, is satisfactory. However, for $n = 20$ and $n = 30$, we already observe a considerable discrepancy (data not shown).

Degree of parallelism

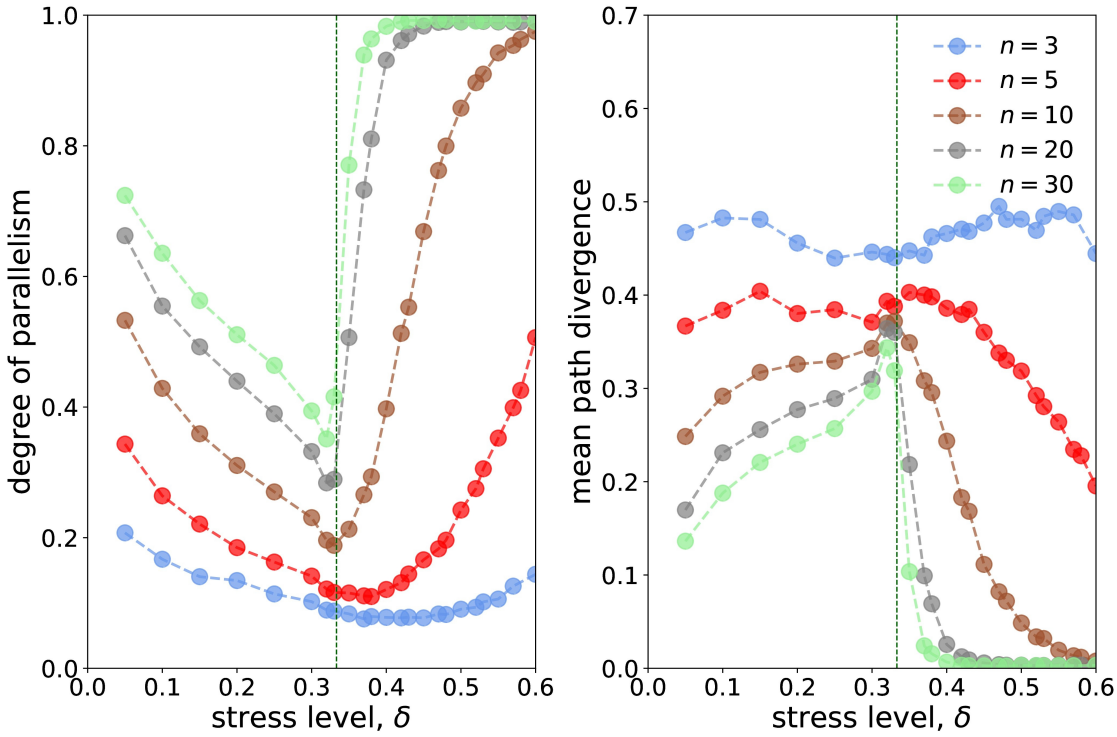
The previous developments about the probability of evolutionary rescue and extinction of populations will be helpful in order to understand better the underlying processes that drive the evolutionary process and affect the degree of parallelism of independent populations.

Initially, both population replicates are isogenic and their sizes equal the carrying capacity K, which is assumed to be the same, i.e., $K_A = K_B = K$. At $t = 0$, the phenotypic optimum is placed at a distance such that the initial fitness of every individual equals $1 - \delta$. The phenotypic optimum is the same for both populations.

To quantify the degree of evolutionary parallelism, we keep track of the evolutionary trajectory of each population and utilize the metrics defined in Section 2.4.1. The left panel of Figure 14 exhibits the degree of parallelism versus the initial population's stress level, δ. Results are shown for distinct values of the number of traits n. As we can see, the larger the n, the larger the degree of parallelism in all the range of δ. On the other hand, the right panel of Figure 14 exhibits the dependence of the mean path divergence versus δ. The measurement is taken by calculating the distance $\Omega(\phi_k^A, \phi_k^B)$, as defined in Eq. (2.7), between the simultaneous evolutionary paths in replicates A and B. There is an evident negative correlation between the degree of parallelism and the mean path divergence. Thus, the larger the number of traits, the smaller the genetic divergence between the evolutionary trajectories. Path divergence is generally negatively correlated to repeatability measures (REIA; CAMPOS, 2020).

From both plots, it is noticeable a change of behavior at the critical point $\delta_c = 1 - 1/W_{max}$, that delimits the domains of non-extinction and rescue/extinction. Except for a small trait number, the degree of parallelism decreases with δ, reaches a minimum at $\delta = \delta_c$, and then features an abrupt rise at the domain of rescue/extinction. Conversely, when looking at the mean path divergence, $\delta = \delta_c$ is the point at which the path divergence reaches a maximum for intermediate and large n. The change of behavior resembles those observed in systems

Figure 14 – Parallelism and mean path divergence as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. From the green vertical line onwards, the population is in a regime of rescue/extinction. The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points.



Source: The author (2024).

displaying a phase transition. This behavior change turns even more abrupt as the number of traits is augmented. Similar qualitative behavior is seen for the mean path divergence.

The characterization of the degree of parallelism in the first regime, the domain of non-extinction, which is delimited by $\delta < 1-1/W_{max}$, is more easily understood. As well known, the probability that a mutation is beneficial decreases with its phenotypic effect within the FGM formulation (HARTL; TAUBES, 1996). Because the probability of parallelism decreases with the number of available beneficial mutations (ORR, 2005b), the drop in the degree of parallelism, as observed here, concurs with previous findings. By augmenting δ, one increases the chance of a mutation being beneficial. Raising δ also decreases the population size in which a positive growth rate is restored, as one can infer from Eq. (2.6). Consequently, this will reduce the supply of beneficial mutations and diminish the occurrence of clonal interference, in which rare large-effect mutations get established (GERRISH; LENSKI, 1998; CAMPOS; WAHL, 2009; BAILEY et al., 2017). Also, higher stress levels may allow mutations of small effects to freely accumulate

independently and not be shared between the two populations in the regime of non-extinction. In this model, where mutation-effect sizes follow an exponential distribution, the presence of many small-effect mutations in the genome could contribute to the decrease in parallelism. Altogether, these factors contribute to the reduction of parallelism with the increase of stress level δ in the domain of non-extinction. A rise in the mean path divergence and mean number of steps toward the optimum results in a decrease in parallelism.

The second regime establishes a clear connection between parallel evolution and evolutionary rescue. When multiple populations or species face a similar threat, those independently evolving traits conferring resistance or resilience are more likely to undergo evolutionary rescue (GOMULKIEWICZ; HOLT, 1995; CARLSON; CUNNINGHAM; WESTLEY, 2014). The independent evolution of similar traits in different lineages can contribute to the overall success and survival of species or populations in the face of environmental threats. The abrupt increase in the degree of parallelism follows a similar rise in the probability of extinction. The effect of increasing stress levels δ is two-fold: large δ means a fast decline in population size, leading the population to more promptly reach critical small sizes (CARLSON; CUNNINGHAM; WESTLEY, 2014), but also signifying that fewer opportunities will be available to generate rescue mutations and restore positive growth rates. The evolutionary rescue upon high levels of stress relies on the existence of a few mutations of large phenotypic effect. This problem is even more evident when the number of available loci is small. In fact, for small genome sizes L, there is a chance that none of the L one-step mutations are rescue mutations (WAHL; CAMPOS, 2023). Therefore, evolving under conditions of severe environmental stress contributes to the increase in repeatability and parallelism at the genotypic level.

2.4.4 Discussion

Together, an integrated study of parallel evolution alongside evolutionary rescue is essential to cope with empirical investigations on rapid evolution, where fewer genes are responsible for adaptive responses to stress. Replicate experiments with microbial systems addressing population resistance to antibiotics (WEINREICH; WATSON; CHAO, 2005; GERSTEIN; LO; OTTO, 2012), salinity concentration (BELL; GONZALEZ, 2011), and temperature (HUANG et al., 2018), are proper systems for comparing our theoretical predictions since in the controlled-lab conditions

the severity of stress can be tuned, whole-genome sequencing is achievable, and factors that affect the per-capita growth rate can be controlled.

As aforesaid, the existence of a mechanism of density regulation leads to observing two distinct regimes, with consequences for the evolutionary outcome and the evolving level of parallelism. In the former, the stress level, quantified by the level of maladaptation, is such that the population can restore positive growth rates by adjusting its population size or through adaptive mutations. In this regime, the population has a null chance of extinction. In the second identified regime, dubbed the rescue/extinction regime, the level of maladaptation is sufficiently high such that the population will be doomed to extinction unless a rescue mutation occurs (GOMULKIEWICZ; HOLT, 1995; ALEXANDER et al., 2014; BELL, 2017).

Notably, we have characterized these two regimes by studying the dependence of the probability of rescue/extinction on the stress level δ. We have developed an analytical approach to estimate the likelihood of rescue/extinction that matches our simulation for geometrically declining populations, i.e., non-density regulated. However, those approximations are also helpful in providing upper and lower bound estimates for the density-regulated case.

Our formalization and analytical approximations are not broadly applicable, limited to studies of parallelism and evolutionary rescue at lower levels of biological organization, where protein, gene, and mutations are expected to constrain the number of possible paths for the adaptive process. In our model, asexual populations adapt solely through *de novo* mutations, but higher levels of parallelism from standing genetic variation are predicted in theory (BURKE; LITI; LONG, 2014) and observed in experiments (MACPHERSON; NUISMER, 2017). Expanding the methods to characterize mutational paths to include standing variation, sexual recombination (THOMPSON; OSMOND; SCHLUTER, 2019), and epistasis are all clear avenues for developing our implications on their roles in genetic parallelism.

- The main results presented in this section have been published in the article: *Freitas, Osmar, and Paulo RA Campos."Understanding Evolutionary Rescue and Parallelism in Response to Environmental Stress." Evolution (2024).* Ref. (FREITAS; CAMPOS, 2024b)

2.5 Epistasis and Incongruence

Epistasis refers to the phenomenon whereby the genetic background (the set of preexisting mutations) influences a mutation's effect on fitness and, thus, on adaptive evolution. In particular, it may also influence a population's fate by constraining the availability of genetic mutations that can lead to rescue in a declining population. Understanding the influence of epistasis in the process of Evolutionary Rescue is the matter of the current contribution.

In this section, we implement the FGM and NK models as a genotype $\rightarrow$ phenotype $\rightarrow$ fitness map to study the conditions for the evolutionary rescue of an asexual population subject to an abrupt environmental change. The motivation for this three-layer mapping is the control of epistasis as a free parameter. We compare individual-based simulations with analytical approximations when possible. We have found that higher levels of epistasis increase the incongruence or non-linearity between maps and determine the fraction of available mutations that can lead to rescue. These results show that landscape topographical properties - influenced by epistatic interactions - play an essential role in population persistence, showing the importance of integrating the genetic bases of adaptation in studying the problem of evolutionary rescue.

2.5.1 Kauffman's NK Model

The NK landscape model of random epistatic interactions was introduced by Kauffman et al. (KAUFFMAN; WEINBERGER, 1989) to study the influence of landscape ruggedness on adaptive evolution. It resembles spin-like models of disordered magnetic materials, common in solid-state physics, where each of the atoms spins can assume one of two possible states $\{\pm 1\}$ (DERRIDA, 1981). Likewise, each organism in the population is represented by a binary sequence, and both systems may exhibit frustration. For each gene (locus) in the genotype, one evaluates its interactions with other genes to estimate its contribution to the phenotype. In the model, K determines the level of interaction between genes and can be adjusted as a parameter to change the degree of ruggedness of the landscape.

As a model for the genotype-phenotype mapping, we consider the NK landscape model (KAUFFMAN; LEVIN, 1987; MACKEN; PERELSON, 1989). The model determines an individual's phenotype from its genotype, denoted as S. Each individual is characterized by a genome

consisting of L loci, represented as $S = (s_1, s_2, \ldots, s_L)$, where each s_α is a binary variable $s_\alpha \in \{0, 1\}$. Therefore, the genotype space consists of 2^L sequences.

The contribution of a given locus j to the phenotype, denoted as ω_j, is a function determined by $\omega_j = g[s_j, \prod(j)]$, relying on the state s_j of locus j and the state of a set of K neighbors, denoted as $\prod(j)$. The K neighbors are selected randomly from the remaining $L - 1$ loci. Therefore, each ω_j has 2^{K+1} distinct arguments randomly drawn from a uniform distribution, defined in the $[0, b]$ range. Finally, the phenotypic value z is estimated as

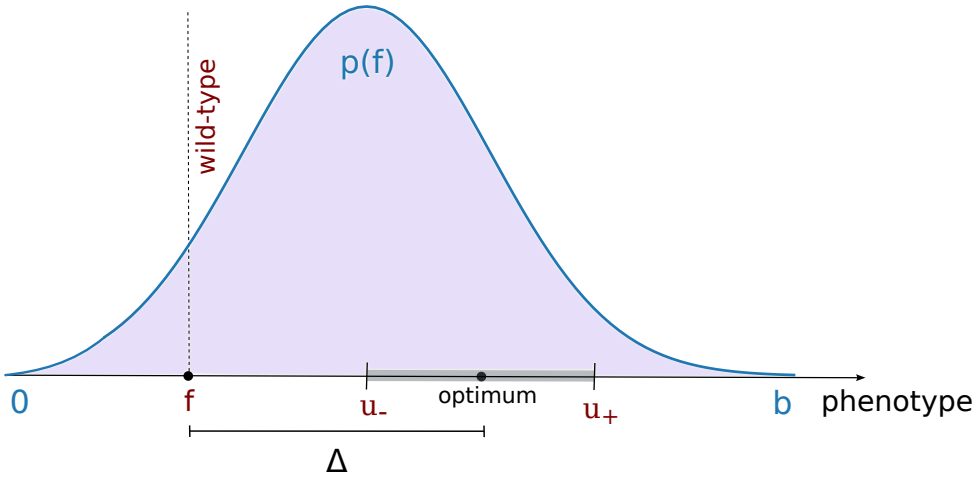
$$z = \frac{1}{L} \sum_{j=1}^{L} \omega_j. \tag{2.19}$$

By changing the parameter K, one varies the degree of ruggedness of the genotype-phenotype landscape: from the additive model, which corresponds to $K = 0$ and is also known as Mt. Fuji landscape (AITA; HUSIMI, 1996), to a completely uncorrelated fitness landscape, which is obtained by making $K = L - 1$. The latter is also known as the House-of-Cards, a biological version of the random-energy spin glass model of Derrida (KINGMAN, 1977; DERRIDA, 1981).

An important feature of the NK landscape model is that it incorporates interdependencies, or epistatic interactions, among the genome's loci. If an allele at a specific locus changes, it impacts not only the fitness contribution of that locus but also the contributions of loci that are epistatically coupled to it. Epistasis is modulated by the parameter K.

Epistasis is directly related to the fitness landscape's topography (BANK, 2022). If mutations have an effect independent of its genetic background (no epistasis), the landscape bears a single adaptive peak. Thus, an optimizing combination of mutations carries the single fittest solution to population survival. In turn, incrementing epistasis leads to increased fitness peaks, giving rise to a rugged landscape i.e. multiple local optima. In most of the applications of the NK model, phenotypic values are used as a proxy for fitness (FREITAS; WAHL; CAMPOS, 2021; REIA; CAMPOS, 2020; KAUFFMAN; LEVIN, 1987). However, here, one assumes an additional layer, i.e., genotype $\rightarrow$ phenotype $\rightarrow$ fitness, and the mapping from genotype to fitness is no longer direct (HWANG; PARK; KRUG, 2017; CIRNE; CAMPOS, 2022). In this multi-layer mapping the number of local fitness peaks may differ between the genotype-phenotype map, and the phenotype-fitness map. Like in Srivastava and Payne (2022), these inherent discrepancies or differences in the topographical properties are referred to as incongruence.

Figure 15 – Illustration of the modelling of evolutionary rescue. In this illustration, the wild-type population has phenotype f by randomly picking a sequence in the genotype space. The optimum is placed a distance Δ from f, such that the fitness of the wild-type population equals $1-\delta$. The interval (u_-, u_+) delimits the phenotypic domain in which the population can restore a positive growth rate $(W > 1)$. Superimposed on the illustration is a drawing of the phenotype distribution, which is a Gaussian distribution.



Source: The author (2024).

2.5.2 Simulation Protocol

Similar to the previous work, an initial isogenic population of size N_0 is set up to simulate evolutionary rescue following an abrupt environmental change. The genotype of the initial population, S_0, with corresponding phenotype z_0, is randomly chosen among the 2^L genotypes that comprise the genotypic space.

At the onset of the simulation, time $t = 0$, a sudden shift occurs in the environment, leading to the relocation of the phenotypic optimum to a position Δ away from z_0, as illustrated in Fig.15. The phenotypic displacement Δ is calculated such that the population's fitness at time $t = 0$ amounts to $1 - \delta$, where δ is described as the initial population's stress level or the level of maladaptation. Notice though, that under density independence, the initial phenotypic distance to the optimum at fixed δ is

$$\Delta = \sqrt{-2 \ln \left[\frac{1-\delta}{W_{max}} \right]}. \tag{2.20}$$

In the absence of density regulation and upon the inexistence of rescue mutations, the population is expected to shrink geometrically and go extinct.

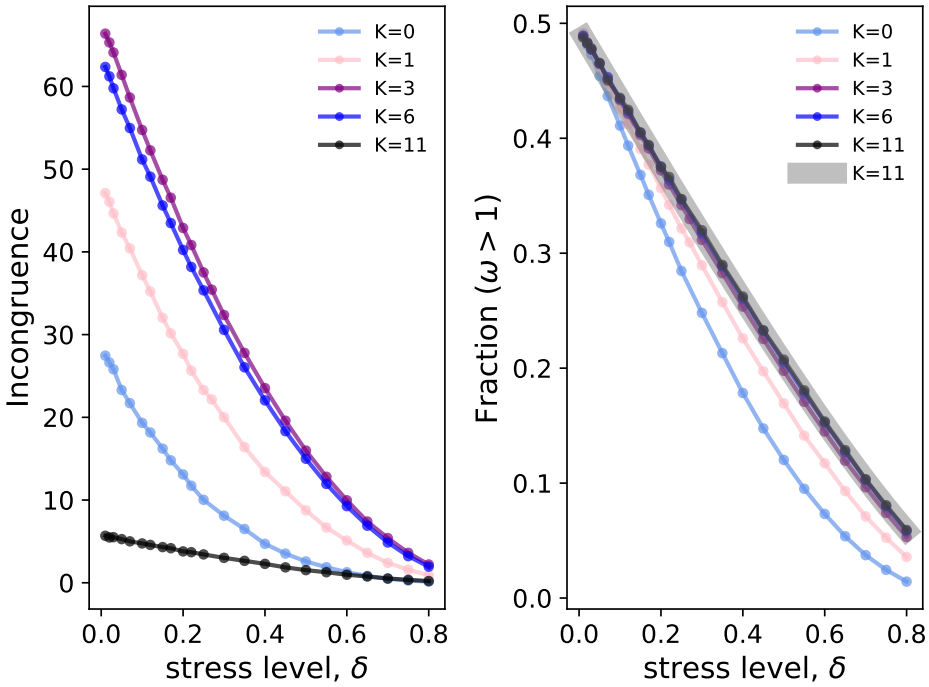
To estimate the probability of extinction/rescue, we let the population evolve up to its extinction or when positive growth rates have been restored. In the latter situation, the simulation stops as soon as the number of individuals of fitness larger than one, $W(z) > 1$, exceeds

200. The results are not sensitive to this particular choice. The probability of rescue is just a fraction of the simulations in which rescue has been found.

Additionally, we analyze the topographical properties of the resulting fitness landscape. The former quantity is the fraction of genotypes in the genotypic space with fitness values larger than one, $W(z) > 1$, meaning positive growth rates. This quantity is of great interest as this fraction directly estimates the number of potential rescue mutations. The second quantity is incongruence, as first presented in Ref. (SRIVASTAVA; PAYNE, 2022). The incongruence is then quantified as the average of the absolute change in the number of peaks $\langle |p_f - p_{gp}| \rangle$, where p_f and p_{gp} are the number of peaks in the fitness landscape and genotype-phenotype map, respectively.

2.5.3 Results

Figure 16 – Incongruence and the fraction of potential rescue mutations as a function of the stress level. Curves denote different degrees of epistasis K. The thick solid line is the analytical result for the uncorrelated case, $K = 11$, obtained from Eq. (2.23). Error bars are smaller than points. Parameters for both cases are $L = 12$, $W_{max} = 1.5$, and $b = 6$.



Source: The author (2024).

Topographical properties

Here, we investigate how the stress level, quantifying the initial phenotypic distance from the wild-type population to the optimum, shapes the fitness landscape. First, it is essential to highlight that as opposed to the original phenotypic Fisher's geometric model (FGM), which displays a single optimum, its genotypic counterpart has multiple local optima due to the discreteness of the genotype space (HWANG; PARK; KRUG, 2017). In the left panel of Fig. 16, we explore the dependence of the incongruence on the stress level, δ. We verified that the number of local optima of the fitness landscape is greater than that observed for the NK landscape model and, as we see from the plot, reaches a maximum at intermediate values of the epistatic parameter K. However, we also observe that the incongruence is significantly reduced as the stress level is augmented, making the fitness landscape smoother at more severe stress levels.

In the right panel of Fig. 16, the fraction of potential rescue genotypes is shown as a function of δ, i.e., those genotypes that can effectively restore positive growth rates. As expected, it is a decreasing function of δ. Except for $K = 0$ and $K = 1$, the curves collapse and coincide with the theoretical prediction, to be discussed below, showing that sufficient epistasis increases the availability of rescue mutations.

For our theoretical prediction, note that the phenotype associated with a given genotype is a sum of L contributions, one for each locus, which, in turn, are random values taken from a uniform distribution $U(0, b)$, with mean $\bar{f} = b/2$ and variance $\sigma^2 = Var[f]$ given by

$$\begin{aligned} Var[f] &= \overline{f^2} - \overline{f}^2 \\ &= \int_0^b \cancel{p(f)}^{1/b} f^2 \, df - b^2/4 \\ &= b^2/12. \end{aligned}$$

Following the Central Limit Theorem (REICHL, 2016), the probability distribution of phenotypic values for the uncorrelated landscape is a Gaussian with the same mean and variance $\sigma_f^2 =$

σ^2/L, which is easily obtained by taking the variance of the mean:

$$\begin{aligned} Var[\bar{f}] &= Var\left[\frac{1}{L}\sum_{i=1}^{L} f_i\right] \\ &= \frac{1}{L^2}\sum_{i=1}^{L} Var\left[f_i\right] \\ &= \frac{1}{L^2} L\, Var\left[f\right] \\ &= \frac{b^2}{12L}. \end{aligned}$$

Therefore, a Gaussian distribution of mean $b/2$ and variance $\sigma_f^2 = \frac{b^2}{12L}$, holds irrespective of the value of the epistatic parameter K:

$$p(f) = \sqrt{\frac{6L}{\pi b^2}} \exp\left(-\frac{6L(f-b/2)^2}{b^2}\right), \tag{2.21}$$

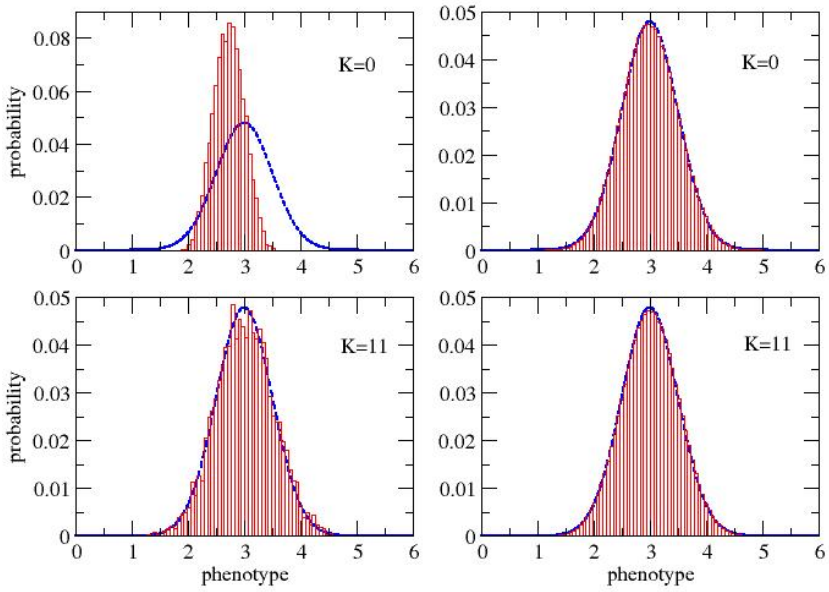
As one samples a single phenotypic distribution, when the phenotypic values are low correlated, as found for intermediate and large K, the distribution of phenotypic value for this sample already approaches the distribution above (Eq. 2.21). But this is no longer true for small K, such as $K = 0$ and $K = 1$, where we have a very high correlation and the distribution for one sample is more likely to explore a smaller domain of possible phenotypic values. This situation is depicted in Figure 17, where we show the distribution of phenotypic values for epistatic parameters $K = 0$ (additive landscape) and $K = 11$ (uncorrelated landscape) for single and multiple samples. The distributions on the left correspond to one sampling of the phenotypic landscape, whereas on the right, one has the resulting distribution over $1,000$ distinct samples of the phenotypic landscape. The dashed lines are the theoretical distribution given by Eq. (2.21). This evidence explains why the fraction of potential rescue genotypes for low K do not collapse with those for intermediate and large K in Fig. 16.

Although analytical expressions are generally troublesome, some derivations are possible in the particular case of the uncorrelated fitness landscape, as the following shows.

An approximation for the case $K = L - 1$

The wild-type population has a phenotype value provided by Eq. (2.21), as its genotype is randomly chosen among the 2^L sequences of the genotype space. In the uncorrelated case, any mutant will also have its phenotype value, $f^{'}$, taken from the same distribution, and therefore, the mutation phenotypic effect will be $u = f^{'} - f$. Because f and $f^{'}$ are both Gaussian

Figure 17 – The distribution of phenotypic values over the entire genotype sequence. The epistatic parameter value is K = 0 (upper panels) and K = 11 (lower panels). In the left panel, the distribution is obtained from a single sample of the phenotypic landscape, whereas in the right panel, the distribution is taken from an ensemble of 1, 000 samples of the phenotypic landscape



Source: The author (2024).

variables, u will also be given by a Gaussian distribution of null mean and variance $2\sigma_f^2$, i.e.,

$$p(u) = \sqrt{\frac{3L}{\pi b^2}} \exp\left(-\frac{3Lu^2}{b^2}\right). \tag{2.22}$$

Not every mutation can potentially steer the population toward the rescue domain. The probability that a random mutation is a potential rescue mutation is simply

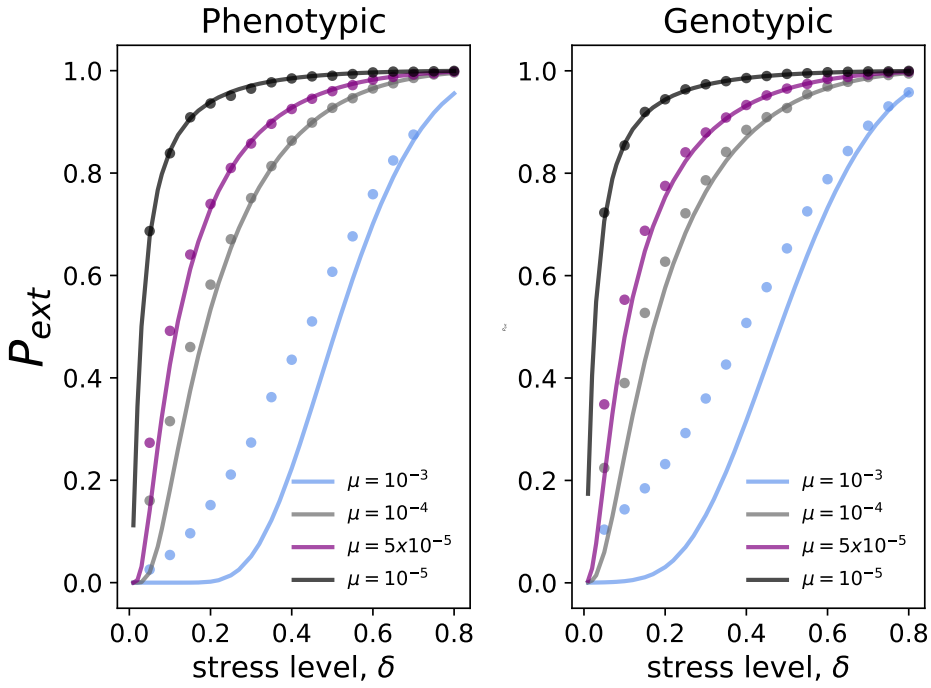
$$\rho = \int_{u_-}^{u_+} p(u)du, \tag{2.23}$$

where the condition

$$W(u) = W_{max} e^{-\frac{1}{2}(\Delta - u)^2} = 1 \tag{2.24}$$

provides the upper and lower bounds of the integral, i.e., within this range, positive growth rates (fitness larger than one) are ensured. The solutions of Eq. 2.24 are $u_\pm = \Delta \pm \sqrt{2 \ln W_{max}}$. In the right panel of Fig. 16, the theoretical prediction for the fraction of potential rescue mutations, Eqs. (2.23) and (2.22), is plotted along with simulation results, where a good agreement is found, except for small K for the reasons outlined in the previous section.

Figure 18 – Probability of extinction for the uncorrelated landscape, $K = 11$, as a function of the stress level for the (left panel) phenotypic and (right panel) genotypic versions of the model. Curves represent distinct values of mutation probability μ. Solid lines are the analytical approximations, Eqs. (2.25) (phenotypic model) and (2.27) (genotypic model). Error bars are smaller than the points. Parameter values are $N_0 = 10^4$, $W_{max} = 1.5$, $b = 6$, and genome size $L = 12$ (genotypic model).



Source: The author (2024).

As already highlighted, only mutants of fitness larger than one can contribute to rescuing the population. As in our previous work, the probability that a randomly chosen mutation rescues the population is again given by Eq. 2.9:

$$\bar{\pi} = \int_{u_-}^{u_+} p(u)\pi(u)du.$$

As already derived, the probability of population extinction for the phenotypic case is

$$P_{ext} = (1 - \bar{\pi})^A, \tag{2.25}$$

where A is the total number of mutants (ORR; UNCKLESS, 2008). As for the genotypic version, L potential rescue mutations exist. Depending on the mutation probability μ and the rate of population decline $1 - \delta$, these L one-step mutations can repeatedly occur up to the population's extinction or rescue. The chance that a single potential rescue mutation arises i times without getting established and rescuing the population is $[1 - \pi(u)]^i$. As in the previous sections, we assume that *de novo* mutations follow a Poisson distribution with an average of

A/L, so all loci are equally likely to be mutated. The likelihood that a single potential rescue mutation effectively rescues the population is determined as follows:

$$R = 1 - \frac{1}{\rho} \int_{u_-}^{u_+} p(u) \exp\left(\frac{-\pi(u)A}{L}\right) du. \tag{2.26}$$

with ρ given by Eq. (2.23). Note that $R \times \rho$ provides the likelihood a potential rescue mutation gets established. Because there are L one-step mutations, the probability of extinction of the population is calculated as

$$P_{ext} = (1 - R\rho)^L. \tag{2.27}$$

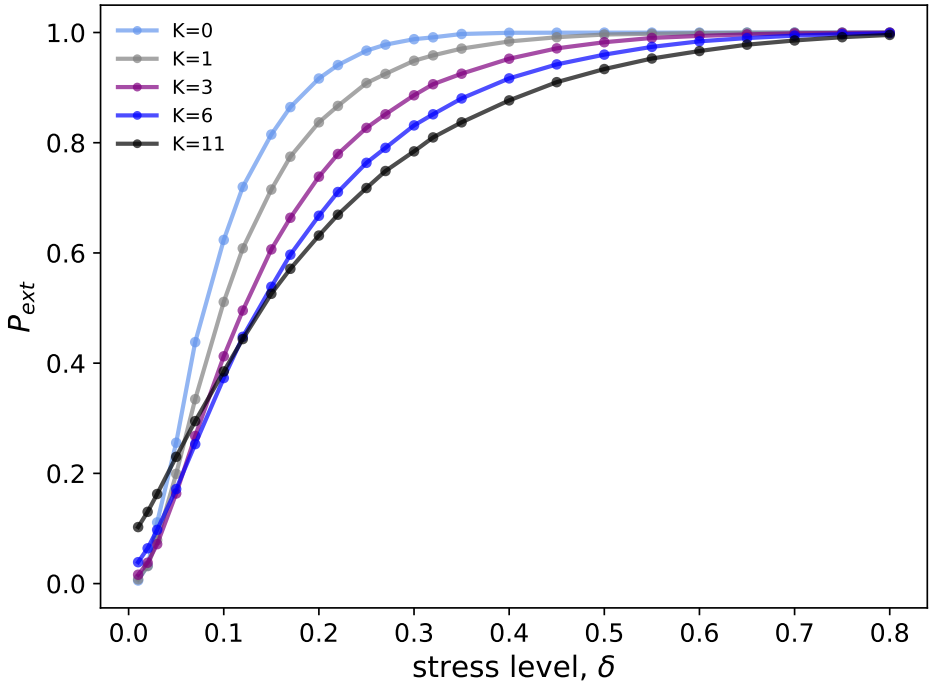
In Figure 18, we show the probability of extinction as a function of the stress level, δ, for a population of initial size $N_0 = 10,000$ and distinction values of the mutation probability μ. For the genotypic model, we have considered genome size $L = 12$. We confront the simulation results with the theoretical prediction, Eq. (2.25) and Eq. (2.27). The theory fits reasonably well the simulation data for low and intermediate μ but not for large mutation probabilities. We hypothesize that the inconsistency at large μ is because of the interference among coexisting mutations (BARRICK; LENSKI, 2013), whose competition hinders fitter variants from increasing in size. The probability of extinction displays a sharp increase with δ at small μ, but the rise of P_{ext} is smoothed with the augmentation of μ.

General case

So far, we have addressed the case of the uncorrelated fitness landscape, which has allowed us to derive analytical approximations and get critical insight into the problem. In Fig. 19, we show the probability of population extinction versus the stress level δ and distinct epistatic parameter values.

The likelihood of extinction is a monotonic increasing function of δ, which holds for all K. Counterintuitively, we notice that, except for very small δ, smaller extinction probabilities occur at larger K. Usually, when we think about the role of turning the fitness landscape more rugged, the evolutionary process becomes more constrained, and more easily, populations get trapped at local fitness peaks, which at first glance could mean some difficulty in establishing positive growth rates. However, an opposite scenario is observed. The reason larger K means an enhanced chance of rescuing the population is two-fold: first, the phenotypic effect of

Figure 19 – Probability of extinction as a function of the stress level for distinct values of the epistatic parameter K. Error bars are smaller than points. Parameter values are $L = 12$, $\mu = 1 \times 10^{-4}$, $N_0 = 10^4$, $W_{max} = 1.5$, and $b = 6$.

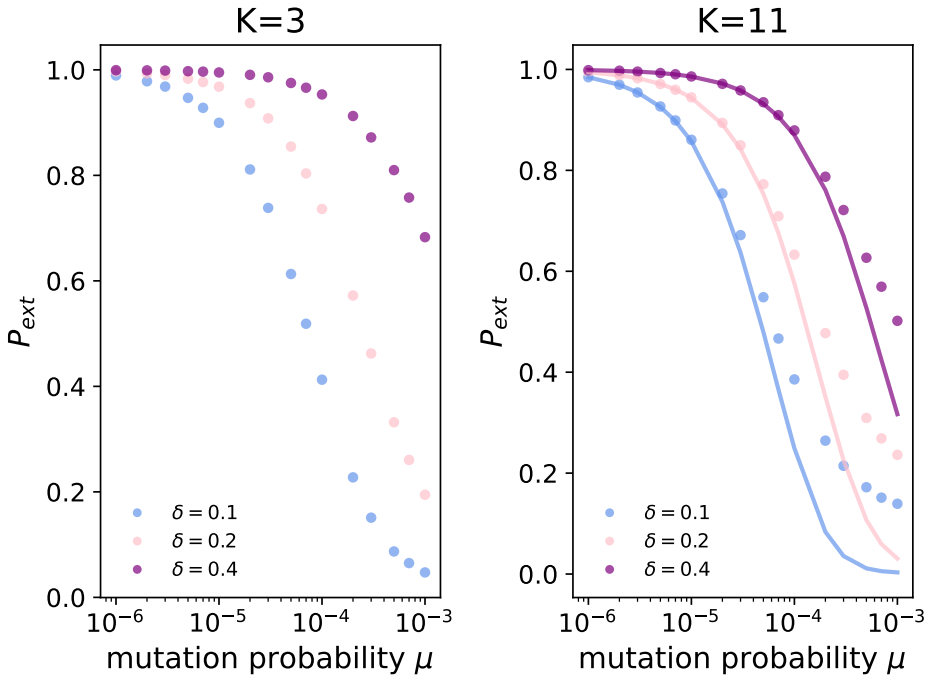


Source: The author (2024).

mutations u is proportional to $(K+1)/L$ (CAMPOS; ADAMI; WILKE, 2002). If the evolutionary process of rescue relies mainly on few-step mutations, then a larger K would ensure a higher chance of producing a phenotypic displacement to move the population for the rescue domain $[f+u_-, f+u_+]$ (see Fig. 15). In other words, on average, there exist local solutions to rescue. Second, the main advantage of a smoother landscape (small K) in a genotype $\rightarrow$ fitness map relies on establishing multiple-step mutations as the population can more easily sweep through the landscape. However, because the three-layer mapping of genotype $\rightarrow$ phenotype $\rightarrow$ fitness leads to incongruence, an increase in the number of peaks in the fitness landscape compared to its phenotypic counterpart curbs much of its advantage in terms of diffusion over the genotypic space through mutations.

We also observe that the different behavior at very small δ, in which lower extinction probabilities do not ensue from very large K, such as $K = 11$, is amplified at large μ. Indeed, this abnormality can be understood in the face of the fundamentals of Fisher's geometric model. When the distance to the optimum phenotype is narrow, mutations of sizable effect are likely to have a detrimental effect (TENAILLON, 2014; RAM; HADANY, 2015), i.e., they can fall outside the rescue domain $[f+u_-, f+u_+]$.

Figure 20 – Probability of extinction as a function of the mutation probability. Curves represent different stress levels δ. Error bars are smaller than points. Solid lines are analytical approximations for the full uncorrelated case ($K = 11$), Eq. (2.27). Parameter values are $L = 12$, $N_0 = 10^4$, $W_{max} = 1.5$, and $b = 6$. The epistatic parameter K is indicated in the subtitle of the panels.



Source: The author (2024).

Figure 20 now explores the dependence of the extinction probability P_{ext} on the mutation probability μ. The extinction probability is a monotonic decreasing function of μ. The larger the stress level δ, the larger the mutation probability μ that characterizes the onset of the rescuing regime. Results are qualitatively the same for $K = 3$ (left panel) and $K = 11$ (right panel). We contrast the simulation data in the right panel with the theoretical prediction, Eq. (2.27). Note that the theoretical predictions fit the simulation data over a broad range of mutation probability μ, which is widened with δ. As mentioned before, the theory ceases to predict the simulation data well when coexistence and interference among mutations exist. An increased δ entails a swift population decline, thus reducing A, the number of mutations up to population extinction. This is why the theoretical prediction is suitable over a broader range of μ with the augmentation of δ.

2.5.4 Discussion

Understanding the role of epistasis in evolutionary rescue is crucial because it can significantly influence the outcomes of rescue attempts, as well as the long-term viability of rescued populations. In this context, the role of epistasis is addressed here as the genotype-phenotype map, provided by the NK landscape model, proposed by Srivastava and Payne (2022).

Our results show that the degree of epistasis can significantly impact the fate of endangered populations. Remarkably, we have shown extinction probabilities are larger when smoother genotype-phenotype maps are used. The reason for this outcome is two-fold: first, the phenotypic effect of mutations is proportional to $(K+1)/L$, and hence, high epistasis can provide sufficiently large mutation effects, thus ensuring the population's rescue. Second, the whole mapping genotype-phenotype-fitness leads to incongruence in the landscapes, making smooth landscapes at the phenotypic level more rugged at the fitness level. Thus, it curbs the advantages of smooth landscapes in an evolutionary context, that is, the accessibility of beneficial mutations that can accumulate gradually and allow populations to move freely across the landscape.

From a mathematical perspective, we derived analytical results for the fraction of potential rescue mutations, i.e., those mutations that can restore positive growth rates. More importantly, we developed a formulation to analytically estimate the probability of extinction in the uncorrelated model version that applies to the case of $K = L - 1$. We also introduced a phenotypic version of the model, which we used to give us some insight into the derivations for the genotypic model.

- The main results presented in this section have been published in the article: *Freitas, Osmar, and Paulo RA Campos. "The role of espistasis in evolutionary rescue." The European Physical Journal E (2024).* Ref. (FREITAS; CAMPOS, 2024a)

2.6 Conclusions

Understanding the eco-evolutionary process known as evolutionary rescue is crucial not only in the face of global climate change (PAULS et al., 2013; SOUZA et al., 2023), but equally important in the study of pathogenic microbes, for example providing insights into the issue of antibiotic resistance, a global health challenge (LIN; KUSSELL, 2016; LARSSON; FLACH, 2022). Although primarily studied assuming abiotic sources of demographic change, the study of evolutionary rescue can also provide a proper framework for the role of biotic agents of demographic change, such as introducing a novel pathogen (SEARLE; CHRISTIE, 2021), a topic closely related to the epidemics of virulent pathogens.

Within the framework of our first study, we investigate density regulation effects on evolutionary rescue and parallel evolution at genetic levels (CHEVIN; LANDE, 2010). Although the contribution of population size to parallelism is the subject of many studies, they are usually restricted to populations of constant size. A meta-analysis from (BAILEY et al., 2017) shows an increase in parallelism with population size from bacteria and fungi, while (SZENDRO et al., 2013) reports a non-monotonic dependence when looking at the repeatability of adaptive pathways. Although some theoretical works address these questions in populations with varying sizes (e.g., in bottlenecked populations (FREITAS; WAHL; CAMPOS, 2021)), growth and decline rates remain deterministic. Thus, a clearer formalization of the effects of density regulation on parallel evolution was missing.

The Fisher Geometric Model (FGM) plays a crucial role in understanding evolutionary rescue and has been proven to offer an adequate framework to the changing environments (MATUSZEWSKI; HERMISSON; KOPP, 2014; TENAILLON, 2014; ANCIAUX et al., 2018). Moreover, by investigating the relationship between genotype, phenotype, and fitness, we can gain essential insights into the molecular mechanisms underlying evolutionary processes and the adaptation of organisms to their environments, with evolution driven by genotype changes and selection based on phenotypes (CATALÁN et al., 2023). In this context, in our second study, we address the mapping genotype $\rightarrow$ phenotype $\rightarrow$ fitness, highlighting the effects of epistasis on the map nonlinear transformations and how it can significantly impact the fate of endangered populations.

In conclusion, these couple of works aim to contribute to understanding population persistence and their evolutionary responses in the face of environmental changes. Investigating the impact of different demographic regimes, as other forms of density regulation, could provide valuable insights into the interplay between population dynamics and genetic adaptation. Replicate experiments are giving new insights into genetic parallelism and the constraints promoting the repeatability of evolutionary outcomes, as well as empirical fitness landscapes. By studying these mechanisms, we can better understand how populations adapt and evolve in response to environmental stress and develop more effective conservation and management strategies to ensure their persistence or halt their spread, as with emerging diseases.

3 SPECIATION IN A NEUTRAL MODEL

Tantos povos se cruzam nessa terra
Que o mais puro padrão é o mestiço
Deixe o mundo rodar que dá é nisso
A roleta dos genes nunca erra
(Mestre Ambrósio)

The evolution of life on Earth shows a great diversification of organisms. This inherent feature of life can be observed under different biological scales, from genomics to species to ecological levels of organization. At the population level, geographic isolation and selection pressure are the two fundamental mechanisms behind the emergence and maintenance of diversity. Still, geographic isolation alone is enough to generate a wide range of biodiversity dynamics and patterns (PRINCEPE et al., 2022; COYNE; ORR, 2004; AGUIAR et al., 2009), highly influenced by population's biogeographic history (GAVRILETS, 2014) and spatial structure (CAMPOS et al., 2013; MARQUITTI; FERNANDES; AGUIAR, 2020).

As populations are separated from each other, the stochastic nature of mutation and genetic drift, together with the impossibility of exchanging genes through reproduction (gene flow), leads to differentiation. As a result, previously identical populations will eventually become so distinct that reproduction among them is impossible - this reproductive isolation is a common criterion to characterize speciation (MAYR, 1963; FITZPATRICK; FORDYCE; GAVRILETS, 2009).

The evolutionary history of population differentiation can be represented by a phylogeny: a tree-like structure where branch ramifications account for speciation events and terminal nodes to the species. Together with ecological and evolutionary data, phylogenetic trees are used to study the processes behind diversification (NEE; MOOERS; HARVEY, 1992; MOOERS; HEARD, 1997; BLUM; FRANçOIS, 2006; MORLON, 2014), including those at microevolutionary scale (COSTA et al., 2019). For instance, the distribution of a phylogeny branches length, over time and among lineages, are common topological metrics used to assert variations in the speciation and extinction rates among species (SHAO; SOKAL, 1990; MOOERS; HEARD, 1997; MORLON, 2014), including those promoted by partial isolation (MARQUITTI; FERNANDES; AGUIAR, 2020).

In this work, we theoretically study the macroevolutionary speciation patterns that emerge from the microevolutionary dynamics of populations inhabiting two patches (or islands). Populations evolve due to recombination, mutation, and migration of individuals between patches. The model is neutral, meaning that neither survival nor reproduction depends on a fixed genotype - there is no selection, thus speciation emerges only from isolation. However, individuals must have a minimal amount of genetic similarity in order to reproduce. This dynamic was studied for a scenario of constant migration probability through time (PRINCEPE et al., 2022), and intermittent migration (PRINCEPE et al., 2023), where migration between patches can only occur in specific periods defined by the empirical data of sea level oscillations (SPRATT; LISIECKI, 2016). In addition to the ecological metrics evaluated in previous studies (richness, β-diversity, and species asymmetry distribution) (PRINCEPE et al., 2022; PRINCEPE et al., 2023), here we track each speciation and extinction event, build the complete and extant phylogenies, and characterize the topological patterns in terms of phylogeny balance, speciation acceleration, and age. Our results highlight how the microevolutionary dynamics subject to different isolation regimes can reflect the macroevolutionary patterns.

3.1 Derrida-Higgs Model of Speciation

A number of statistical models of evolving populations in neutral landscapes are widely used in the literature (GAVRILETS, 2004). The interest of this approach lies in the possibility of clarifying evolutionary phenomena by means of methods and concepts developed in statistical mechanics. In this context, the model introduced by Derrida and Higgs (HIGGS; DERRIDA, 1991) is particularly interesting as a simple model containing the concept and dynamics of biological speciation.

In this neutral speciation model, individuals are characterized by a biallelic genome of length G and alleles $\{S_1, S_2, ..., S_G\}$, with $S_i = \pm 1$. For simplicity, individuals are haploid and hermaphrodite. At each generation, M offspring are born, replacing their parental generation. In this way, the population size through time does not vary. The population is characterized by an $M \times M$ matrix Q whose components measure the degree of genetic similarity between each pair of individuals α and β:

$$q^{\alpha\beta} = \frac{1}{G}\sum_{i=1}^{G} S_i^{\alpha} S_i^{\beta}, \quad \text{with} \quad q^{\alpha\alpha} = 1. \tag{3.1}$$

Similarity ranges from completely different ($q^{\alpha\beta} = -1$) to identical pairs ($q^{\alpha\beta} = 1$). Alternatively, the genetic distance between the individuals, measuring the number of genes bearing different alleles, is the Hamming distance $H^{\alpha\beta} = G(1 - q^{\alpha\beta})/2$.

In discrete non-overlapping generations and under random mating, the chance of selecting a partner is simply $1/M$. For large genome size $G >> 1$, one can show (see Appendix B) that, given enough time, the overlaps $q^{\alpha\beta}$ converge to a stationary distribution centered at

$$q^{\alpha\beta}_{t\to\infty} \to q_0 \approx \frac{1}{(1 + 4\mu M)}, \tag{3.2}$$

where μ is the mutation probability per locus at each birth event. This approximation holds for μ and $1/M$ much smaller than one, which is always the case for real populations.

The particularity of the Derrida-Higgs model is that reproduction occurs between pairs of individuals, α and β ($\alpha \neq \beta$), if and only if $q^{\alpha\beta} > q_{min}$, i.e. if parents share a minimum similarity q_{min} of their genome. If $q_{min} > q_0$, the population divides into groups of individuals whose average similarity is larger than q_{min}, i.e. there is no gene flow between groups. Therefore, the Derrida-Higgs model incorporates the definition of species as reproductively isolated groups.

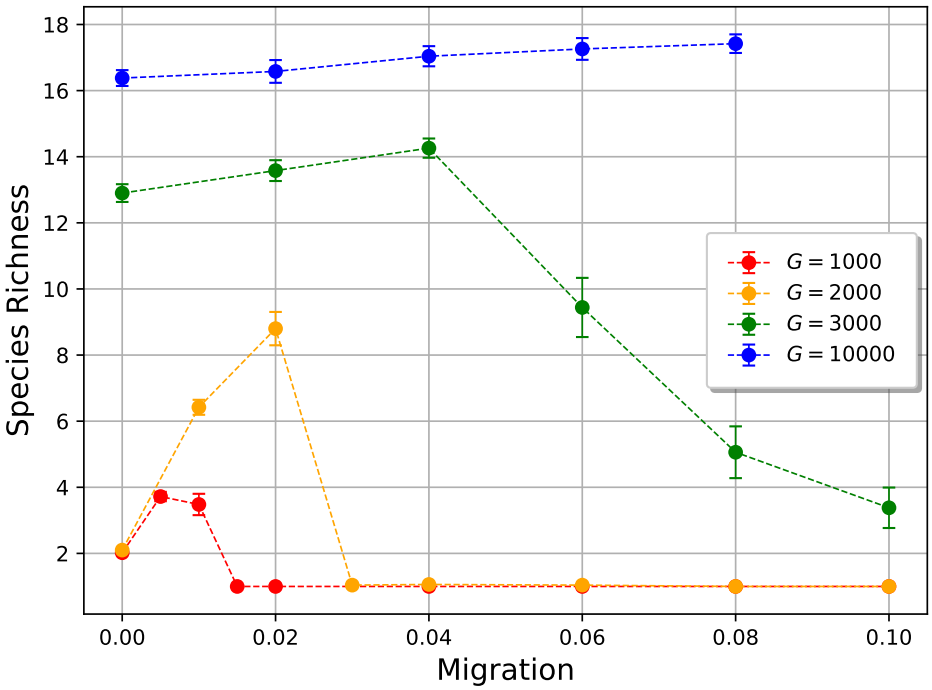
The original model is restricted to sufficiently large genomes and the emergence of *sympatric* speciation - divergence in the absence of geographic isolation. The conditions for sympatric speciation in a finite genome of size G only hold above a critical value of G, which depends on population size, mutation rate and q_{min} (AGUIAR et al., 2009). An extension of the model for geographically structured populations of infinite genome was proposed by Manzo and Peliti (MANZO; PELITI, 1994), and have shown the effects of geographic isolation in the process of allopatric speciation. Recent analytical and numerical progress was made by Principe et al (PRINCEPE et al., 2022) for the case of two insular populations of finite genome, where individuals could be exchanged between patches with migration probability ε.

In this latter case, individuals are characterized by their genomes and by the patches they occupy: q is the similarities between individuals inhabiting the same island, and p for different areas. For ε, μ and $1/M$ all much smaller than 1, similarities also converge to a stationary distribution q_0 and p_0, and the total number os species in the islands, N_T, can be estimated as a function of the expected number of species in each island N:

$$N_T = N\left(2 - \frac{p_0}{q_0}\right). \tag{3.3}$$

The explicit derivations of these results are in Refs (PRINCEPE et al., 2022). [1] In a qualitative analysis, when there is no migration $\varepsilon = 0$ the islands are independent, and one expects that the similarity between individuals of different islands tends to zero, thus $p_0 \to 0$, and $N_T = 2N$. When migration dominates over mutations $\varepsilon >> \mu$, the two islands behave as a single island with double population size, thus $p_0 \to q_0$ and $N_T = N$. This behavior is summarized in Figure 21, for variations in genome size G.

Figure 21 – Example results from Derria-Higgs model on a two-deme population, showing how the total species richness of both islands N_T changes with migration rate ε and genome size G.



Source: The author (2024).

For genome sizes of $G \leq 2000$, the expected behavior of Equation 3.3 is shown. Under complete isolation $(\varepsilon = 0)$, we have one species in each island and $N_T = 2$, and a single species shared by both islands for a higher enough migration rate. Interestingly, low levels of migration increase the diversity of the system.

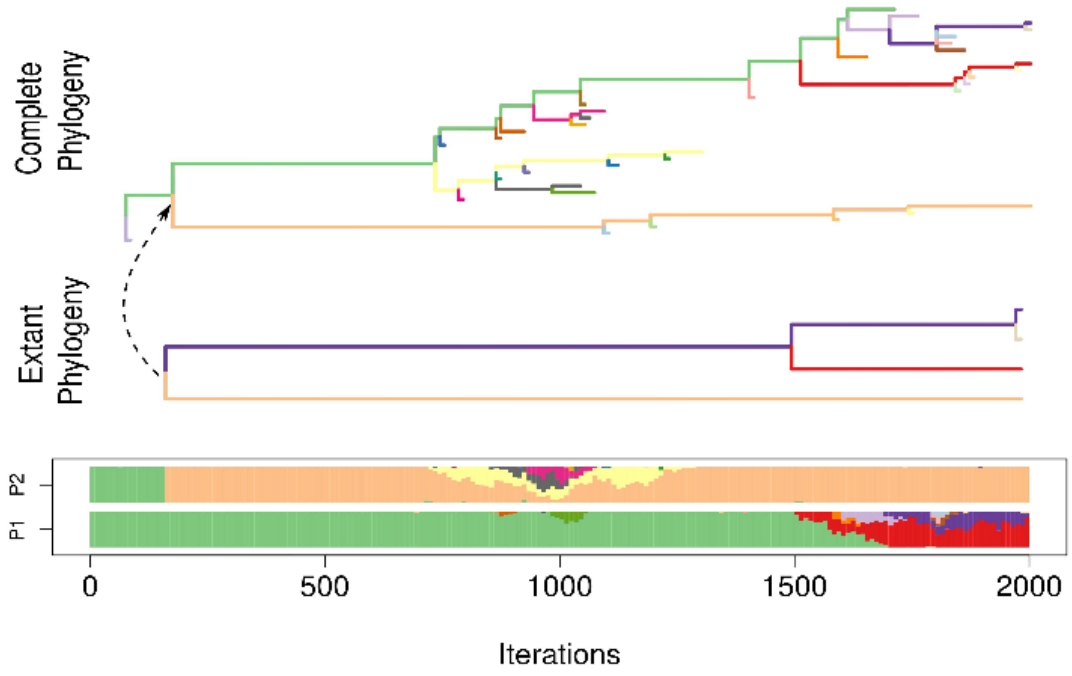
On the other hand, larger genome sizes induce increased variation in the speciation process, with species being generated inside each island even under complete isolation - we call this differentiation process pure sympatric speciation. To study the effects of gene flow on speciation without the biases of such a process, we avoid this regime.

[1] Reproduced in Appendix B of this thesis for completeness.

3.2 Phylogenetic Trees

Accordingly to our previous discussion and the models presented, we define speciation by the impossibility of gene flow between populations. Over time, individuals can accumulate genetic differences until reproductive isolation is reached ($q^{\alpha\beta} > q_{min}$). A speciation event occurs each time a group of compatible individuals splits into subgroups, i.e. no individual is genetically compatible with another individual from another subgroup, breaking the possibility of gene flow between populations. When this happens, each subgroup receives a new species identity (colours in Fig. 22). Every ten generations, one identifies species, recording the time they emerged and went extinct, as well as their ancestral species. This procedure enables us to build the phylogenetic tree. We have analyzed the complete and extant phylogenies: the former includes the sequence of all speciation and extinction events (Fig. 22); in the latter, information about extinct species is missing. We note that other methods for generating phylogenies from individual-based models are known (COSTA et al., 2018).

Figure 22 – Example of the model evolution under constant migration. There are three plots: the upper and middle panels correspond to the complete and extant phylogenies, respectively. The plot at the bottom illustrates species distribution abundance in each patch over time. In each speciation event, the most abundant species maintains its ancestral color.



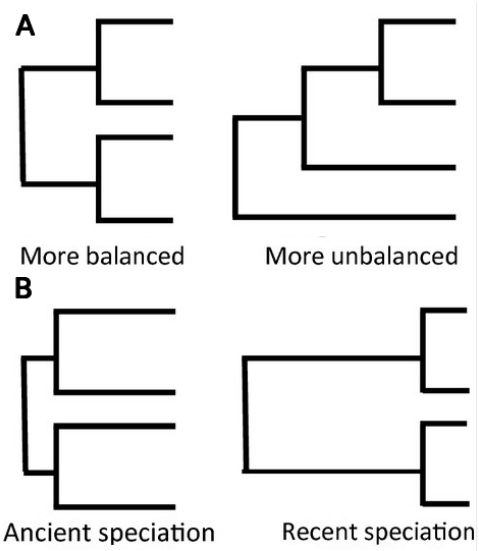
Source: The author (2024).

An overall observation is that much information can be lost when analyzing the phylogeny of extant species instead of the complete phylogeny. In these instances, we observe that recent

speciation events may last a few generations, probably because the similarity of two very similar species falls below the speciation threshold again, allowing the fusion/hybridization into a single species (BOTELHO; MARQUITTI; AGUIAR, 2022) or because the abundance of the new species is so small that it quickly becomes extinct by drift. Here we do not differentiate fusion from extinction and all events in which a species disappears are counted as an extinction event. When three or more speciation events arise from the same species the phylogenies contain polytomies (multifurcation) (MOOERS; HEARD, 1997; HOELZER; MEINICK, 1994) (see, for example, the extant phylogeny in Fig. 27C).

Phylogenies are the result of speciation and extinction events along population history, and its topology contains information about the diversification process. For example, early stochastic models of population differentiation, such as Yule (YULE, 1925) and birth-death models (KENDALL, 1948), utilize a constant diversification rate in their processes. As a result, their phylogenies exhibit a specific tree topology where ramifications are uniformly distributed over time and among branches. On the contrary, if some lineages speciate more than others it affects the phylogeny balance (Fig 23.A). If speciation events are not uniformly distributed in time, this variation clusters the ramifications of the tree (Fig 23.B).

Figure 23 – Exemplification of two topological features of phylogenies that carry information about the diversification process. A) The distribution of ramifications among branches - the phylogeny balance. B) The change in speciation rate over time.



Source: The author (2024).

In our model, lineage speciation and extinction rates are not imposed as parameters but emerge naturally from the population-level dynamics. To search for its signal in the phylogeny, we utilize the topological metrics of tree balance and acceleration of the speciation rate.

Phylogeny balance (J):

Balance measures how much speciation and/or extinction rate vary among lineages. We use the J metric to evaluate the phylogeny balance statistics (LEMANT et al., 2022). It retains direct mathematical relationships with common balance indices (like Sackin's Index (SACKIN, 1972)), but J is a normalized metric that enables meaningful comparison of trees with different numbers of leaves. Furthermore, J is not restricted to bifurcating trees, as its definition addresses polytomies - a ramification in more than 2 branches.

Let $\{\ell_1, \ell_2, ...\ell_n\}$ be the terminal nodes (species) in a tree with n nodes (Figure 24 below). Tree balance is defined as

$$J = \frac{1}{\sum_{k=1}^{n} \ell_k} \sum_{i=1}^{n} \ell_i \, W_i. \tag{3.4}$$

The sum is made over the number of leaves ℓ_i at the end of each internal node labeled i. Let η_i be the number of branches or ramifications directly sprouting from node i, and one define $W_i \in [0, 1]$ as the normalized Shannon entropy of the leaves:

$$W_i = \sum_{j=1}^{\eta_i} W_{ij} = \sum_{j=1}^{\eta_i} -p_{ij} \, log_{\eta_i} \, p_{ij}, \tag{3.5}$$

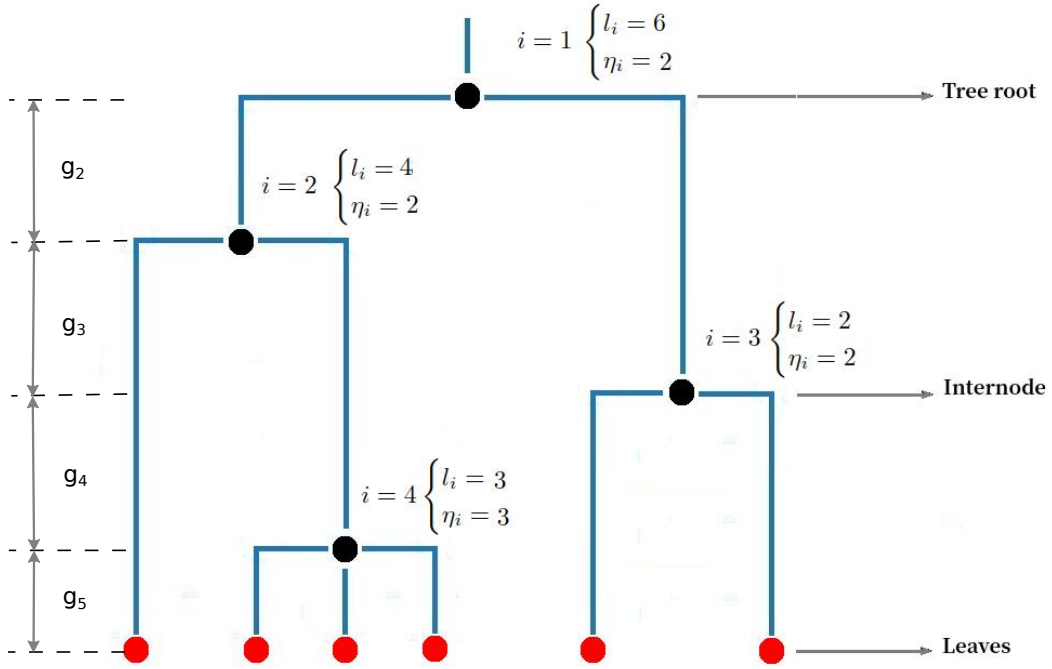
where $p_{ij} = \ell_j / \ell_i$ is the proportion of total leaves at the end of each branch. Note that $\sum_j p_{ij} = 1$, and if all p_{ij} are equal for a given internal node i, all descendants are split into equal subtrees, resulting in $W_i = log_{\eta_i} \eta_i = 1$ for maximum balance. An unequal distribution reduces the balance value, with the limit case $W_i = 0$ if and only if i spawns only one descendant, thus $\eta_i = 1$. Figure 24 shows an example of the parameters used to calculate J from a sample tree.

Acceleration of speciation (α):

Branch lengths represent the time between speciation events, their distribution tells us if such events happen uniformly over evolutionary time or are clustered at specific periods. Let $g_2, g_3, ..., g_k$ be the distances between each node of a phylogeny with $k-1$ speciation events. A reliable index proposed by Pybus and Harvey (2000) to analyze topological aspects of branch lengths is the γ-statistics, defined as:

$$\gamma = C\left[\frac{2}{N-2} \frac{\sum_{j=2}^{N-1} T(j)}{T(N)} - 1\right], \quad \text{with} \quad \begin{cases} T(j) = \sum_{k=2}^{j} k g_k, \\ C = \sqrt{3(N-2)}. \end{cases} \tag{3.6}$$

Figure 24 – A sample tree exhibiting the necessary and sufficient parameters for the calculus of both topological metrics. The balance (J) requires the number of leaves ℓ_i and ramifications η_i for each node i. Acceleration of the speciation rate (α) requires the distance between each internal node g_i.



Source: The author (2024).

Here, kg_k is used to measure the relative position of nodes within the phylogeny, $T(j)$ is the sum of branch lengths up to node $j+1$, and $T(N)$ is the total distance of all tree branches. If the population speciates at a constant rate b per branch, then $g_k = 1/bk$ and $\gamma = 0$. We follow the work of Costa et al. (2019), and write the time between speciation events as $g_k = \frac{k^{-\alpha}}{b}$, where α accounts for the change in this time or the net acceleration of speciation. The choice of α is twofold: (1) γ-statistics is sensible to the number of leaves in the tree N, while $\alpha(\gamma, N)$ can be uniquely determined given γ and N; (2) both α and γ correctly predicts a constant speciation rate per branch, as in a pure-birth model ($\alpha = 1$, $\gamma = 0$), but only α can predict a constant speciation rate per unit time, as in a Yule model ($\alpha = 0$, $\gamma < 0$). Under this framework, values that differ from random branching processes show an accumulation of speciation events at the end of population history for $\alpha > 1$, and at the beginning for $\alpha < 0$. In this sense, α tells us if phylogeny results from more ancient or recent speciation.

3.3 Barriers to gene flow

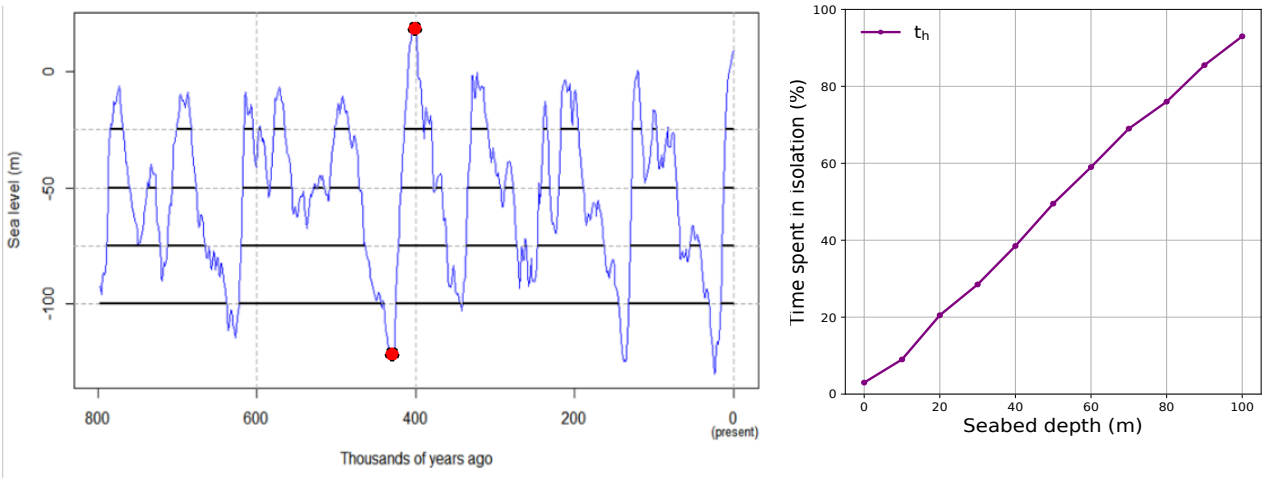
The isolation imposed by barriers may be complete but temporary, alternating between periods of connection and seclusion. There are many empirical examples where the intermittence of the barrier can play an essential role in populations' ecological and evolutionary dynamics. These studies range from oceanic islands (HE et al., 2019), watersheds (BAGGIO et al., 2017), sky islands (PAN et al., 2019) and even host-pathogen interaction (hosts seen as patches of pathogens) (GORDO et al., 2009; BOEGER et al., 2022; D'BASTIANI et al., 2023). In particular, geological landscapes associated with climatic variations may induce or act as temporal barriers (HUGHES; EASTWOOD, 2006; EZARD; PURVIS, 2016; PERES et al., 2020). The frequency and duration of those periods of isolation and gene flow may influence in complex ways the speciation process (NEE; MOOERS; HARVEY, 1992; AGUILéE; LAMBERT; CLAESSEN, 2011), leaving a signature in the phylogenetic data.

We model two migratory scenarios: (1) the patches are permanently connected, and migration is constant through time; (2) the patches have periods of connection and isolation, and migration can only occur during the connection period. We considered the historical sea-level variation over the past 800 thousand years to establish the pattern of connection between the patches (SPRATT; LISIECKI, 2016), where each iteration of the model represents 400 years.

The two patches are isolated when the sea level is above a threshold h (understood as the seabed that halts migration between patches); otherwise, they are connected with a constant migration rate, regardless of how below the sea level is from the threshold. Then, the deeper the threshold (lower values of h), the longer the net time spent in isolation (Figure 25).

In classical biogeography (RIDLEY, 2003), biologists classify speciation in a spectrum from sympatric to allopatric speciation. Allopatry requires geographic isolation, sympatry does not. Furthermore, allopatric processes can involve dispersal and/or vicariance. Speciation by dispersal involves isolation by distance and is commonly used in models with explicit spatial structure. Vicariance involves geographical isolation promoted by barriers. Since our model involves intermittent barriers, we refer to vicariance when speciation is happening due to or along the emergence of a barrier.

Figure 25 – A) Sea-level oscillation over the previous 800 thousand years obtained from Spratt2016. Horizontal dashed lines represent the thresholds h, in which the patches isolate when the sea level is above them (highlighted by the solid black lines). For example, despite seabed depth, the red dots mark the moments when the patches are fully connected and isolated, respectively. B) The relation between the time spent in isolation and seabed depth is approximately linear.

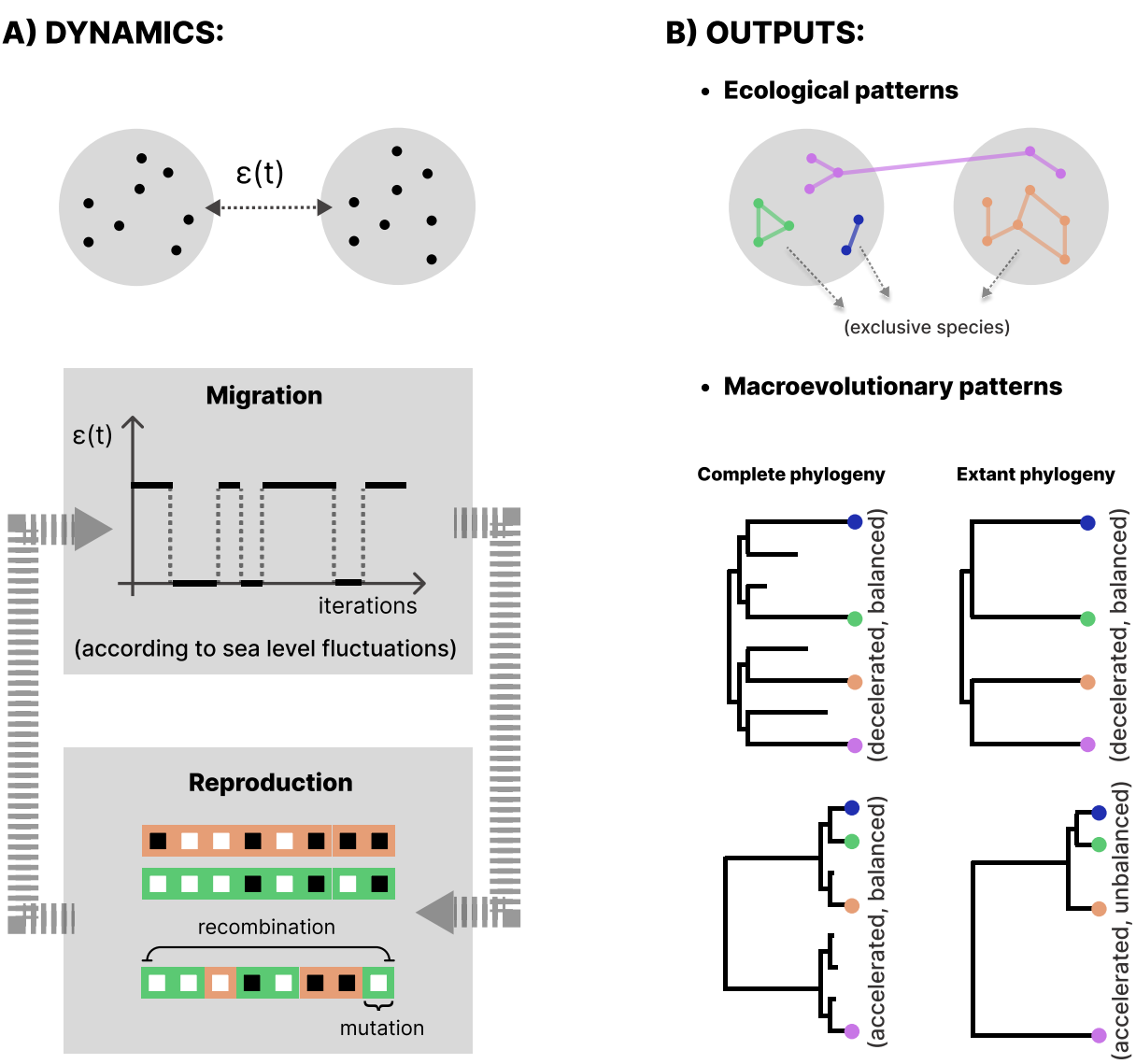


Source: The author (2024).

3.4 Simulation protocol

Initially, one equally allocates M genetically identical individuals over the two patches. Over $T = 2000$ iterations, the population can reproduce and migrate, as detailed below. At each iteration, $M/2$ offspring is born in each patch, replacing their parental generation. In this way, the population size through time does not vary. Reproduction occurs between pairs of individuals, α and β ($\alpha \neq \beta$), within the same patch and whose genetic dissimilarity is equal to or smaller than q_{min}. For each birth event, one randomly chooses a focal parent α. This focal parent has a maximum of $M/2$ trials to find a genetically compatible mate β, also randomly defined, with replacement, between the individuals in the same patch. The first individual β that satisfies $q^{\alpha,\beta} \leq q_{min}$ reproduces with α, and they generate one offspring. The offspring inherits locus by locus the allele of α or β with equal probability. Additionally, each locus can mutate with probability μ. This process is repeated until exactly $M/2$ births per habitat patch are obtained. Regardless of the scenarios, during periods of connection, each individual can migrate to the other patch with probability ε per iteration (Fig. 26A).

Figure 26 – Schematic overview of the study. (A) The model considers an initial population occupying two patches. At each model iteration, the individuals can migrate and reproduce. We evaluate a constant and an intermittent (as illustrated) migration. Reproduction promotes population variability due to recombination and mutation. (B) The model outputs allow us to identify species distribution as well as the complete (including extinctions) and extant (without extinction) phylogenies.



Source: The author (2024).

For each simulation, we evaluate the macroevolutionary patterns (balance and acceleration) in the phylogenies' topology, the number of speciation events, the number of species at a given time N_T, and the distribution of species among patches. To measure the distribution of species among the patches we use the percentage of exclusive species in the patches through the β-diversity index:

$$\beta = \frac{2N_T - N_1 - N_2}{N_T}, \tag{3.7}$$

where N_i is the number of species in patch i. When $\beta = 0$, all species are present in both patches, while $\beta = 1$ means the species are exclusive in each patch.

We also measure the richness asymmetry distribution between patches in a given time, $\Delta N = |N_1 - N_2|/(N_1 + N_2)$. In a recent study (PRINCEPE et al., 2022), the authors observed that the asymmetry could be higher than expected by chance, even under patch equivalence and constant migration.

All statistical analyses were performed in R v.4.1.2 (Posit team, 2022). We used drop.fossil function of the ape package in R (PARADIS; SCHLIEP, 2019) to built the extant trees.

3.4.1 Parameter values and data treatment

For all simulations, we assume a total population size of $M = 400$ individuals, a genome size of $G = 2000$ loci, a mutation probability $\mu = 0.001$ per locus, a total of $T = 2000$ generations and the maximal genetic dissimilarity of $q_{min} = 0.9$ - therefore a difference of $H^{min} = 100$ genes. For this set of parameters, pure sympatric speciation is unlikely to occur when the patches are entirely isolated ($\varepsilon = 0$) (as was shown in Fig 21).

We vary migration rate ε from 0 to 0.2 and sea level threshold h from -100 to 0 (the sea level data vary between $-130.0m$ to $+19.2m$ (SPRATT; LISIECKI, 2016)). We also considered the scenario without isolation, equivalent to $h > 19.2m$. To facilitate the comparison between different sea level thresholds, the results are presented in terms of the fraction of time in isolation, t_h, and the mean migration rate, $\bar{\varepsilon} = \varepsilon(1 - t_h)$, ranging from 0 to 0.08. The relation between t_h and h, for $-100 \leq h \leq 0$, is approximately linear, $t_h \approx |h|/100$ (see Fig. 25B).

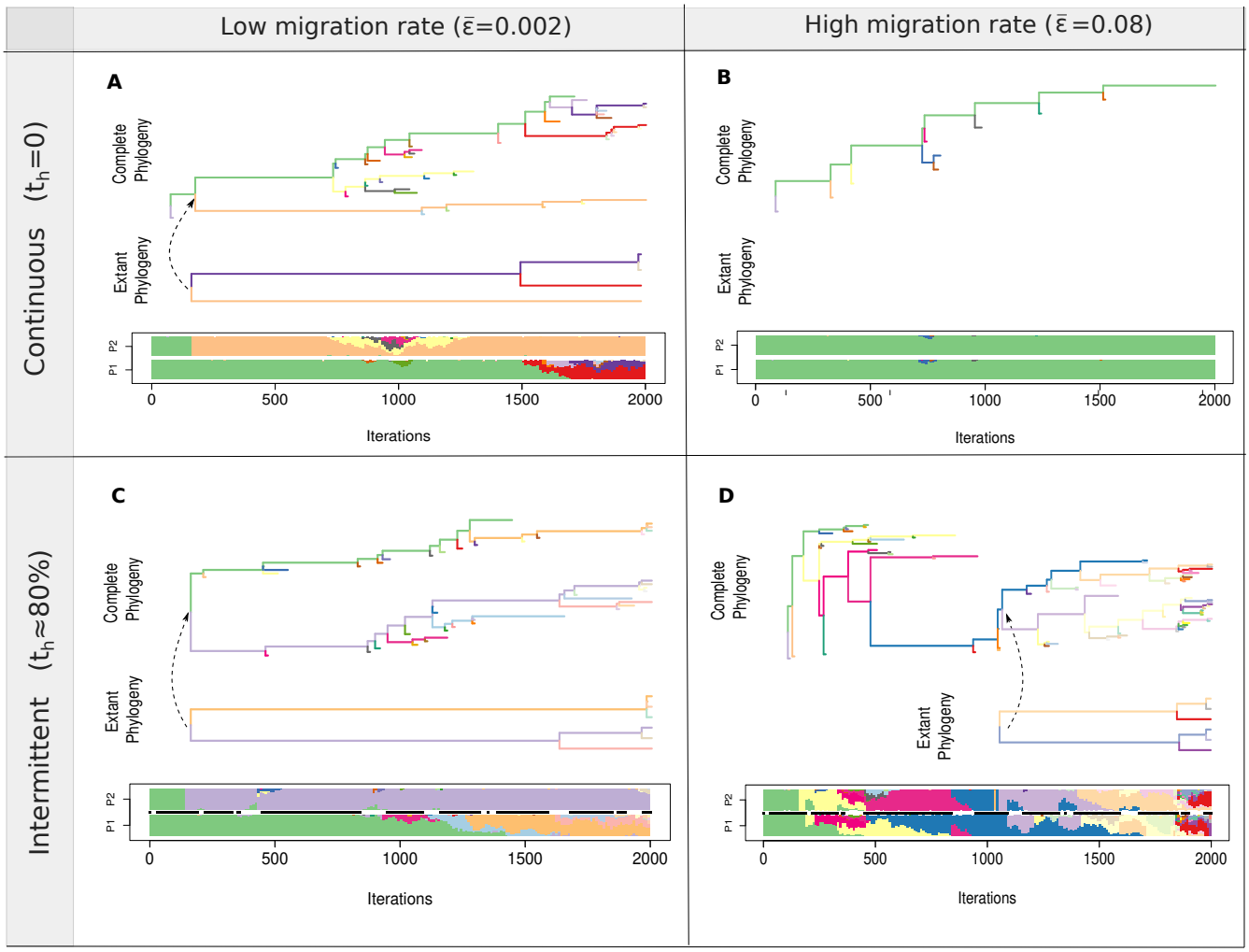
For each set of parameters, we realized 50 independent replications. As the calculation of J and phylogeny age require at least one speciation event, while α-statistics two events, we report only the cases in which at least 20% of the replications attended this criterion.

To better understand how the above metrics relate to each other, we measure their correlation: for each isolation time (t_h) we calculated the Pearson correlation for the data set of the independent runs, comprising the aforementioned mean migration rate interval. Results are reported at the end of simulations ($T = 2000$).

3.5 Results

Instances of the evolutionary trajectory under our model produce different evolutionary dynamics (Fig. 27). Although they are just examples without a proper statistical analysis, some visible elements help to understand the results.

Figure 27 – Examples of the model evolution under constant (top) and intermittent (bottom) migration; as well as low (left) and high (right) migration. For each example, there are three plots: upper and middle panels correspond to the complete and extant phylogenies, respectively. The plot at the bottom illustrates species distribution abundance in each patch over time. In each speciation event, the most abundant species maintains its previous color. The horizontal black line in the lower plot indicates when patches are isolated. In all plots, the most abundant species keep the ancestral color. The values of migration probability and the percentage time in isolation ($t_h \approx |h|/100$) are indicated in the figure headings.



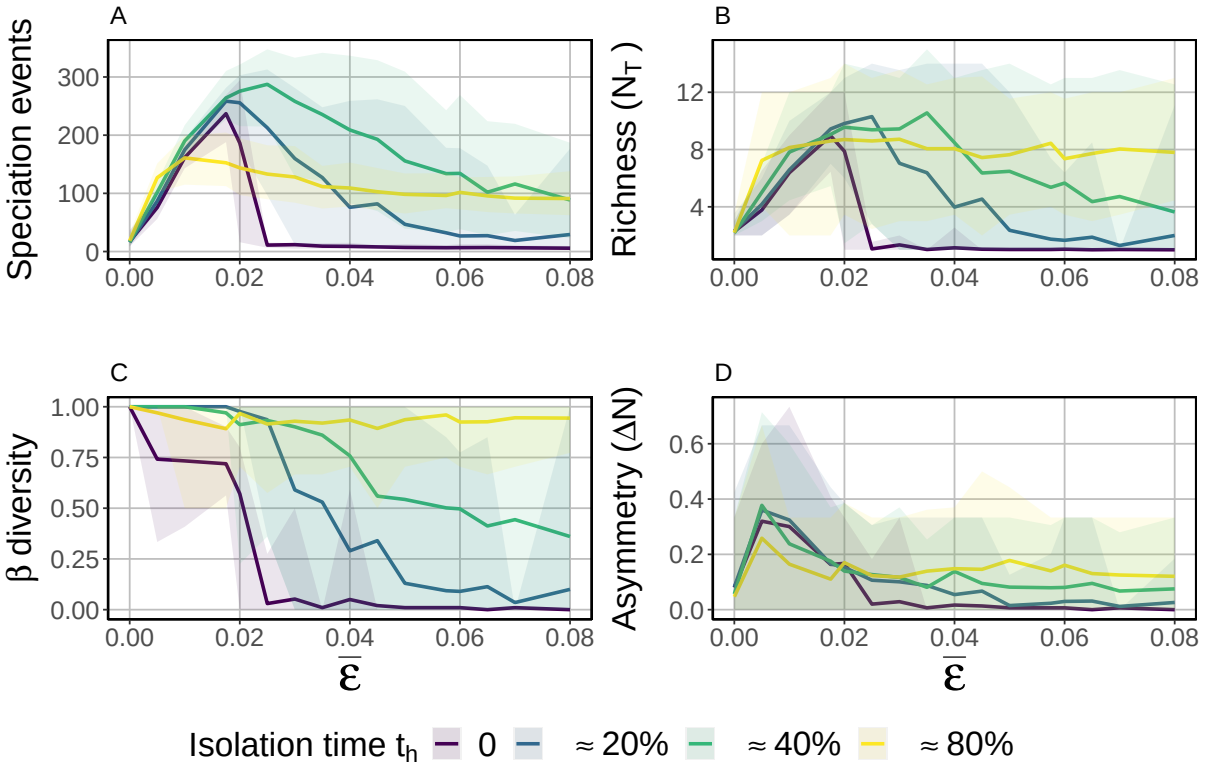
Source: The author (2024).

Two modes of speciation can be observed: (i) speciation induced by migration (when speciation occurs during connection phase); (ii) vicariance, the speciation caused by the barrier's emergence. Speciation induced by migration can occur if migration is low (Fig. 27A), but

it is prevented under high migration (Fig. 27B). This speciation induced by migration was previously observed by (PRINCEPE et al., 2022), in which a few individuals migrate to another island and found a new population or impose several genetic novelties that result in sympatric speciation. For example, in Fig. 27A, the second speciation event (yellow) owes to a founder population (the yellow species that occur in $P2$ emerged from the green species from $P1$) and, subsequently, a cascade of speciation takes place in sympatry. On the other hand, speciation promoted by vicariance is the predominant process under high migration: during the migrating period, the patches share the same species, and during the isolation, speciation events occur (Fig. 27D, and also discussed by (PRINCEPE et al., 2023)).

Statistical analysis over an ensemble of independent evolutionary trajectories unveils different scenarios in terms of the mean migration rate ($\bar{\varepsilon}$) and time in isolation (t_h). We report the ecological and macroevolutionary patterns at the end of simulations (Fig. 28 and 29).

Figure 28 – Number of speciation events, species richness, β diversity, and species asymmetry distribution as a function of the mean migration rate ($\bar{\varepsilon}$) for different intermittent regimes (colors). The calculus of all metrics is made only by the end of the simulation ($T = 2000$). The solid curves represent the average over 50 replications, and the shadow area the confidence interval of 90%.



Source: The author (2024).

Despite the high number of species generated throughout history (Fig. 28A), the number of species coexisting at a particular generation (for example, at the end of simulations, as in

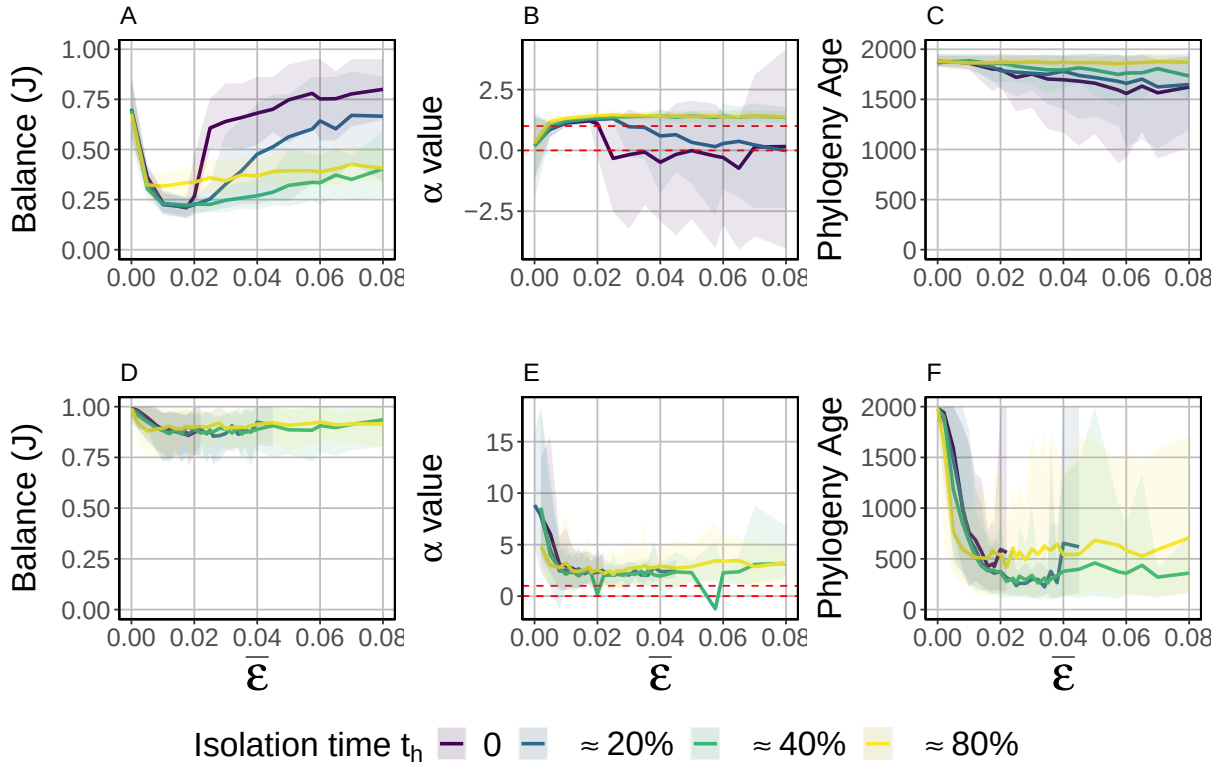
Fig. 28B) is of the order of units. Thus, over a long time, the number of extinction events is practically the same as speciation events (Fig. 28A).

The species distribution among patches at the end of simulations shows a decrease in β-diversity with mean migration rate for all but the most isolated case (Fig. 28C). This decrease is expected since migration between patches favors the homogenization of species distribution. However, as β-diversity is measured at the end of the simulation and the last iterations are in isolation ($\forall t_h > 0$, see Fig. 25), for $t_h \approx 80\%$ these last iterations are time enough to promote speciation, keeping β-diversity in high values. The β-diversity for $t_h \approx 20\%$ and $t_h \approx 40\%$ decreases more slowly than $t_h = 0$ because the intermittence of the migration promotes speciation (see Fig. 28A) and also because the last time in isolation can promote speciation for not all but part of the species. The asymmetry of the species distribution over the patches varies with the migration intensity in a very similar fashion, despite isolation time (Fig. 28D). The higher asymmetry occurs under low migration probability ($\bar{\varepsilon} \approx 0.005$). Although the difference is subtle, it reveals that as the migration increases, the species distribution obtained from the simulations tends to be more symmetrical.

The phylogenetic metrics reveal different macroevolutionary patterns (Fig.29). In panel Fig. 29A, the phylogeny balance J displays a non-linear dependence on the migration $\bar{\varepsilon}$, declining for low values of migration rate regardless of the isolation regime, up to reaching a minimum at $\bar{\varepsilon} \approx 0.02$, which corresponds to the points at which species richness and speciation events achieve their maximum. Then, the phylogeny becomes more balanced with increasing migration. In this regime, the signature of the intermittent barriers on the phylogeny balance becomes prominent, so that the phylogeny balance tends to decrease with the time spent in isolation.

The balance of the complete phylogeny indicates the speciation mode. When migration is low, the species distribution that emerges from the dynamics is not symmetrical, thus one patch has more abundant species. The chance that two (or more) immigrated individuals reproduce is higher if their patch of origin has one abundant species, as it increases the likelihood of the two individuals being compatible and establishing a new species (PRINCEPE et al., 2022). This dynamics of speciation reduces the balance of the phylogeny (Fig. 29A) since new species are generally emerging from one abundant species. Under sufficiently high migration, species disperse over the patches and occupy them equally (symmetry). The emergence of a barrier

Figure 29 – Macroevolutionary patterns (phylogeny balance, speciation acceleration, and age) for complete phylogenies versus the mean migration rate($\bar{\varepsilon}$) for different intermittent regimes (colors). The solid curves represent the average value of the metrics and shadow areas indicate the confidence interval of 90%. Following the α definition, the dashed red lines indicate the interval in which the results do not differ from random branching models.



Source: The author (2024).

in this scenario promotes speciation by vicariance, and symmetrical speciation events tend to occur, increasing the balance of the phylogeny.

Looking at the acceleration of speciation metric, α, one also observes that at low migration rates, the results are unaffected by the duration of isolation events. Like for J, at intermediate and large migration rates, the effect of the isolation time, t_h, on the metric comes about, and now we see that for more isolated populations ($t_h \approx 40\%$ and $t_h \approx 80\%$), α saturates at a rate of $\alpha \approx 1.25$, meaning that speciation events are more numerous in recent history. On the other hand, under reduced isolation ($t_h = 0$ and $t_h \approx 20\%$) and $\bar{\varepsilon} > 0.02$, speciation events are more uniformly distributed over evolutionary time ($0 < \alpha < 1$).

In complete phylogenies, the first speciation event equals the phylogeny age (Fig. 29C). At low $\bar{\varepsilon}$, it happens almost at the same iteration time despite isolation, an indication of the mean waiting time for allopatric speciation. Increasing migration rate tends to increase the time needed to speciate. The exception is for the most isolated case ($t_h \approx 80\%$) since a long

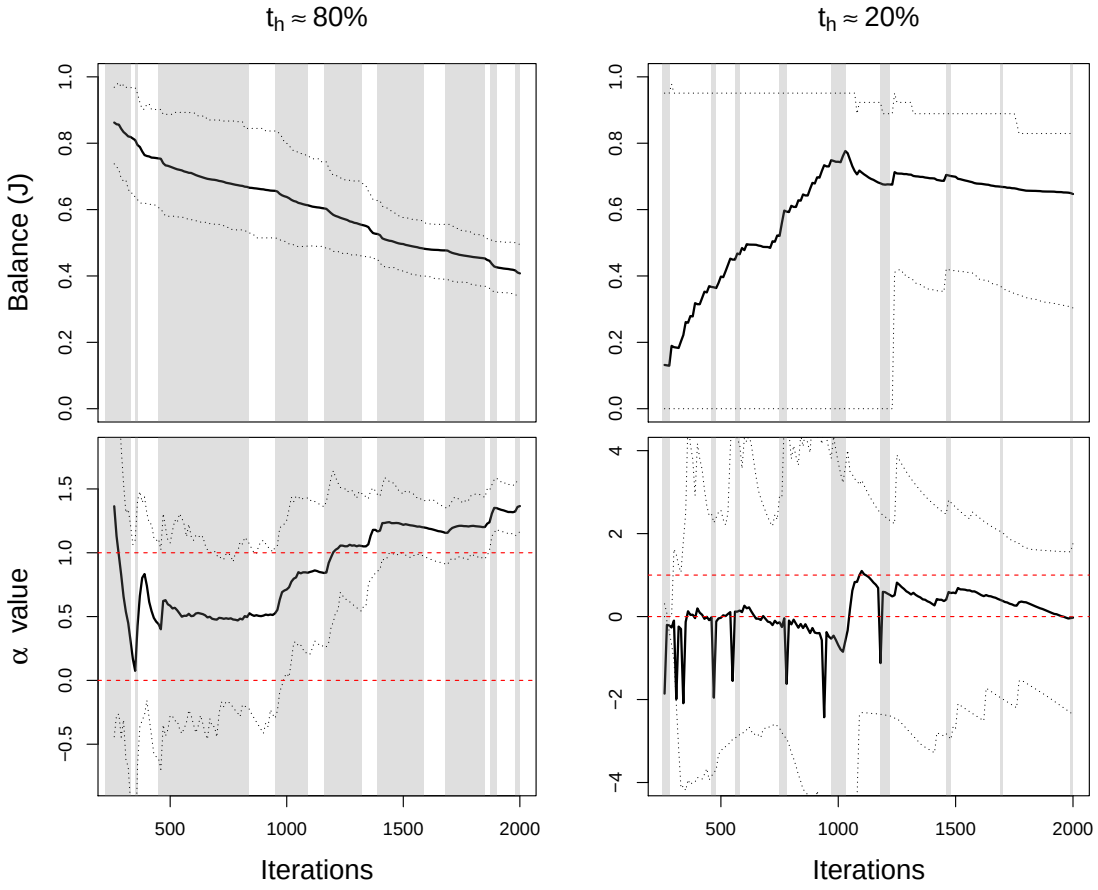
period of isolation is imposed at the beginning of the dynamics (see Fig. 25).

The values of balance, speciation acceleration, and phylogeny age of extant phylogeny have different tendencies from complete phylogeny (Fig. 29D-F). Regardless of migration intensity and isolation period, the extant phylogenies are balanced, accelerated and the information of the first speciation event tends to be lost as the migration increases; that is, the age decreases. Thus, the scenarios with different isolation times were not distinguishable on extant phylogenies. However, is important to highlight that the extant phylogenies presented a very low number of species, which may reduce the statistical significance of these results.

Recent studies have shown that the emergence of a barrier initially favors speciation, followed by a reduction of the speciation rate, probably due to the decline of opportunity for new speciation (ALENCAR; QUENTAL, 2021). On the other hand, cycles of gene flow during speciation promoted by intermittent barriers were shown to generate species at an exponential rate (HE et al., 2019).

To see the effect of barrier emergence, we report the balance and speciation acceleration over time for complete phylogenies, for the same migration rate ($\bar{\varepsilon} = 0.08$), and when the population spends more time in isolation ($t_h \approx 80\%$) and in connection ($t_h \approx 20\%$) (Fig. 30).

Figure 30 – Evolution of phylogeny balance and acceleration of speciation for combinations of isolation $t_h \approx \{80, 20\}$ and migration rate $\bar{\varepsilon} = 0.08$. Solid lines represent the average of 50 replications, while dotted ones delimit the confidence interval of 90%. Following the α definition, the dashed red lines indicate the interval in which the results do not differ from random branching models. Background gray bands indicate when patches are isolated for each t_h.



Source: The author (2024).

Comparing the balance between the two cases there is a distinct tendency: for high isolation, barrier emergence induces a pulse of speciation events – speciation by vicariance – whose rate is captured by its acceleration; while for low isolation, speciation is induced during periods of connectance – speciation induced by migration – and its rate sharply decreased with barrier emergence. We see the initial time period of isolation/connectance seems determinant to establish the genetic variation within and among patches, thus we see a higher initial tree balance for $t_h \approx 80\%$ in contrast with $t_h \approx 20\%$.

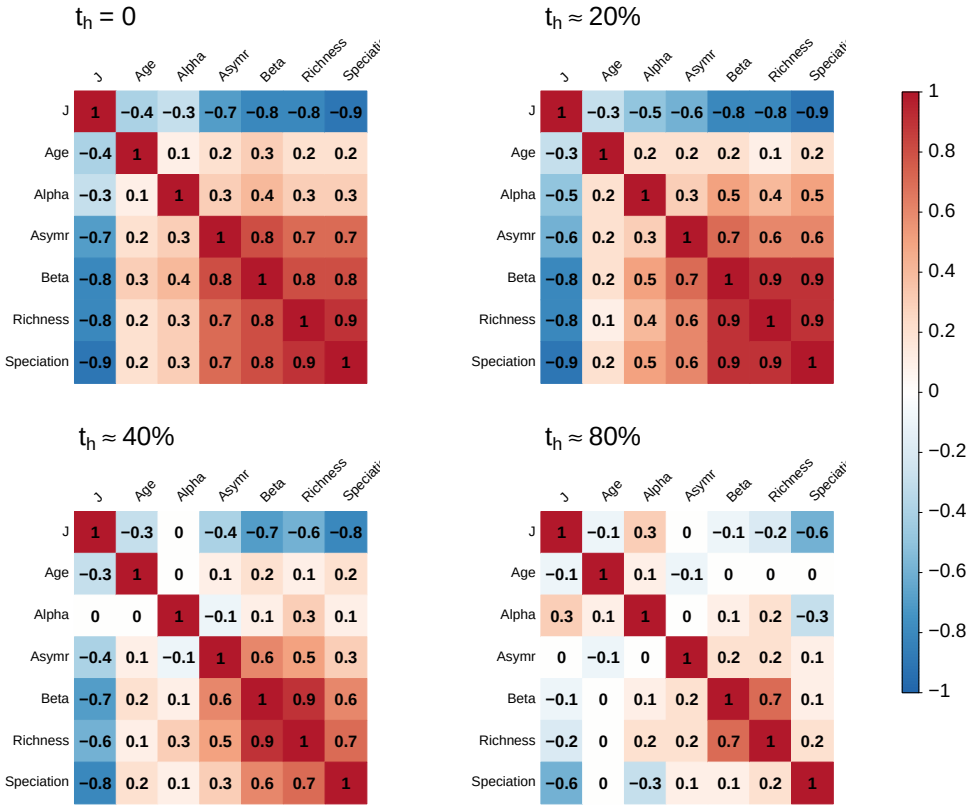
Acceleration of speciation tends to a positive stationary value while the phylogeny balance tends to decrease regardless of the scenarios. In fact, for a long enough time, the phylogeny balance should tend to zero since each extinct branch contributes with the decrease of J. Both α and J metrics depend on all historical events of extinctions and speciations, and they tend

to converge as the recent events become only an insignificant fraction of the whole history.

Our results showed that at each emergence of a barrier, the speciation rate first quickly increases and then slows down. However, when the time in isolation is insufficient to decrease acceleration to values before the vicariance, each new isolation event tends to increase the acceleration.

Next, we investigate how the results from Figure 28 and 29 correlate with each other. Interestingly, we observe, in general, that the degree of correlation drops off with the period of isolation t_h (Fig. 31). For example, our results unveil that all metrics are poorly correlated at the more extreme value of t_h, corresponding to $t_h \approx 80\%$. The correlations between the metrics related to the phylogenies (J, Age, and α) are relatively weak regardless of the isolation period. Nevertheless, in a low isolation period, the phylogeny balance J displays a substantial correlation with the remaining metrics: asymmetry, β diversity, species richness, and speciation number. However, for scenarios of a high isolation period, J and α can provide important complementary information along with those related to species composition.

Figure 31 – Correlation tables between the analyzed metrics for the four scenarios of isolation (t_h indicated in the figure). The correlation was obtained comprising the whole interval of migration values $\bar{\varepsilon} \in (0.002, 0.08)$, with values reported at the end of simulations ($T = 2000$) for all complete phylogenies.



Source: The author (2024).

3.6 Conclusion

In this insular model of speciation by genetic differentiation, we investigate macroevolutionary patterns through topological phylogenetic metrics.

The imbalance of empirical phylogenies often has been associated with heterogeneous ecological processes among lineages (BLUM; FRANçOIS, 2006; CARON; PIE, 2020). However, theoretical models have already shown that unbalanced phylogenies can emerge when speciation involves isolation by distance (PIGOT et al., 2010; COSTA et al., 2019), while balanced phylogenies from sympatric (COSTA et al., 2019) and vicariant (PIGOT et al., 2010; MARQUITTI; FERNANDES; AGUIAR, 2020) speciation. In line with these theoretical studies, our statistical analysis of the complete phylogenies reinforces the need to consider nonecological processes when interpreting macroevolutionary patterns.

The role of the geographical barriers and the duration of the isolation period is also evident when we look at the acceleration of the speciation process. Accelerated speciation occurs ($\alpha > 1$) for high isolation times (Fig. 29B). These results are at odds with those expected in null models such as (YULE, 1925) and birth-death models (KENDALL, 1948), where the speciation rate can be constant or equal among all lineages. Although these null models do not deal with the microevolutionary dynamics, at first glance, an agreement between the two approaches was expected as, in both cases, individuals are ecologically equivalent (ALENCAR; QUENTAL, 2021), but see (MCPEEK, 2008).

Previous theoretical studies have shown that in scenarios of smooth ecological pressure (GASCUEL et al., 2015) (near the neutral regime), the longer the time spent in isolation, the more unbalanced and accelerated the phylogeny; in contrast, shorter times lead to balanced and decelerated trees. Although our model differs in many respects, we observe this behavior for a sufficiently high migration.

The simplified assumptions we made to study the mechanisms of diversification with intermittent gene flow have some limitations. Regarding the parameters, population size ($M = 400$), mutation rate ($\mu = 0.001$), genomic size ($B = 2000$), and time for one iteration (400 years, although it implies in 2.5×10^{-6} mutation per locus per year) are far from realistic. This set of parameter combinations reproduces the expected empirical scenario of vicariance: two species emerge when the two islands are isolated. However, the number of coexisting

species (richness) was rarely superior to 10 because extinctions are likely to occur under low populations. Consequently, extant phylogenies generated by the model lack sufficient richness for a meaningful statistical analysis. Extant phylogenies tend to be balanced and accelerated regardless of the duration of intermittent barriers; being indistinguishable regardless of isolation time. Although statistical inference of these extant trees is fragile, we hypothesize that these two general features will prevail when compared with the complete ones, given that the removal of extinct species implies the loss of many ancient speciation events, resulting in long branches connecting recent speciation events, thus giving rise to “tippy" trees (see for example Fig. 27C). Furthermore, such losses also affect the increase in phylogenetic balance, since all extinct species represent branches without speciation (LEMANT et al., 2022). Modifications on the model structure, as increasing the number of islands, could be done to promote higher richness. Nevertheless, our model allows a better understanding of the speciation mechanism under intermittent regimes and its consequence on the complete phylogeny. It showed some possible effects of migration intensity and isolation time on the speciation process.

In our model, lineage speciation and extinction rates are not imposed as a parameter but emerge naturally from the population-level dynamics and are influenced by two major factors: species abundance and abiotic paleoenvironmental conditions. Here we stress a scenario where paleoenvironmental conditions are directly related to population isolation mechanisms and, thus, to the diversification process (CARNAVAL et al., 2009). As such, we highlight the importance of integrating phylogeny data with historical paleoenvironmental data in order to explore its role in ecological biodiversity better.

- The main results presented in this section have been published in the article: *Freitas, Osmar, Paulo RA Campos, and Sabrina BL Araujo. "Patch biogeography under intermittent barriers: macroevolutionary consequences of microevolutionary processes." Journal of Evolutionary Biology (2024).* Ref. (FREITAS; CAMPOS; ARAUJO, 2024)

4 SPECIATION IN A METAPOPULATION

Biogeography does more
than ask where organisms live;
it asks why they are found there.
(MacArthur & Wilson)

Speciation events can occur either by geographic isolation or selective pressure imposed on organisms in a population. Geographic isolation is the most accepted and simple mechanism to explain speciation events, however, it is not enough to describe the great diversity of species on our planet.

Selective pressure promotes speciation when inducing adaptation to different environmental conditions, leading to population divergence. In this sense, speciation is expected to be triggered by events of environmental change as well, as a response to climate adaptation (QVARNSTRÖM et al., 2016), but also as a result of human activities and urbanization (THOMPSON; RIESEBERG; SCHLUTER, 2018; HENDRY; NOSIL; RIESEBERG, 2007). On the other hand, species unable to cope with a scenario of fast and intense environmental alterations may go extinct.

To analyze both selection and geographic effects on speciation, it is convenient to use structured population models such as a metapopulation (HANSKI, 1998; GAVRILETS, 2004). In this scenario, populations are distributed among patches connected by migration and each patch may harbor different environmental conditions that maximize survival. Straightforwardly, one can now quantify and separate the contribution of each mechanism and asses the diversity within and among the subpopulations.

In this chapter, we investigate the effect of geographic isolation and selective pressure on the diversity and persistence of a metapopulation in a time-varying environment. For this, we use an individual-based model with genetic ground, where the population evolves in response to viability selection, density-dependent reproduction, and migration. We assess the diversity pattern and the degree of divergence at equilibrium and how it differs when evolving under different environmental perturbations. Our results show the conditions in which environmental pressure and its temporal variation can favor species diversity maintenance or extinction, and allow us to discuss the impacts of the intense and rapid climatic oscillations occurring on our planet.

4.1 Model

As in our previous approaches (see Chapter 2), our model is phenotypic with a genetic basis, in which individual traits represent a point in phenotypic or trait space. Mutations correspond to random displacements in trait space and are assumed to act additively. Selection acts on the probability of individual survival - their fitness or viability - and is proportional to the distance from an individual's traits to the corresponding optimum traits on each patch.

Differently, here, environmental changes are modelled in a perturbative way through fluctuations in the optimum phenotype of every patch. These changes are expected to increase the mismatch of individuals' phenotype with the optimum in their respective patch, thus causing a drop in individual viability. The possibility of exchanging patches through migration regardless of an optimum match acts as a double edge: well-adapted individuals may find it hard to survive in different environments; ill-adapted ones may have the luck to find more suitable places. In harsh conditions, in which either the selective pressure is too intense, or the rate of environmental changes is too high, the risk of population extinction is enhanced regardless of migration (CHEVIN; LANDE; MACE, 2010). Together, these processes may influence the level of genetic variation and diversity of the metapopulation.

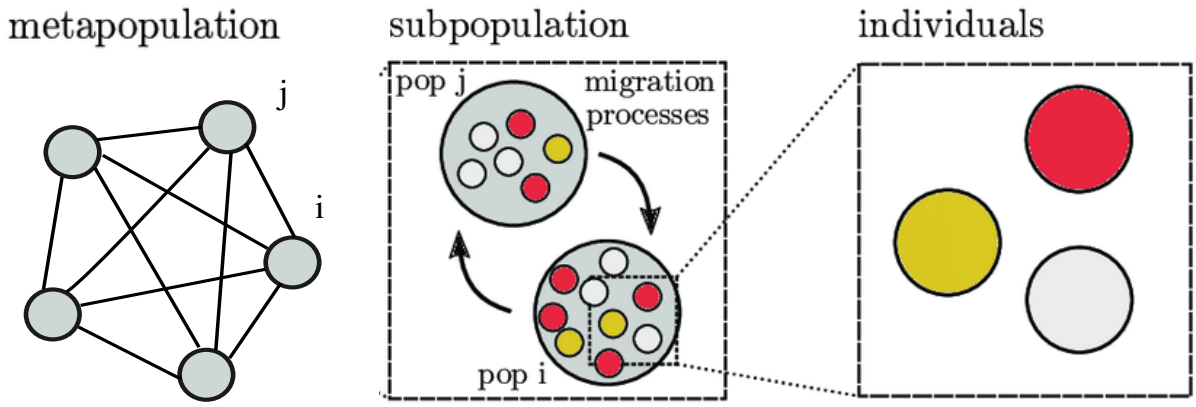
4.1.1 Metapopulation

A metapopulation is a spatially structured model in which local populations inhabit patches connected by dispersal. The patches can represent different spatial locations or resources, as hosts in the host-parasite interactions (BRAGA et al., 2018). One can picture a network where population reproduction and adaptation are restricted within each node (see Figure 32).

Earlier mathematical models on metapopulation dynamics have focused on the alternate states of occupancy and vacancy of its patches, and have shown how local extinction and recolonization influence demographic persistence, the coexistence of interacting species, and genetic variation (BÜRGER, 2006; LEVINS, 1969; HANSKI, 1998).

In our mechanistic modeling, a metapopulation is distributed over a constant number of M patches under sustained environmental change. As our focus is on environmental changes and not on the different landscape structures, we adopt a homogeneous configuration for the fragmented metapopulation. The patches have the same carry capacity K and form a fully

Figure 32 – Representation of a metapopulation model, consisting of connected patches and the individuals that inhabit them. Colors allude to the differences in population composition. In this particular case of a homogeneous metapopulation, all the patches are mutually connected.



Source: Modified from (HACKL; DUBERNET, 2019).

connected network. Individuals inside the patch can migrate at a probability m per generation and are equally likely to move to one of the $M-1$ remaining patches.

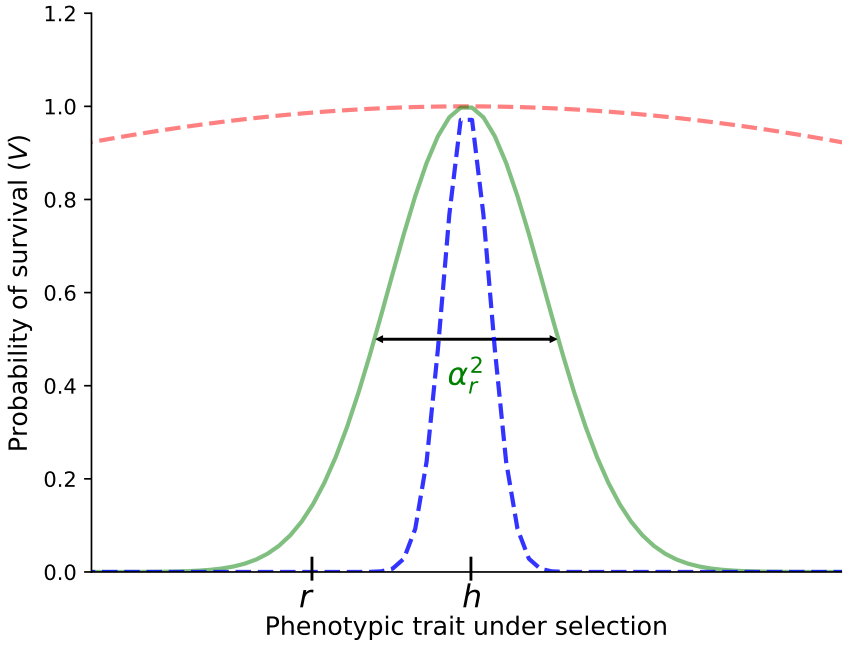
4.1.2 Fitness landscape

At the phenotypic level, a trait under selection determines the probability of individual survival in each patch. For L traits, we represent an individual's phenotype as a L-dimensional vector, $\vec{r} = r_1, r_2, ..., r_L$, where r_i is the trait value of the i-th trait, and $\vec{r}$ is the phenotype vector associated with a given genotype. We use Fisher's Geometric model (FISHER, 1930; ORR, 2005b; SOUSA et al., 2016; BRAGA et al., 2018), to define the fitness of an individual as a function of the Euclidean distance between $\vec{r}$ and $\vec{h}$ - the traits that maximize population survival in a given patch. Under Gaussian stabilizing selection, the viability of an organism j with phenotype $\vec{r_j} = (r_{j,1}, r_{j,2}, \ldots, r_{j,L})$ in a given patch k is calculated as

$$V_j^k = \exp\left(-\frac{\sum_\ell (r_{j,\ell} - h_{k,\ell})^2}{2\alpha_r^2}\right), \tag{4.1}$$

where $h_{k,\ell}$ represents the optimum value of trait ℓ at patch k, which is subject to temporal change. The parameter α_r is the width of the viability function, and is associated with the strength of selection; the smaller the α_r the stronger the selection is. In the limit $\alpha_r \to 0$, only those individuals whose phenotype matches the patch optimum phenotype survive, whereas in the limit $\alpha_r \to \infty$ one recovers the neutral selection regime (Figure 33).

Figure 33 – Example of the one-dimensional Gaussian viability function for different widths α_r (colours). On the trait-axis, we have the individual phenotype at point r and the optimum phenotype of the patch it inhabits h. Selective pressure goes with $1/\alpha_r$, thus the individual has zero probability of survival for the blue curve, low probability for the green, and almost one for the red. In the limit $\alpha_r \to \infty$ one has $V \to 1$, thus individuals are always viable as there are effectively no selection mechanisms.



Source: The author (2024).

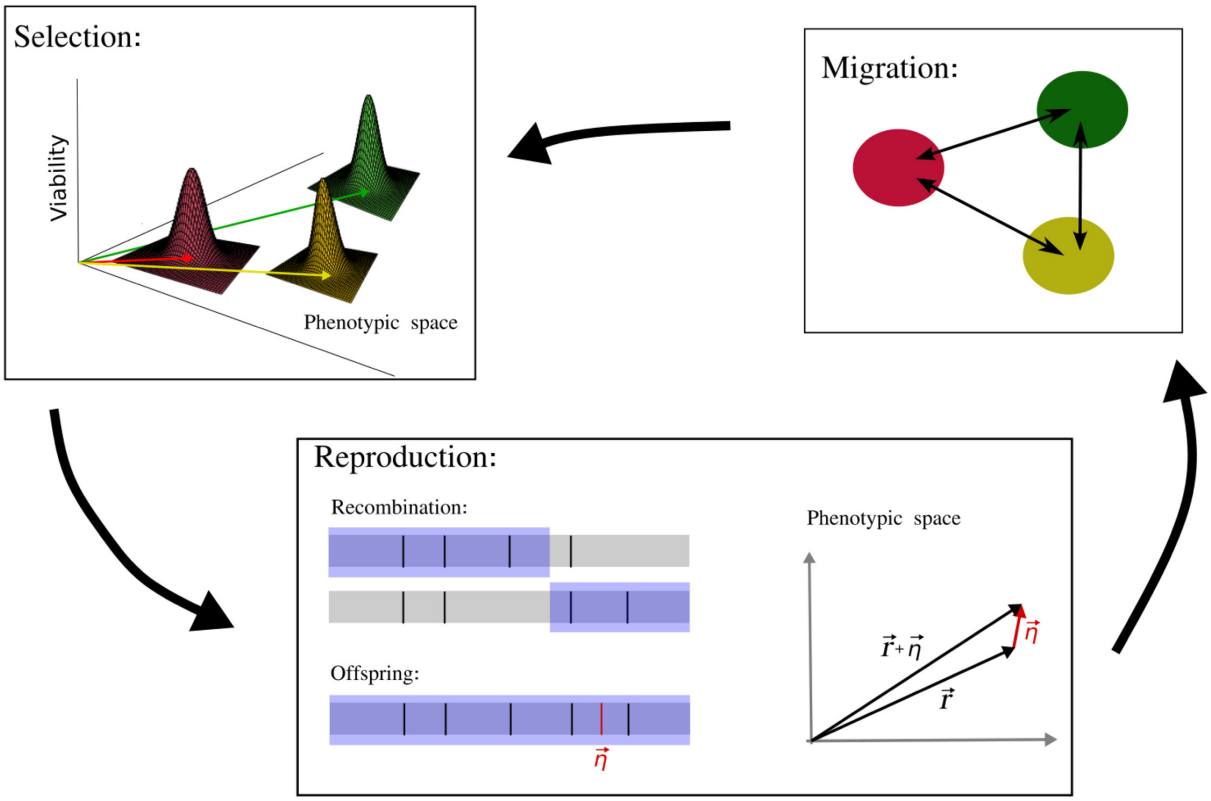
Importantly, the width of viability selection α_r is responsible for establishing the tolerance limits of the individuals. The lower the α_r, the stronger the selection and the narrower the phenotypic range the individuals are viable. Note that selection is independent and equally strong in all L traits.

4.1.3 Population growth and speciation

Individuals are haploid and characterized by an infinitely large genome, such that the infinite sites assumption applies (see Chapter 1). In this case, every mutation, which occurs at a rate μ per genome per generation, is unique. Therefore, there is no back mutation. Once a mutation arises, the phenotype of the mutant becomes $\vec{r} + \vec{\eta}$ (Fig. 34 presents an illustration of the process). So a mutation can affect the L traits, resulting in a displacement $\vec{\eta}$ in the phenotypic space. Because we make use of the additive assumption, mutations are combined additively, such that the phenotype arising from two mutations is simply $\vec{r} + \vec{\eta_1} + \vec{\eta_2}$ (MARTIN; ELENA;

LENORMAND, 2007; HWANG; PARK; KRUG, 2017). Once a mutation occurs, each component of the displacement vector $\vec{\eta}$ is obtained from a Gaussian distribution of null mean and variance σ^2_{mut}. Henceforth, σ_{mut} is set at $\sigma_{mut} = 0.01$. [1]

Figure 34 – Scheme of the model dynamics. Individuals are distributed among patches, illustrated by green, red, and yellow colors. At each generation, individuals go through Selection, Reproduction, and Migration. Selection: patches impose a survival probability (viability) on their inhabitants. The position of the peaks, on the phenotypic space $\vec{h}$ (illustrated by each colored vector) denotes the optimum phenotype, which changes from patch to patch. Reproduction: occurs sexually among individuals inhabiting the same patch and are not genetically too dissimilar, i.e., their genetic distance must be smaller than a critical value g. The offspring of a couple is a product of parental genomic recombination plus a possible mutation. Migration: occurs between patches with a probability m per individual.



Source: The author (2024).

Reproduction occurs locally inside each patch. Each individual randomly selects a potential partner from the same patch. Reproduction ensues once the Hamming distance, the genetic distance between two genotypes, is smaller than g. In the infinite site approach, the Hamming distance is simply given by

$$\pi_{ij} = n_i + n_j - 2n_{i,j}, \tag{4.2}$$

where n_i is the number of mutations of genotype i, and $n_{i,j}$ is the number of mutations shared by the two genotypes. Then, if the condition $\pi_{ij} < g$ is met, there is exchange of genetic

[1] A high rate, but still in the range of realistic values as found for Drosophila melanogaster and microbes (Drake 1999; Gao, Pan, Hu, Ma, Wu, Shao, Barton, Woodruff, Zhang, and Fu 2011.

information, in which the one-point crossover scheme is adopted for the recombination (see Fig. 34). Accordingly, the parental genotypes are cut once in a randomly chosen position of the genome and recombined to form the offspring (KLUG; PARK; KRUG, 2019). Non-overlapping generations and a density-dependent population regulation are assumed (HASSELL, 1975), such that the number of copies of a given offspring is obtained from a Poisson distribution of mean value (HENRIQUES-SILVA et al., 2015)

$$\rho = \frac{\lambda}{(1+\gamma N_t)^{\beta}}, \tag{4.3}$$

with

$$\gamma = \frac{\lambda^{1/\beta} - 1}{K}. \tag{4.4}$$

In Eq. (4.3), λ refers to per capita growth rate, N_t is number of successful mating in the patch. Note that γ is chosen such that the minimum value of ρ is one if $N_t = K$. In fact, K makes the role of a carrying capacity and is assumed to be the same for all patches. The parameter β defines the type of competition, and is set at $\beta = 1$, corresponding to a contest competition (OLIVEIRA et al., 2020).[2]

As in our previous chapter, we define that a speciation event occurs when gene flow between two subpopulations is not possible. Therefore, a species is a subpopulation where all members have their Hamming distance, π, larger or equal to g, relative to any other group,i.e. no member of distinct subpopulations can mate and recombine. The latter condition means that there are no connected paths in genotype space connecting the two genomes, i.e., there are no intermediate species connecting them.

4.1.4 Enviromental perturbations

In the regime of environmental variations, the fitness landscape is not static, such that the phenotype favored by each patch changes periodically. Every τ generations the phenotype favored by patch k, $\vec{h_k} = (h_{k,1}, h_{k,2}, \ldots, h_{k,L})$ undergoes a shift in the phenotype space. Each component $h_{k,j}$ changes by an amount ε_j obtained from a Gaussian distribution of null mean and variance σ_h^2.

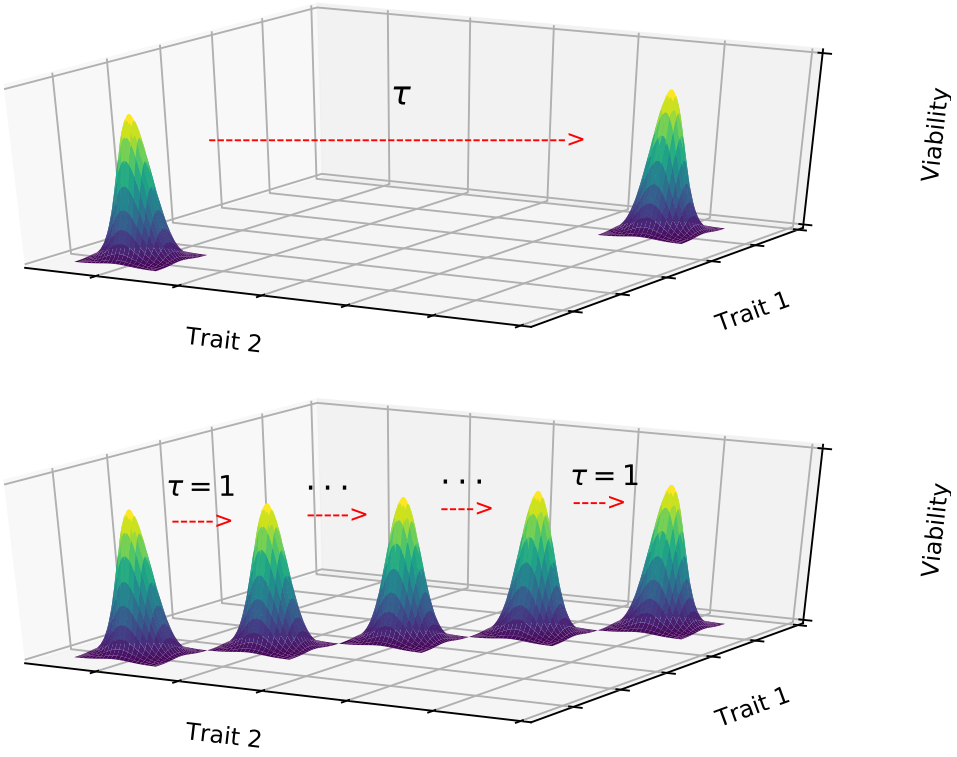
Given our focus on the frequency and magnitude of environmental fluctuations, it seems natural to inquire about the role of gradual vs. discontinuous and abrupt changes in the

[2] Note that, in Chapter 2, the per capita growth rate was given by the fitness itself W. But here, fitness V is taken solely as the probability of survival; and the growth rate emerges from the demographic regulation.

dynamics. It seems plausible that a gradual phenotype displacement must be drawn from a distribution with $\sigma_{gradual} < \sigma_h$. But how much smaller must $\sigma_{gradual}$ be, and how can we relate it to the different time-scales for a reasonable statistical comparison?

We borrow an idea from the Central Limit Theorem (REICHL, 2016). If X is a random variable from a probability density $P(X)$, the theorem states that the average of a large number of statistically independent events N, will be a Gaussian centered at $\bar{x}$, and with standard deviation $\sigma_x/\sqrt{N}$ (see Appendix C for a formal derivation of the theorem). In our case, the number of events of gradual change equals τ generations. As the displacement of the optimum phenotype is additive over time, the accumulated distribution of gradual changes matches the distribution under discontinuous change whenever $\sigma^2_{gradual} = \sigma^2_h/\tau$ in our model, which assumes a non-constant displacement per event. For the sake of clarity, Fig. 5 shows a sketch of the dynamics under the two scenarios under the elucidative example of constant displacement. After time τ, the net displacement of the two processes will be the same.

Figure 35 – A mere illustration of moving optimum for a given patch. In this example, the displacement of the optimum phenotype is constant during each event, whereas in our simulations the displacement in each trait is taken from a Gaussian distribution. When we compare gradual (corresponding to $\tau = 1$) vs. abrupt environmental changes, the standard deviation for the gradual process is chosen such that over a time τ the distribution of the net displacement for the two processes are exactly the same.



Source: The author (2024).

4.2 Simulation and Measurements

The model is comprised of a recurrent cycle (see also Fig. 34): viability selection, recombination followed by reproduction, and migration. Concerning the details of each stage of the life cycle:

1. Selection: each individual survives this step with a probability provided by Eq. 4.1. Those individuals who do not survive will be removed. Therefore, those individuals whose phenotype is more distant from the optimum phenotype will be removed with a higher probability.

2. Reproduction: for each surviving individual, another individual from the same patch is randomly selected (with reposition). If the Hamming distance between the two genotypes is smaller than g, the matting is successful; the genomes of these two individuals are recombined and a mutation can occur with probability μ. This new genome is then reserved until all individuals in the patch have had a chance to mate. Then, copies of these new genomes recompose the population in the patch. The number of copies of each genome is defined by a Poisson distribution of parameter ρ, as defined in Eq. 4.3. Those individuals that failed to recombine cannot reproduce and will not be present in the next generation.

3. Migration: every individual migrates to a random patch with probability m per generation, and with a probability $1-m$ remains in the original patch. In the case of migration, the recipient patch is chosen randomly among the $M-1$ available patches.

As initial conditions, we start the simulations with an isogenic population, i.e. with the same genotype and phenotype. Each patch begins at its full carry capacity K. Though the landscape is heterogeneous, and thereby each component of the optimum phenotype $\vec{h}_k$ in patch k is chosen from a Gaussian distribution with standard deviation $\sigma = 0.3$.

We assess all our measures after the metapopulation has evolved for $30,000$ generations, ensuring that measurements are taken after an equilibrium regime has been attained. Table 2 presents the values of the parameters used in our simulations.

Parameters	Description	Values
M	number of patches	**10**
L	number of traits	**3**, 5, 7
μ	mutation rate	**0.01**
m	migration rate	0.003 - **0.005** - 0.007
α_r	width of the viability function	0.2 - 4.0
σ_h	std deviation of environmental changes	0.02, **0.05**, 0.07, 0.10
τ	period of environmental changes	**200**, 500, 1000
K	carrying capacity per patch	500
λ	per capita growth rate	**3**
g	minimum shared mutations	**10**

Table 2 – Parameters of the model. The third column either defines the value or range of values of each parameter explored along the work. In bold, the most prevalent used value of each parameter.

One of the main quantities of our interest is the number of species, hereafter species richness. The species richness relies on genetic measurements of the pool of extant distinct genotypes in the metapopulation. The other two quantities assessed in all simulations are genetic and phenotypic dissimilarities, which correspond to the average Hamming distance, π, and average distance in the phenotypic space, π_P, respectively. For both quantities, a sampling of size N^* is taken from the whole population and hence proceeds with the pairwise measurements and average estimates. For each pair of genomes i and j, the Hamming distance between them is just $\pi(i,j) = n_i + n_i - 2n_{i,j}$, where, as before, n_i is the number of mutations of genotype i, and $n_{i,j}$ is simply the number of mutations shared by the two genotypes. On the other hand, the phenotypic distance between two sampled individuals is given by

$$\pi_P(i,j) = \sqrt{\sum_k (r_{i,k} - r_{j,k})^2}, \tag{4.5}$$

and provides the Euclidian distance between the two samples individuals in the phenotypic space, whose average must be divided by the total number of possible pairwise comparisons $n(n-1)/2$:

$$\pi_p = \frac{\sum \pi_p(i,j)}{N^*(N^*-1)/2}, \tag{4.6}$$

where N^* is sampling size, and the sum runs over all pairs of the sampling.

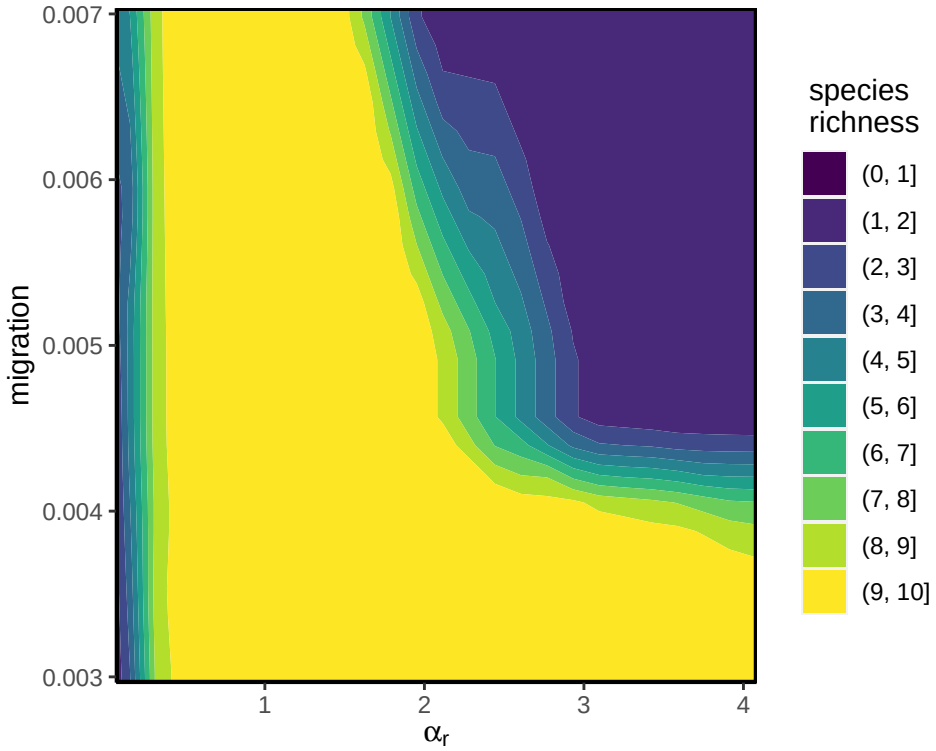
4.3 Results

4.3.1 The role of migration in shaping diversity

As previously mentioned, selective pressure and geographic isolation affect population divergence and thus shape diversity. In this metapopulation model where each of the M patches has a different phenotypic optimum, one can expect that the severity of selection will promote population divergence as different traits are better adapted to different patches.

We visit the parameter space of both metrics and observe the species richness. Figure 36 displays a heat map for the species richness in terms of the viability width α_r and migration probability m.

Figure 36 – Species richness in terms of the width α_r and migration. The selective pressure is inverse with α_r, thus from right to left we go from the neutral regime where effectively there is no selection mechanism, to a domain of increased pressure where populations are at risk of local and global extinction. The values of the remaining parameters are those in bold in table 2.



Source: The author (2024).

In the neutral regime ($\alpha_r > 3$), species richness is defined by migration probability alone. For very low migration ($m \leq 0.004$) populations are sufficiently isolated, such that the average number of migrants exchanged between patches is not enough to prevent population diver-

gence. In this extreme case, speciation is promoted solely by the stochastic and independent accumulation of different genetic variations between subpopulations of different patches. As a result, the maximum number of species equals the number of patches. As m is increased, we enhance gene flow between clusters of subpopulations, thus the average species richness decreases until migration is high enough to homogenize the whole metapopulation into a single species.

As selective pressure increases we leave the neutral regime, and individuals less adapted to their patch begin to be ruled out by viability selection, thus reducing diversity within each patch and enhancing diversity among patches. Notably, in the extreme case where $\alpha_r < 0.5$, diversity drops considerably and irrespective of migration rate. Selection strength is too punishable for individuals not close enough to the optimum, especially for those already adapted to their patch but forced to migrate to another (remember that migration is random and individuals do not evaluate if it is better to migrate or not). In this narrow range, metapopulation dynamics prevail with alternate episodes of local extinction and recolonization, and even global extinction (as shown further below in Figure 39).

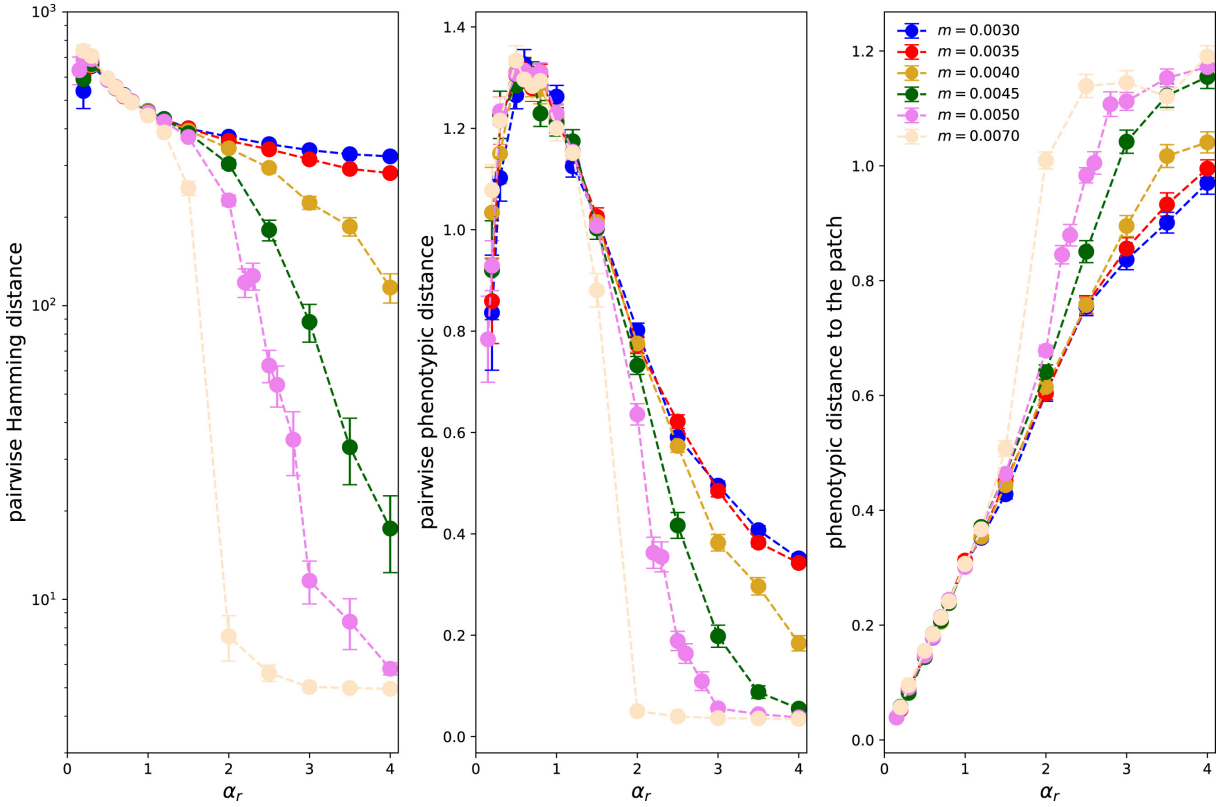
For a finer analysis of the role of migration in population divergence, we assess the level of genetic and phenotypic divergence within the population, and their distance to the optimum phenotype of their respective patch Figure 37. For all the metrics, we acknowledge the effect of migration from intermediate levels of α_r and forward, with no substantial change in the domain of severe pressure ($\alpha_r \leq 1.5$).

The role of migration in the genotypic domain is central to determining reproductive isolation and speciation. As the migration increases, genetic and phenotypic distance among the sampled individuals reduces significantly (please note the logarithmic scale in Fig. 37), leading to low species diversity. The correlation between the distances is somewhat expected as the effect of mutations on phenotype is additive.

The right panel exhibits the average phenotypic distance from the individuals in the sampling to the optimum phenotypes of the patches they lie in. Contrary to what has been observed regarding the pairwise phenotypic distance, the phenotypic distance to the patch's optimum phenotype grows with migration. This behavior is set up on the onset of the neutral regime, and the divergence between the curves increases with α_r.

Altogether, these outcomes suggest that as individuals increase their motility over the metacommunity they become phenotypically more similar. At the same time, they are phenotypically more distant from the optimum phenotype in their patch. This phenomenon occurs

Figure 37 – Average pairwise Hamming distance (left panel), average pairwise phenotypic distance (middle panel) and phenotypic distance to the optimum imposed by the inhabited patch (right panel) versus the width of the viability function, α_r. Migration probabilities are shown in the legends. The values of the remaining parameters are those in bold in table 2.



Source: The author (2024).

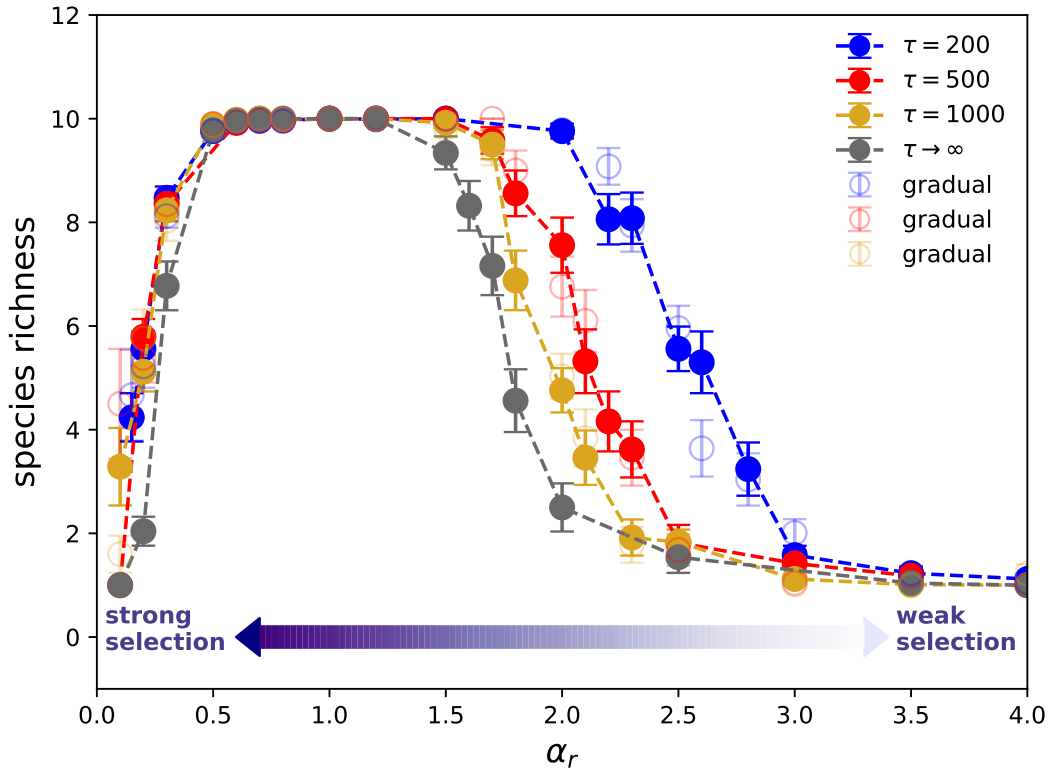
because the selection strength is reduced, allowing the population to survive in a suboptimal condition (ARAUJO et al., 2015; BROOKS; HOBERG; BOEGER, 2019), experiencing different selective pressures more frequently. Under higher migration, our results show that it is selectively advantageous to have a phenotype that reasonably matches most of the patches, as a generalist of sorts.

4.3.2 Diversity and Environmental fluctuations

In Figure 38, we investigate how the strength of selection shapes the species richness when the environment changes. Migration probability m is set at $m = 0.005$, guaranteeing a single species in the neutral regime when there is no change. Results are shown for different values of τ, the time to environmental changes, with $\tau \to \infty$ meaning no change at all.

The species richness is a one-humped function of α_r (the width of selection viability). As τ

Figure 38 – The dependence of the species richness on the width of the viability function, α_r. The selection strength or pressure can be understood as $1/\alpha_r$. Different curves denote distinct values of the period of disturbance events, τ. From right to left $\tau = 200$, $\tau = 500$, $\tau = 1000$, and $\tau \to \infty$. The open symbols correspond to the gradual counterpart of the discontinuous environmental changes upon the condition that $\sigma^2_{gradual} = \sigma^2_h/\tau$. The values of the remaining parameters are those in bold in Table 2.

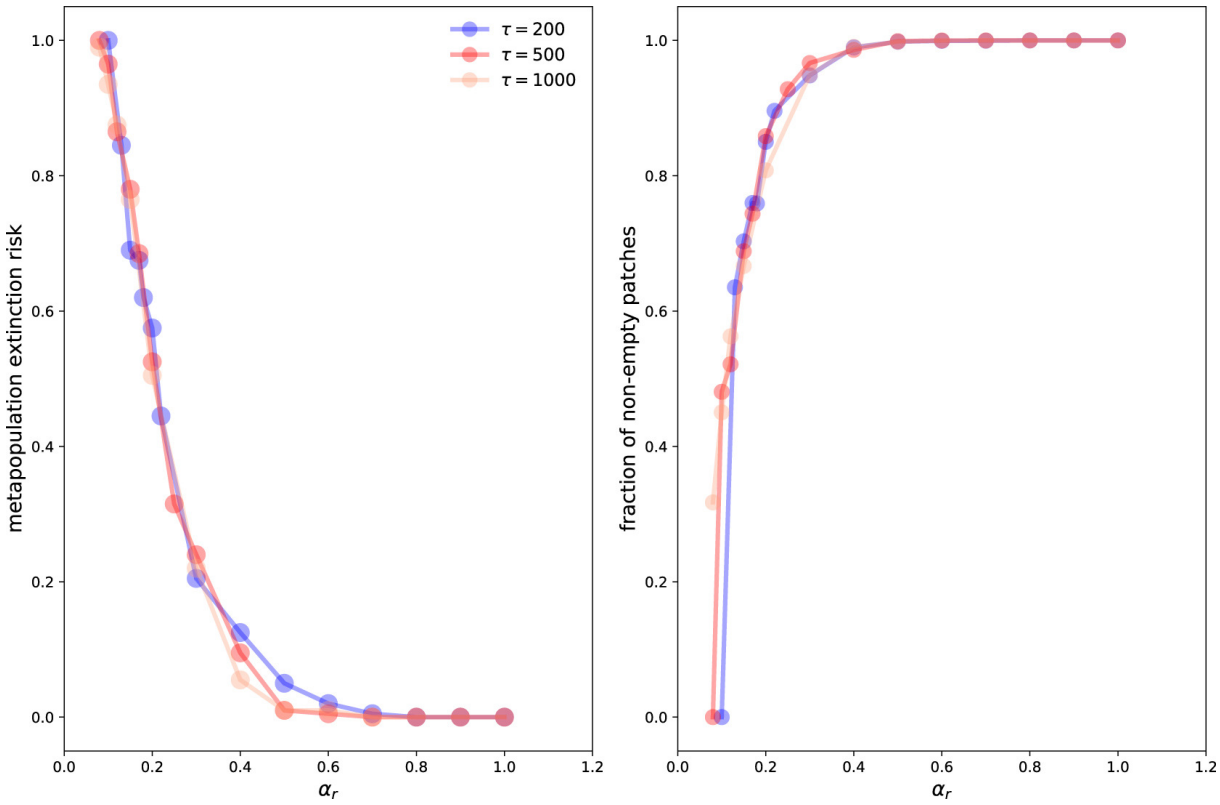


Source: The author (2024).

decreases (i.e., environmental changes are more frequent), the curves are shifted towards higher values of α_r. The extent of the interval where the species richness remains in its maximum value and the regime of weak selection also depends on τ. Thus, the homogenizing effect of dispersal depends on the rate of environmental changes.

Selective pressure goes with $1/\alpha_r$, being responsible for establishing the tolerance limits of the individuals. When α_r is small, meaning the selection is strong, the shift of the optimum phenotypes can be disruptive at a global scale, leading to population extinction. As we report in Figure 39). while the likelihood of global extinctions shrinks when α_r is augmented, events of local extinction, i.e., those in which a patch becomes empty, are still probable. Local extinctions can lead to a continuous process of extinction followed by recolonization of patches, entailing the loss of genetic variation, the so-called founder effect (BARTON; CHARLESWORTH, 1984), and hence the reduced species richness. On the other hand, when α_r increases, selection becomes softer, turning the drift of the optimum phenotype less harsh. The softness of selec-

Figure 39 – Metapopulation extinction risk (left panel) and fraction of non-empty patches (right panel) versus the width of the viability function, α_r. The periods of disturbance events, τ, are indicated in the legend.The values of the remaining parameters are those in bold in Table 2.



Source: The author (2024).

tion first leads to increased species richness within the metapopulation, as local populations can now respond to selection and evolve towards the optimum phenotype, thus avoiding local extinction. Therefore, at small and intermediate values of α_r, the variation of optimum phenotypes among patches, which is expected to be more significant than the resulting variation due to environmental changes within the patch, contributes to making subpopulations more reproductive isolated and effectively reducing gene flow. Thus, at intermediate α_r, local adaptation differentiates the subpopulations and, importantly, the degree of divergence depends on the rate of environment fluctuation.

In Figure 38, we also compare gradual vs. abrupt environmental changes. For abrupt changes, we set the magnitude at $\sigma_h = 0.05$ and different values for the time between environmental changes τ were simulated. In the same plot, we present results for their gradual counterparts (open symbols). As aforementioned, gradual environmental changes are subject to the condition $\sigma_{gradual} = \sigma_h/\sqrt{\tau}$. In a nutshell, we can state that the curves nearly collapse, meaning that the pattern of speciation is practically independent of the mode and rate of

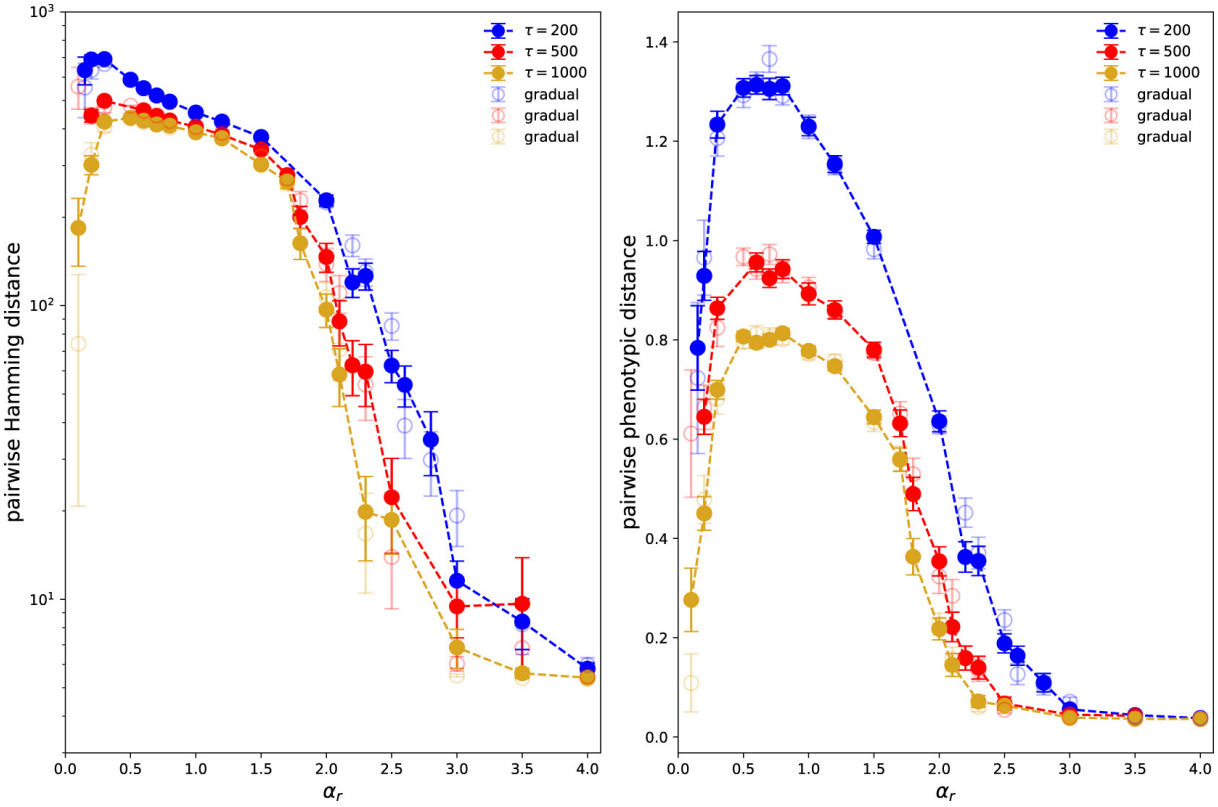
environmental changes. Quite surprisingly, these outcomes demonstrate that, at least for fluctuating environments, what matters is the magnitude of the net displacement in the phenotypic space over a given time interval.

Although this equivalence is expected for the moving optimum, it is not a trivial matter that the population, evolving at its own time-scale and mechanisms, would bear similar results at the speciation level. Note that the metapopulation is constantly under viability selection. Gradual changes are more frequent and less drastic (σ_{grad} at each generation), thus the selection response to environmental changes is smoother, and evolution towards the optimum phenotype can be achieved through a smaller number of generations. On the other hand, when the events take a long time to occur and are more pronounced (σ_h at each τ generations), there is an abrupt drop in the viability of the local population, thus the response to selection is intense. More generations are expected for the population to reach viability around the new optimum phenotype. Despite that, the rate at which these changes occur seems to have no noticeable effect once the overall effect is the same.

In order to comprehend the macroevolutionary scenario, it is important to see how it relates to the metapopulation dynamics at a finer scale. Figure 40 shows the dependence of the genetic and phenotypic distance among individuals on the width α_r.

Both distances have a similar pattern of the species richness (Fig. 2): they first increase with α_r reaching a maximum and then decrease. Although strong selection imposes differentiation among individuals, it also increases mortality, extinguishing individuals from some patches (Fig. 3). Then, a little reduction in the selection strength avoids extinctions and increases the overall differentiation. Interestingly, different from its genetic counterpart, phenotypic distance is sensitive to environmental fluctuations, and its highest phenotypic distance occurs under the smaller period of environmental changes, probably because the higher number of changing events increases the chance that the optimum phenotypes are farther from each other. Once again, results for abrupt environmental changes match their gradual counterparts. The shape of the curves is basically unaltered, and a collapse of the curves is still observed.

Figure 40 – Average pairwise Hamming distance (left panel) and average pairwise phenotypic distance (right panel) as a function of the width α_r. Different values of the time to environmental changes are encoded by different colors, as indicated in the legends. Data are compared to their gradual counterparts, in which $\sigma_{gradual} = \sigma_h/\sqrt{\tau}$. The values of the remaining parameters are those in bold in table 2.



Source: The author (2024).

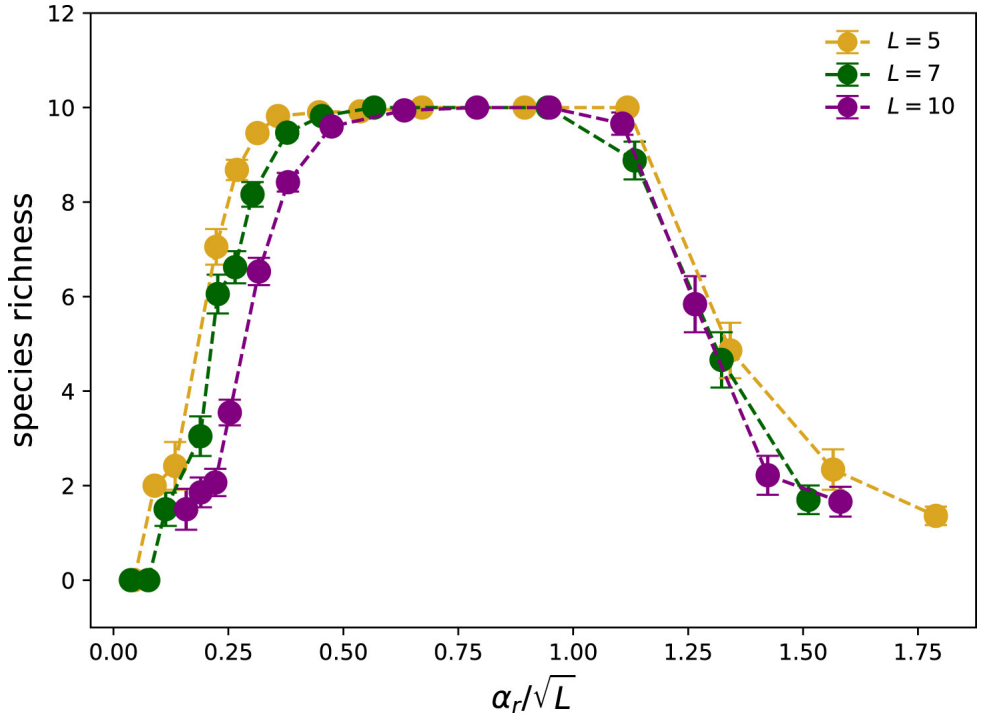
4.3.3 Further investigation of parameter space

In this last section, we extend the results of the model in two ways: 1) exploring their sensibility to model parameters - the number of traits L and patches' carry capacity K; and 2) modifying the selection mechanism to include directional selection, the effects of a global trait and an ingenuous implementation of phenotypic plasticity.

Dependence on the number of traits

The number of traits, L, is the dimensionality of the phenotypic space and the number of terms in the argument of the Gaussian viability function, as defined in Eq. (4.1). From our knowledge of classical statistical mechanics, and similar reasoning applied to random walks, the squared distance in the numerator of the Gaussian function should scale with L. Because α_r appears in the equation as $1/\alpha_r^2$, we expect a rescaling factor of α_r/L^2.

Figure 41 – Dependence on the number of traits. Species richness vs. the rescale quantity $\alpha_r/\sqrt{L}$, where L is the number of traits (values indicated in the legend). The values of the remaining parameters are those in bold in table 2.



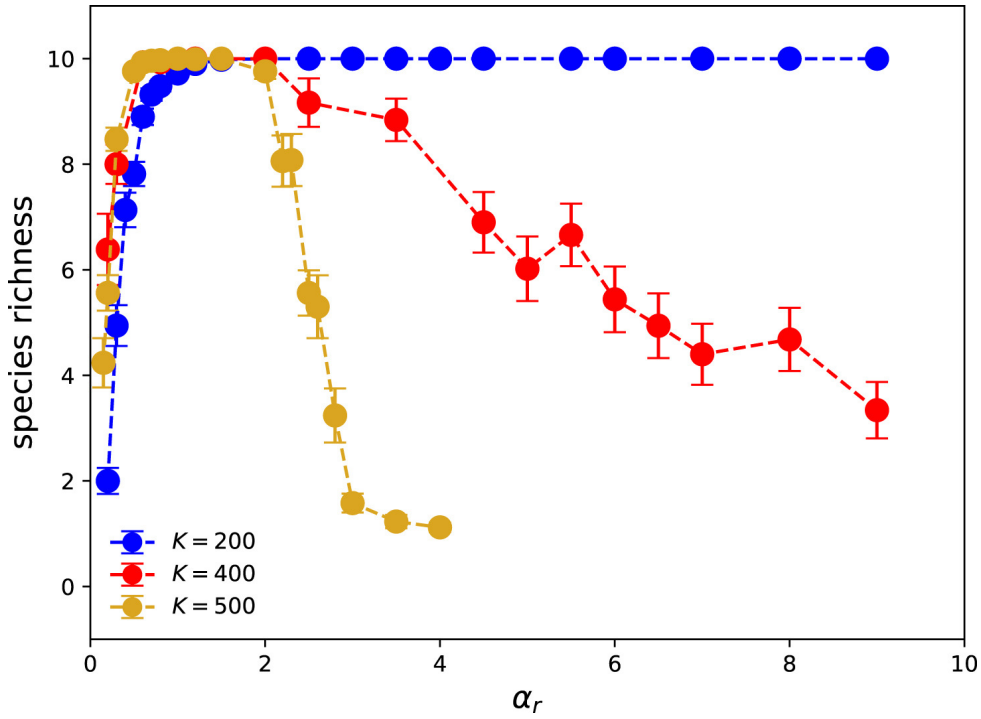
Source: The author (2024).

In fact, we observe in Figure 41 a collapse of the curves for different values of L, showing that the effect of the number of traits is to shift the whole curve, keeping its shape unaltered, by a factor proportional to $\sqrt{L}$.

Dependence with carry capacity

For the sake of completeness, in Fig. 42 we show the effect of the carrying capacity on the pattern of speciation. We observe that changing the carrying capacity K while keeping the migration rate constant is similar to changing migration while keeping the carrying capacity fixed. $K = 500$ corresponds to the prevalent value used in our simulations. When the carrying capacity is reduced to $K = 400$, the onset of the neutral regime shifts towards higher α_r, and when it is further reduced to $K = 200$, the species richness plateaus at the highest diversity for the majority of α_r. This behavior is quite similar to the one displayed in Figure 36 for migration.

Figure 42 – Dependence on the carrying capacity K. Species richness vs. the width α_r for different values of K, indicated in the legend. The values of the remaining parameters are those in bold in table 2.



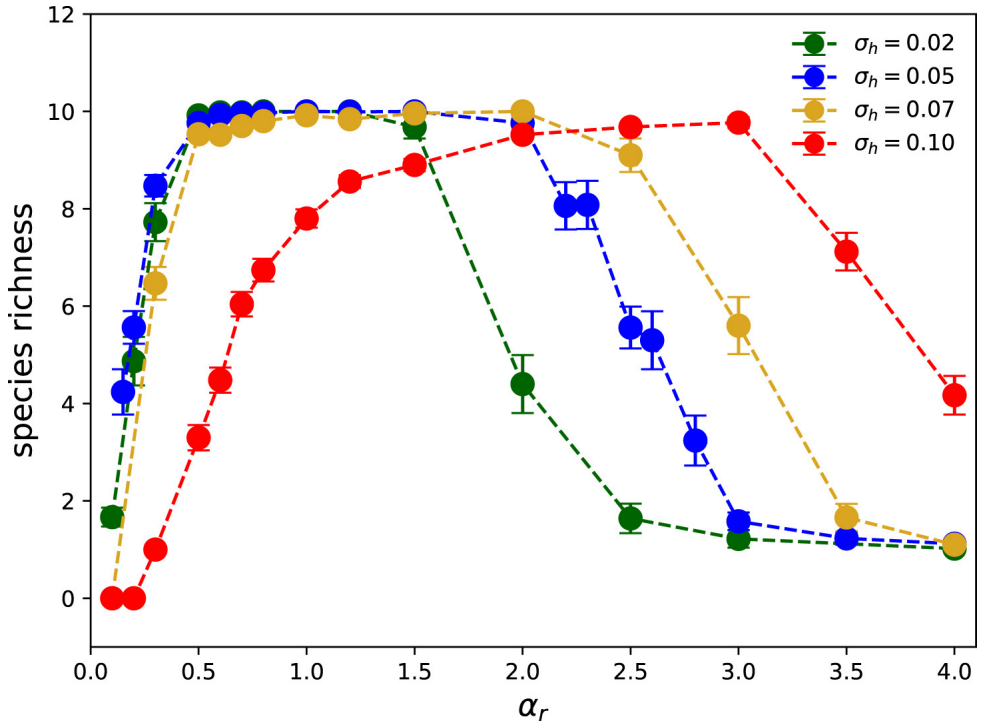
Source: The author (2024).

Furthermore, as discussed in Chapter 1, genetic drift is stronger in populations of small sizes, which enhances their genetic divergence among patches. These results lead us to conclude that what matters is the number or flux of migrants per generation ($N\,m$, where N is the population size at a given generation) since this number is directly proportional to both migration rate and carrying capacity.

Dependence on the magnitude of environmental changes

In Figure 43, we observe the influence of the variance of environmental changes, σ_h^2 on species richness. The rise of σ_h^2 might bring about more severe drops of the viability within a patch at the time of the optimum phenotype shift, especially when selection is intense. We notice that the domain of α_r, in which the events of local extinction followed by recolonization are relevant, is broadened with σ_h^2. This region is characterized by the growth of the species richness with α_r.

Figure 43 – Dependence of the species richness on the width of the viability function, α_r. The standard deviations for environmental changes are indicated in the legend. The values of the remaining parameters are those in bold in Table 2.



Source: The author (2024).

Directional selection - gradual vs. abrupt changes

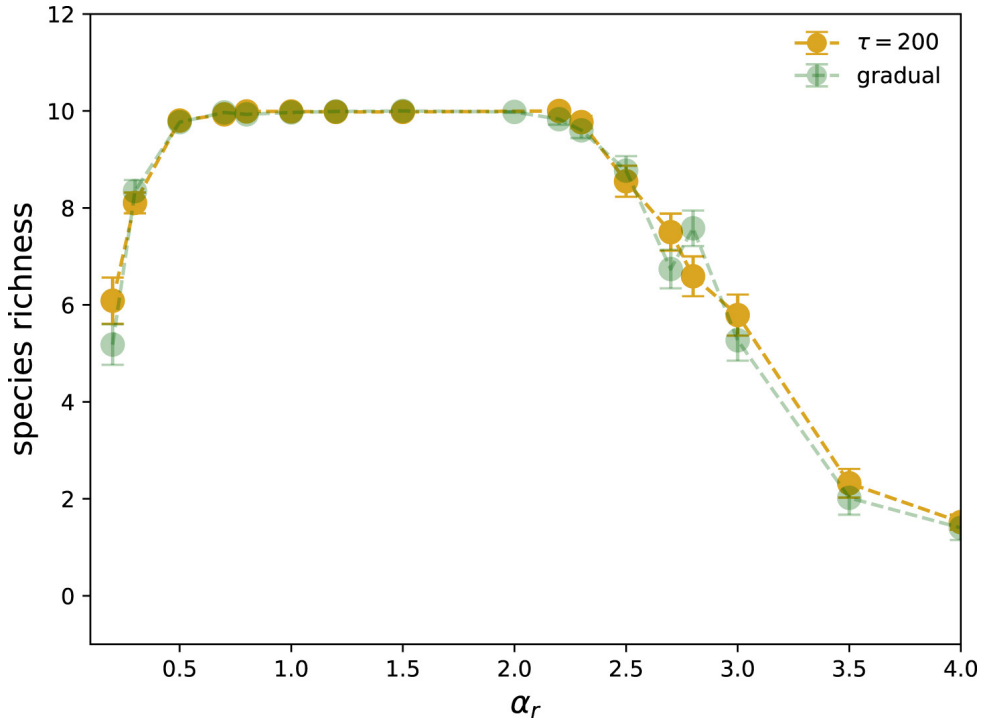
Here we will consider a slight modification of our modeling and include a directional component of selection, such that the population faces a combination of directional and stabilizing selection. With directional selection at a single trait j, the $j-$th component of the optimum phenotype at patch k at the $n-$th environmental change becomes

$$h_{k,j} = h^0_{k,j} + \mathcal{V}n + \Upsilon, \tag{4.7}$$

where $\mathcal{V}$ denotes the directional component of selection, Υ is the sum of n Gaussian variable of null mean and variance σ_h^2, and $h^0_{k,j}$ is its initial value. In order to make a correct comparison between gradual and abrupt changes, $\mathcal{V}_{gradual}=\mathcal{V}/\tau$, and as already established $\sigma^2_{gradual} = \sigma_h^2/\tau$.

Figure 44 displays the species richness vs. α_r under a metapopulation subject to directional and stabilizing selection. In the plot, we have chosen the directional component slightly lower than the fluctuations $\mathcal{V} = \sigma_h/10$. Results are presented for abrupt changes taking place every $\tau = 200$ generations and gradual changes under our rescale, as emphasized above. Once again,

Figure 44 – Dependence of the species richness on the width of the viability function, α_r under directional ($\mathcal{V} = 0.005$) and stabilizing selection. The two curves differ by abrupt or gradual environmental disturbances $\sigma^2_{gradual} = \sigma^2_h/\tau$, with $\sigma_h = 0.05$. The values of the remaining parameters are those in bold in table 2.



Source: The author (2024).

the agreement between the two scenarios is remarkable. Therefore, the pattern of speciation remains unaltered even when directional selection components are considered.

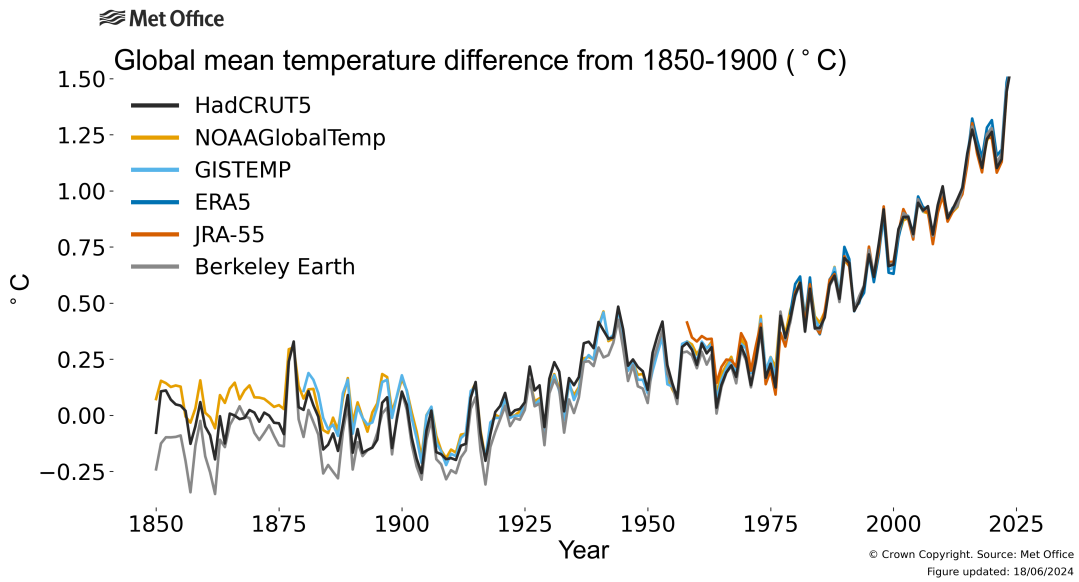
Global vs. local traits

In the model, local adaptation manifests as a phenotypic optimum, a L-dimensional position in trait space, unique to each patch. Biologically speaking, we are assuming that the conditions or resources necessary to strive in each patch are not only different but also independent. On the other hand, abiotic pervasive factors such as temperature (Fig 45) and atmospheric CO_2 concentration can directly affect ecosystems' function as a whole and drive population evolution at a global-scale.

That said, it seems natural to inquire how adaptation to a global selective trait pervasive to all metapopulation, e.g. temperature, may alter the emergence of diversity. To that, we maintain our protocol of drawing a phenotypic position $\vec{h}$ from a Gaussian distribution with standard deviation $\sigma = 0.3$, but now the $j-$th component of the optimum phenotype of all

[3] <https://climate.metoffice.cloud/current$_warming.html$ >

Figure 45 – Annual global average temperatures expressed as the difference from pre-industrial conditions based on the 1850-1900 average. Four different data sets are shown - HadCRUT5, NOAAGlobalTemp, GISTEMP, and Berkeley Earth - as well as two reanalyses - ERA5 and JRA-55. There is good agreement on the overall evolution of global temperatures and year-to-year variability.



Source: Public image accessed in 2025. [3]

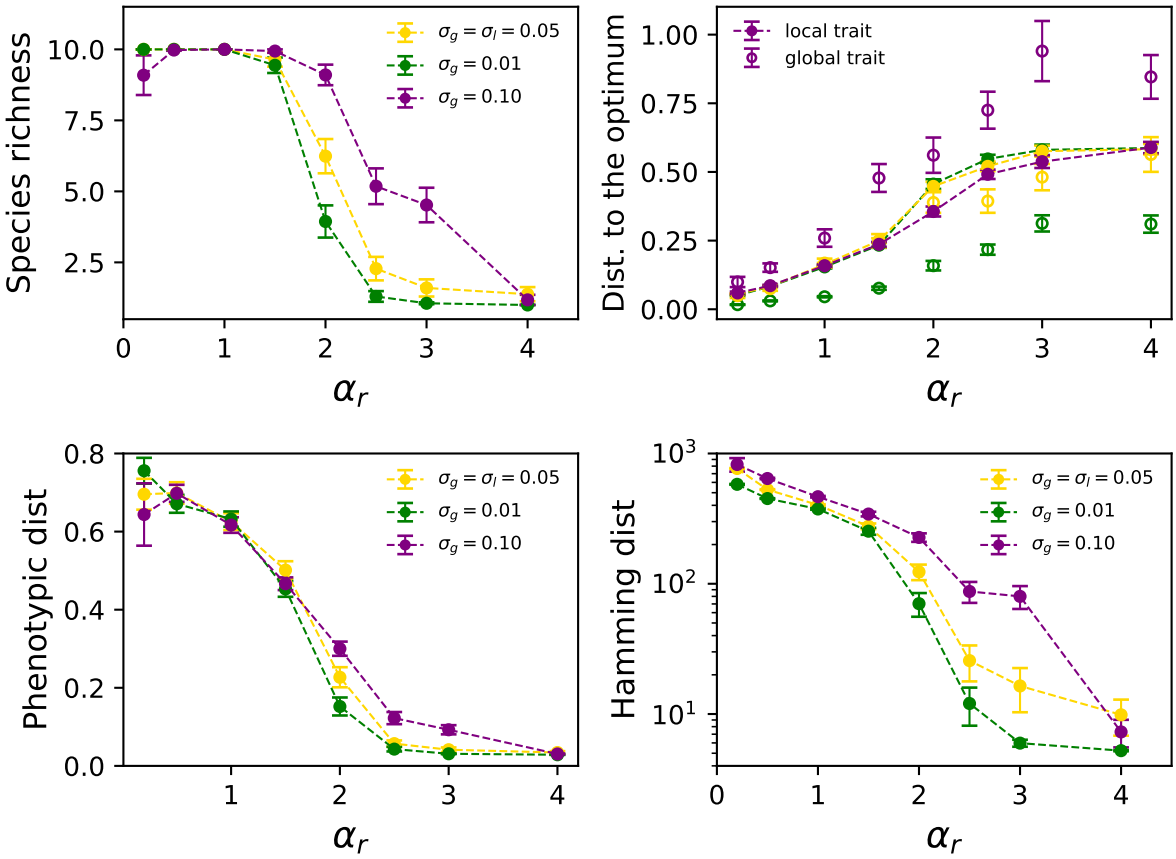
patches is the same, i.e. this single trait is global and felt by all metapopulation irrespective of the patch the individuals are inhabiting.

In Figure 46, we show the pattern of species richness and the metrics of population divergence at genotypic and phenotypic levels, for different magnitudes of environmental changes relative to the local and global traits, hereafter σ_l and σ_g. In the results, σ_l is fixed at 0.05. Species richness, genotypic and phenotypic divergence show a monotonic response with σ_g, with increased magnitude on the global trait leading to increased levels of population divergence among the sampled individuals. On the upper right panel, we separate the phenotypic distance among the global and local traits. Population can more easily follow the global trait the lower the σ_g. Interestingly, on the intermediate values of α_r the mismatch on the global trait results in populations more closely adapted to their local traits.

Effects of phenotypic plasticity

The capacity of the population to adapt to environmental changes is intrinsically related to how fast population variation can emerge from the dynamics. In our model, mutation and recombination are the main drivers of population variation and both happen at a time-scale of discrete generations - from parents to their offspring.

Figure 46 – Species richness, average pairwise Hamming distance, average pairwise phenotypic distance and phenotypic distance to the optimum imposed by the inhabited patch versus the width of the viability function, α_r. Dependence on the magnitude of global trait variation σ_g is indicated in the legend. In the upper right panel, we distinguish the distance to local and global traits - full and empty points, respectively. The values of the remaining parameters are those in bold in table 2.



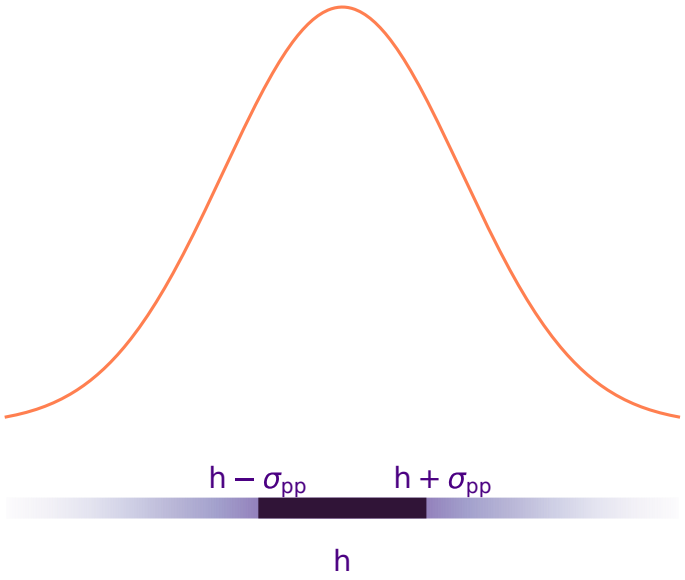
Source: The author (2024).

In reality, as a response to biotic and abiotic changes, some individuals are capable of altering their phenotypic traits during their lifetime without changes to their genetic counterparts. The ability to express different phenotypes from the same set of genotypes is called phenotypic plasticity (see Chapter 1 for discussion).

Here we imbue our metapopulation with such capacity by extending the range in which the optimum phenotype is reachable. We assume that individuals near enough of the phenotypic optimum by a distance σ_{pp} may reach the benefits of being exactly in the optimum through phenotypic plasticity (see scheme in Figure 47). In other words, if $|\vec{r} - \vec{h}| < \sigma_{pp}$ then $V = 1$.

Importantly, over the years, we realized this approach has been rather ingenious. Firstly, because plasticity is not a property of the environment but of the organism (LALAND et al., 2015). Secondly, the use of plasticity comes with an inherent fitness cost due to the biochemical

Figure 47 – One-dimensional representation of our modeling of phenotypic plasticity. The optimum phenotype is centered at a point h of the trait axis. Here, individuals whose phenotype is at a given distance from the optimum (purple interval), are considered sufficiently adapted to survive viability selection Eq. 4.1. In other words, if $|r - h| < \sigma_{pp}$ then $V = 1$.

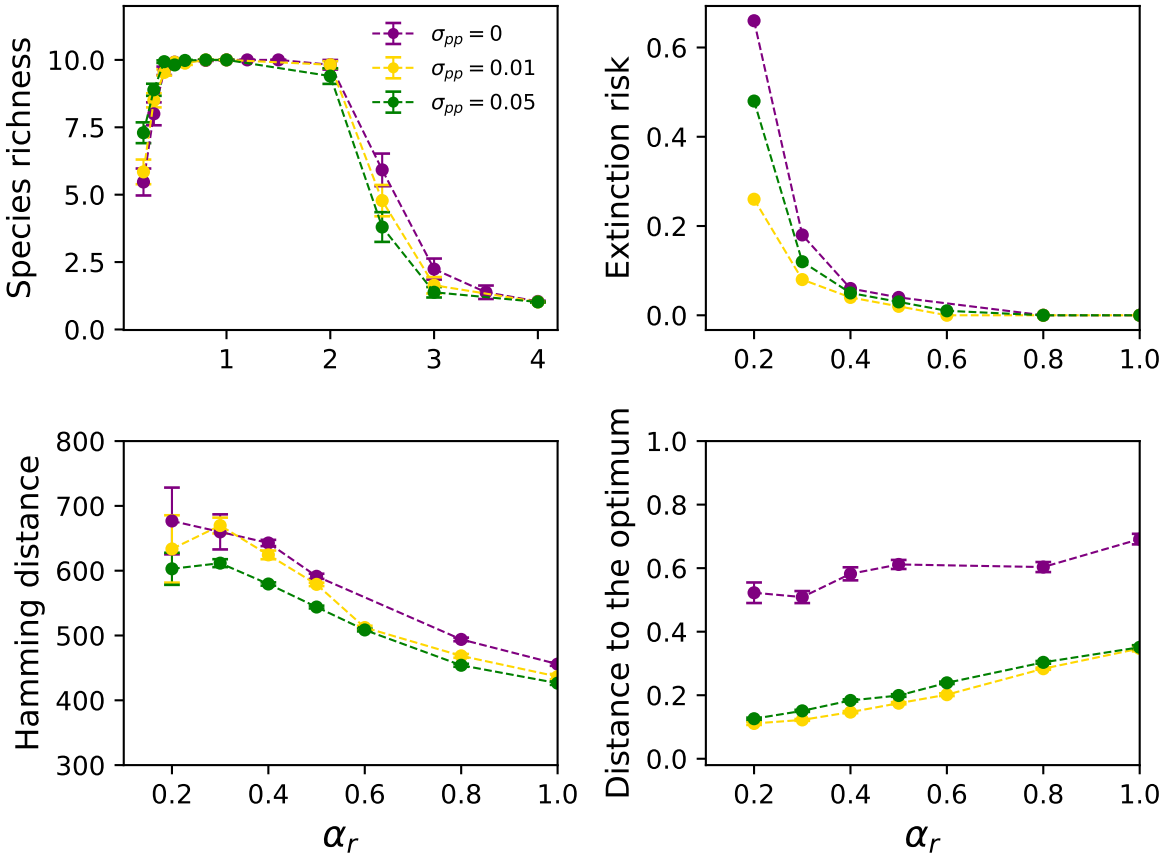


Source: The author (2024).

and physiological machinery that individuals must have to sense the environment during their lifetime and build a proper response.

Regardless, Figure 48 shows our results for different levels of plasticity given by the parameter σ_{pp}. We can see that through plasticity the populations have their divergence reduced both at genotypic and phenotypic level, which results in less number of species maintained in the interval of intermediate selective pressure. Although this difference is not so drastic at species richness and hamming distance, it is more evident at the phenotypic divergence, with both cases equally closer to their patch optimum. As expected, with this implementation, populations are less punished for not being in the patch optimum which is reflected in the decrease of metapopulation extinction risk for higher σ_{pp}.

Figure 48 – Species richness, average pairwise Hamming distance, phenotypic distance to the optimum and extinction risk for the effect of phenotypic plasticity σ_{pp} versus the width of the viability function, α_r. The values of the remaining parameters are those in bold in table 2.



Source: The author (2024).

4.4 Conclusions

There is a long-standing debate about the role of climatic and ecological changes in shaping biodiversity. Little is known whether the biotic responses are a consequence of fast, rapid, or gradual climate change (BOTTA et al., 2019). While some empirical studies suggest that evolutionary responses are slower relative to environmental change, i.e., environmental changes are abrupt (DAVIS; SHAW; ETTERSON, 2005), taking as a special case climate change (HUNTLEY et al., 1991), other studies claim that such changes are gradual or even that both modes co-occur (OSGOOD; WHITE; BAUM, 2021). Biological responses to environmental change occur through migration and phenotypic or molecular evolution, and when failing, populations and species can go locally or even globally extinct (DAVIS; SHAW; ETTERSON, 2005; LORENZEN et al., 2011).

We proposed a metapopulation model that employs the framework of Fisher's geometric model to study how the pattern of diversification and speciation is influenced by environmental changes and their modes of occurrence. Quite surprisingly, we found that the pattern of diversification and speciation is only dependent on the magnitude of the environmental changes in a given time interval, regardless of whether the changes are abrupt or ensue from repeated processes of smaller variations. According to our approach, the magnitude of environmental changes corresponds to each patch displacements of the optimum phenotype. We draw the displacements from a Gaussian distribution of null mean and fixed variance. By rescaling the variance, here using the reasoning of the Central Limit Theorem, one can generate distributions of events that are additive and of smaller magnitudes that are equivalent to those resulting from single events. Because the Central Limit Theorem ensures that the distribution of a sum of independent stochastic variables is Gaussian-distributed, regardless of the distribution of the stochastic variables, our claim is not dependent on the particular choice for the distribution of the magnitude of environmental changes. Our claim remains unaltered even in the presence of directional selection.

The aforementioned achievements are clearly of difficult empirical verification in natural populations, but one expects them to be feasible in well-controlled lab conditions. Nevertheless, our framework is a stepping-stone for future investigation regarding the effects of different metapopulation structures for studying the effects of landscape fragmentation, proper modeling of phenotypic plasticity, and quantification of both the founder effect and the evolutionary rescue processes happening in the domain of strong selection.

- The main results presented in this section have been published in the article: *Freitas, Osmar, Sabrina BL Araujo, and Paulo RA Campos. "Speciation in a metapopulation model upon environmental changes." Ecological Modelling (2022).* Ref. (FREITAS; ARAUJO; CAMPOS, 2022)

5 OVERALL CONCLUSIONS

We are most likely to protect what we appreciate,
and we are mostly likely to appreciate what we understand.
By learning about the processes and diversity of life,
we also become more aware of ourselves
and our place in the biosphere.
(*Neil Campbell)*

In this thesis, we have addressed the implications of environmental changes to population adaptation and diversity through stochastic evolutionary models.

In the first work, we study population persistence in the form of evolutionary rescue, where adaptation must happen sufficiently fast to ensure population survival. First, we show that a density-dependent growth can indeed obscure the observation of rescue in declining populations, since extinction can be prevented by adjusting its size to better cope with the stressful environment. A merit of our model is to show a clear delineation of two regimes separated by a critical level of stress, where, in the latter, the population can only persist through the establishment of rescue mutations. The probability of extinction increases with stress because the availability of rescue mutations decreases. Therefore, the order in which mutations arise in the population becomes more essential to rescue in more stressful environments. Next, we make some simplifications of our model to study the effects of epistatic genetic interactions in evolutionary rescue. Although high levels of epistasis constrain the accessibility of evolutionary trajectories over fitness landscapes, high epistasis can provide beneficial mutations with sufficiently large fitness effects that can rescue the population. In this tradeoff between mutation availability and its effects, we show that low to intermediate values of epistasis have an increased risk for population extinction.

Our second work studies population divergence in an insular system subject to isolation mechanisms driven by environmental events of geographic barriers. We show how gene flow, in the form of migration between islands, influences speciation. By registering the speciation dynamics in a phylogeny, we infer signatures of the intermittent geographical events in the population history. In our model, higher rates of speciation are expected when populations

spend too much time in isolation, and the distribution of speciation events among lineages involves the distribution of species abundances among islands, which are dependent on the intermittent regimes of connection and isolation.

At last, we study a metapopulation model under environmental fluctuations, where both migration and local adaptation influence population divergence and persistence. We make an extensive statistical analysis on the effects of gene flow and selective pressure on population diversity, reported by species richness but also by their genotypic and phenotypic divergence. We study the influence of environmental perturbations in function of selective pressure, going from a neutral regime where divergence is solely promoted by migration, to a stronger selective regime where local and global extinction risk are enhanced. We see that more frequent perturbations promote higher divergence at the phenotypic level, which in turn is accompanied by genetic divergence, resulting in more species diversity. We compare the effect of abrupt and gradual disturbances and show that, in our model, the pattern of speciation is only dependent on the net effect of the environmental disturbances, despite the rate at which the events occur.

The works here discussed have their own limitations and the potential to be expanded in a number of interesting ways, as described in their sectorial conclusions. My only hope is that they can offer us insights and reflections into the processes affecting life's diversity and resilience.

REFERENCES

AGUIAR, M. A. M. D.; BARANGER, M.; BAPTESTINI, E.; KAUFMAN, L.; BAR-YAM, Y. Global patterns of speciation and diversity. *Nature*, Nature Publishing Group, v. 460, n. 7253, p. 384–387, 2009.

AGUILéE, R.; LAMBERT, A.; CLAESSEN, D. Ecological speciation in dynamic landscapes. *Journal of Evolutionary Biology*, v. 24, n. 12, p. 2663–2677, 2011. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1420-9101.2011.02392.x>.

AITA, T.; HUSIMI, Y. Fitness spectrum among random mutants on mt. fuji-type fitness landscape. *Journal of theoretical biology*, Elsevier, v. 182, n. 4, p. 469–485, 1996.

ALENCAR, L. R. V. de; QUENTAL, T. B. Linking population-level and microevolutionary processes to understand speciation dynamics at the macroevolutionary scale. *Ecology and Evolution*, v. 11, n. 11, p. 5828–5843, 2021. ISSN 20457758.

ALEXANDER, H. K.; MARTIN, G.; MARTIN, O. Y.; BONHOEFFER, S. Evolutionary rescue: linking theory for conservation and medicine. *Evolutionary applications*, Wiley Online Library, v. 7, n. 10, p. 1161–1179, 2014.

ANCIAUX, Y.; CHEVIN, L.-M.; RONCE, O.; MARTIN, G. Evolutionary rescue over a fitness landscape. *Genetics*, Oxford University Press, v. 209, n. 1, p. 265–279, 2018.

ARAUJO, S. B.; BRAGA, M. P.; BROOKS, D. R.; AGOSTA, S. J.; HOBERG, E. P.; HARTENTHAL, F. W. V.; BOEGER, W. A. Understanding host-switching by ecological fitting. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 10, n. 10, p. e0139225, 2015.

BAGGIO, R. A.; STOIEV, S. B.; SPACH, H. L.; BOEGER, W. A. Opportunity and taxon pulse: the central influence of coastal geomorphology on genetic diversification and endemism of strict estuarine species. *Journal of Biogeography*, v. 44, n. 7, p. 1626–1639, 2017. ISSN 13652699.

BAILEY, S. F.; BLANQUART, F.; BATAILLON, T.; KASSEN, R. What drives parallel evolution? *BioEssays*, v. 39, n. 1, p. e201600176, 2017.

BAILEY, S. F.; RODRIGUE, N.; KASSEN, R. The Effect of Selection Environment on the Probability of Parallel Evolution. *Molecular Biology and Evolution*, v. 32, n. 6, p. 1436–1448, 2015.

BANK, C. Epistasis and adaptation on fitness landscapes. *Annual review of ecology, evolution, and systematics*, Annual Reviews, v. 53, p. 457–479, 2022.

BARRETT, R. D.; HENDRY, A. P. Evolutionary rescue under environmental change? *Behavioural responses to a changing world*, Oxford University Press, p. 216–233, 2012.

BARRICK, J. E.; LENSKI, R. E. Genome dynamics during experimental evolution. *Nature Reviews Genetics*, Nature Publishing Group, v. 14, n. 12, p. 827–839, 2013. ISSN 14710056. Available at: <http://dx.doi.org/10.1038/nrg3564>.

BARTON, N. H.; CHARLESWORTH, B. Genetic revolutions, founder effects, and speciation. *Annual review of ecology and systematics*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 15, n. 1, p. 133–164, 1984.

BELL, G. Evolutionary rescue and the limits of adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, The Royal Society, v. 368, n. 1610, p. 20120080, 2013.

BELL, G. Evolutionary rescue. *Annual Review of Ecology, Evolution, and Systematics*, Annual Reviews, v. 48, p. 605–627, 2017.

BELL, G.; GONZALEZ, A. Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science*, American Association for the Advancement of Science, v. 332, n. 6035, p. 1327–1330, 2011.

BLOUNT, Z. D.; LENSKI, R. E.; LOSOS, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science*, v. 362, p. eaam5979, 2018.

BLUM, M. G. B.; FRANçOIS, O. Which Random Processes Describe the Tree of Life? A Large-Scale Study of Phylogenetic Tree Imbalance. *Systematic Biology*, v. 55, n. 4, p. 685–691, 08 2006. ISSN 1063-5157. Available at: <https://doi.org/10.1080/10635150600889625>.

BOEGER, W. A.; BROOKS, D. R.; TRIVELLONE, V.; AGOSTA, S. J.; HOBERG, E. P. Ecological super-spreaders drive host–range oscillations: Omicron and risk space for emerging infectious disease. *Transboundary and Emerging Diseases*, v. 69, n. 5, p. e1280–e1288, 2022. ISSN 18651682.

BOLNICK, D. I.; BARRETT, R. D.; OKE, K. B.; RENNISON, D. J.; STUART, Y. E. (Non)parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, v. 49, n. 1, p. 303–330, 2018. ISSN 1543-592X.

BOTELHO, L. L.; MARQUITTI, F. M. D.; AGUIAR, M. A. M. de. Extinction and hybridization in a neutral model of speciation. *Journal of Physics A: Mathematical and Theoretical*, IOP Publishing, v. 55, n. 38, p. 385601, aug 2022. Available at: <https://dx.doi.org/10.1088/1751-8121/ac88a5>.

BOTTA, F.; DAHL-JENSEN, D.; RAHBEK, C.; SVENSSON, A.; NOGUÉS-BRAVO, D. Abrupt change in climate and biotic systems. *Current Biology*, Elsevier, v. 29, n. 19, p. R1045–R1054, 2019.

BRAGA, M. P.; ARAUJO, S. B.; AGOSTA, S.; BROOKS, D.; HOBERG, E.; NYLIN, S.; JANZ, N.; BOEGER, W. A. Host use dynamics in a heterogeneous fitness landscape generates oscillations in host range and diversification. *Evolution*, Wiley Online Library, v. 72, n. 9, p. 1773–1783, 2018.

BROOKS, D. R.; HOBERG, E. P.; BOEGER, W. A. *The Stockholm paradigm: climate change and emerging disease*. [S.l.]: University of Chicago Press, 2019.

BÜRGER, R. *Genetic Structure and Selection in Subdivided Populations by F. Rousset*. [S.l.]: Wiley Online Library, 2006.

BURKE, M. K.; LITI, G.; LONG, A. D. Standing Genetic Variation Drives Repeatable Experimental Evolution in Outcrossing Populations of Saccharomyces cerevisiae. *Molecular Biology and Evolution*, v. 31, n. 12, p. 3228–3239, 2014. ISSN 0737-4038. Available at: <https://doi.org/10.1093/molbev/msu256>.

CAMPOS, P. R.; ADAMI, C.; WILKE, C. O. Optimal adaptive performance and delocalization in nk fitness landscapes. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 304, n. 3-4, p. 495–506, 2002.

CAMPOS, P. R.; WAHL, L. M. The effects of population bottlenecks on clonal interference, and the adaptation effective population size. *Evolution*, Blackwell Publishing Inc Malden, USA, v. 63, n. 4, p. 950–958, 2009.

CAMPOS, P. R. A.; ROSAS, A.; OLIVEIRA, V. M. de; GOMES, M. A. F. Effect of landscape structure on species diversity. *PLOS ONE*, Public Library of Science, v. 8, n. 6, p. 1–10, 06 2013. Available at: <https://doi.org/10.1371/journal.pone.0066495>.

CARLSON, S. M.; CUNNINGHAM, C. J.; WESTLEY, P. A. Evolutionary rescue in a changing world. *Trends in Ecology & Evolution*, Elsevier, v. 29, n. 9, p. 521–530, 2014.

CARNAVAL, A. C.; HICKERSON, M. J.; HADDAD, C. F.; RODRIGUES, M. T.; MORITZ, C. Stability predicts genetic diversity in the brazilian atlantic forest hotspot. *Science*, American Association for the Advancement of Science, v. 323, n. 5915, p. 785–789, 2009.

CARON, F. S.; PIE, M. R. The phylogenetic signal of diversification rates. *Journal of Zoological Systematics and Evolutionary Research*, v. 58, n. 4, p. 1432–1436, 2020. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jzs.12379>.

CATALÁN, P.; GARCÍA-MARTÍN, J. A.; AGUIRRE, J.; CUESTA, J. A.; MANRUBIA, S. Entropic contribution to phenotype fitness. *Journal of Physics A: Mathematical and Theoretical*, IOP Publishing, v. 56, n. 34, p. 345601, 2023.

CHEVIN, L.-M.; LANDE, R. When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population? *Evolution*, Blackwell Publishing Inc Malden, USA, v. 64, n. 4, p. 1143–1150, 2010.

CHEVIN, L.-M.; LANDE, R.; MACE, G. M. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS biology*, Public Library of Science San Francisco, USA, v. 8, n. 4, p. e1000357, 2010.

CIRNE, D.; CAMPOS, P. R. Rate of environmental variation impacts the predictability in evolution. *Physical Review E*, APS, v. 106, n. 6, p. 064408, 2022.

CONNALLON, T.; CLARK, A. G. Balancing selection in species with separate sexes: insights from fisher's geometric model. *Genetics*, Oxford University Press, v. 197, n. 3, p. 991–1006, 2014.

COSTA, C. L.; LEMOS-COSTA, P.; MARQUITTI, F. M.; FERNANDES, L. D.; RAMOS, M. F.; SCHNEIDER, D. M.; MARTINS, A. B.; De Aguiar, M. A. Signatures of Microevolutionary Processes in Phylogenetic Patterns. *Systematic Biology*, v. 68, n. 1, p. 131–144, 2019. ISSN 1076836X.

COSTA, C. L.; MARQUITTI, F. M.; PEREZ, S. I.; SCHNEIDER, D. M.; RAMOS, M. F.; AGUIAR, M. A. de. Registering the evolutionary history in individual-based models of speciation. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 510, p. 1–14, nov 2018. Available at: <https://doi.org/10.1016%2Fj.physa.2018.05.150>.

COYNE, J. A.; ORR, H. A. *Speciation*. 1. ed. [S.l.]: Sinauer Associates, Inc., 2004. ISBN 0878930892.

CROW, J.; KIMURA, M. *An Introduction to Population Genetics Theory*. Burgess Publishing Company, 1970. ISBN 9780808729013. Available at: <https://books.google.com.br/books?id=MLETAQAAIAAJ>.

DARWIN, C. *On the Origin of Species by Means of Natural Selection*. London: Murray, 1859. Or the Preservation of Favored Races in the Struggle for Life.

DAVIS, M. B.; SHAW, R. G.; ETTERSON, J. R. Evolutionary responses to changing climate. *Ecology*, Wiley Online Library, v. 86, n. 7, p. 1704–1714, 2005.

De Visser, J. A. G.; KRUG, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, Nature Publishing Group, v. 15, n. 7, p. 480–490, 2014. ISSN 14710064.

DEANGELIS, D. L.; GRIMM, V. Individual-based models in ecology after four decades. *F1000prime reports*, v. 6, p. 39, 2014.

DERRIDA, B. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, APS, v. 24, n. 5, p. 2613, 1981.

DESAI, M. M.; FISHER, D. S. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*, v. 176, n. 3, p. 1759–1798, 2007. ISSN 00166731.

DOBZHANSKY, T. *Genetics of the evolutionary process*. [S.l.]: Columbia University Press, 1970.

DROSSEL, B. Biological evolution and statistical physics. *Advances in physics*, Taylor & Francis, v. 50, n. 2, p. 209–295, 2001.

D'BASTIANI, E.; PRINCEPE, D.; MARQUITTI, F. M. D.; BOEGER, W. A.; CAMPIãO, K. M.; ARAUJO, S. B. L. Effect of host-switching on the ecological and evolutionary patterns of parasites. *Systematic Biology*, 04 2023. ISSN 1063-5157. Syad022. Available at: <https://doi.org/10.1093/sysbio/syad022>.

EZARD, T. H. G.; PURVIS, A. Environmental changes define ecological limits to species richness and reveal the mode of macroevolutionary competition. *Ecology Letters*, v. 19, n. 8, p. 899–906, 2016. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12626>.

FELLER, W. *An introduction to probability theory and its applications, Volume 2*. [S.l.]: John Wiley & Sons, 1991.

FISHER, R. A. *The genetical theory of natural selection*. London: Claredon Press, 1930.

FITZPATRICK, B.; FORDYCE, J.; GAVRILETS, S. Pattern, process and geographic modes of speciation. *Journal of evolutionary biology*, Wiley Online Library, v. 22, n. 11, p. 2342–2347, 2009.

FLATHER, C. H.; HAYWARD, G. D.; BEISSINGER, S. R.; STEPHENS, P. A. Minimum viable populations: is there a 'magic number' for conservation practitioners? *Trends in ecology & evolution*, Elsevier, v. 26, n. 6, p. 307–316, 2011.

FRAGATA, I.; BLANCKAERT, A.; LOURO, M. A. D.; LIBERLES, D. A.; BANK, C. Evolution in the light of fitness landscape theory. *Trends in ecology & evolution*, Elsevier, v. 34, n. 1, p. 69–82, 2019.

FREITAS, O.; ARAUJO, S. B.; CAMPOS, P. R. Speciation in a metapopulation model upon environmental changes. *Ecological Modelling*, Elsevier, v. 468, p. 109958, 2022.

FREITAS, O.; CAMPOS, P. R. The role of epistasis in evolutionary rescue. *The European Physical Journal E*, Springer, v. 47, n. 7, p. 49, 2024.

FREITAS, O.; CAMPOS, P. R. Understanding evolutionary rescue and parallelism in response to environmental stress. *Evolution*, Oxford University Press US, v. 78, n. 8, p. 1453–1463, 2024.

FREITAS, O.; CAMPOS, P. R.; ARAUJO, S. B. Patch biogeography under intermittent barriers: macroevolutionary consequences of microevolutionary processes. *Journal of Evolutionary Biology*, Oxford University Press US, v. 37, n. 12, p. 1488–1498, 2024.

FREITAS, O.; WAHL, L. M.; CAMPOS, P. R. Robustness and predictability of evolution in bottlenecked populations. *Physical Review E*, APS, v. 103, n. 4, p. 042415, 2021.

FRY, J. D. Laboratory experiments on speciation. *Experimental evolution: Concepts, methods, and applications of selection experiments*, University of California Press, p. 631–656, 2009.

GASCUEL, F.; FERRIÈRE, R.; AGUILÉE, R.; LAMBERT, A. How Ecology and Landscape Dynamics Shape Phylogenetic Trees. *Systematic Biology*, v. 64, n. 4, p. 590–607, 2015. ISSN 1076836X.

GAVRILETS, S. Evolution and speciation on holey adaptive landscapes. *Trends in ecology & evolution*, Elsevier, v. 12, n. 8, p. 307–312, 1997.

GAVRILETS, S. Speciation in metapopulations. In: *Ecology, genetics and evolution of metapopulations*. [S.l.]: Elsevier, 2004. p. 275–303.

GAVRILETS, S. Models of speciation: Where are we now? *Journal of Heredity*, v. 105, n. S1, p. 743–755, 2014. ISSN 14657333.

GERRISH, P. J.; LENSKI, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica*, Springer, v. 102, p. 127–144, 1998.

GERSTEIN, A. C.; LO, D. S.; OTTO, S. P. Parallel Genetic Changes and Nonparallel Gene–Environment Interactions Characterize the Evolution of Drug Resistance in Yeast. *Genetics*, v. 192, n. 1, p. 241–252, 2012. ISSN 1943-2631. Available at: <https://doi.org/10.1534/genetics.112.142620>.

GILLESPIE, J. *Population Genetics: A Concise Guide*. Johns Hopkins University Press, 1998. (A Johns Hopkins paperback : Science). ISBN 9780801857553. Available at: <https://books.google.com.br/books?id=zb1qAAAAMAAJ>.

GOMULKIEWICZ, R.; HOLT, R. D. When does evolution by natural selection prevent extinction? *Evolution*, JSTOR, v. 49, n. 1, p. 201–207, 1995.

GORDO, I.; GOMES, M. G. M.; REIS, D. G.; CAMPOS, P. R. Genetic diversity in the sir model of pathogen evolution. *PloS one*, Public Library of Science San Francisco, USA, v. 4, n. 3, p. e4876, 2009.

HACKL, J.; DUBERNET, T. Epidemic spreading in urban areas using agent-based transportation models. *Future internet*, MDPI, v. 11, n. 4, p. 92, 2019.

HANSKI, I. Metapopulation dynamics. *Nature*, Nature Publishing Group, v. 396, n. 6706, p. 41–49, 1998.

HARMAND, N.; GALLET, R.; JABBOUR-ZAHAB, R.; MARTIN, G.; LENORMAND, T. Fisher's geometrical model and the mutational patterns of antibiotic resistance across dose gradients. *Evolution*, Blackwell Publishing Inc Malden, USA, v. 71, n. 1, p. 23–37, 2017.

HARTL, D. L.; TAUBES, C. H. Compensatory nearly neutral mutations: selection without adaptation. *Journal of Theoretical Biology*, Elsevier, v. 182, n. 3, p. 303–309, 1996.

HASSELL, M. Density-dependence in single-species populations. *The Journal of animal ecology*, JSTOR, p. 283–295, 1975.

HE, Z.; LI, X.; YANG, M.; WANG, X.; ZHONG, C.; DUKE, N. C.; WU, C. I.; SHI, S. Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. *National Science Review*, v. 6, n. 2, p. 275–288, 2019. ISSN 2053714X.

HENDRY, A. P.; NOSIL, P.; RIESEBERG, L. H. The speed of ecological speciation. *Functional ecology*, Wiley Online Library, v. 21, n. 3, p. 455–464, 2007.

HENRIQUES-SILVA, R.; BOIVIN, F.; CALCAGNO, V.; URBAN, M. C.; PERES-NETO, P. R. On the evolution of dispersal via heterogeneity in spatial connectivity. *Proceedings of the Royal Society B: Biological Sciences*, The Royal Society, v. 282, n. 1803, p. 20142879, 2015.

HIGGS, P. G.; DERRIDA, B. Stochastic models for species formation in evolving populations. *Journal of Physics A: Mathematical and General*, IOP Publishing, v. 24, n. 17, p. L985, 1991.

HOELZER, G. A.; MEINICK, D. J. Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology Evolution*, v. 9, n. 3, p. 104–107, 1994. ISSN 0169-5347. Available at: <https://www.sciencedirect.com/science/article/pii/0169534794902070>.

HUANG, C.-J.; LU, M.-Y.; CHANG, Y.-W.; LI, W.-H. Experimental Evolution of Yeast for High-Temperature Tolerance. *Molecular Biology and Evolution*, v. 35, n. 8, p. 1823–1839, 2018. ISSN 0737-4038. Available at: <https://doi.org/10.1093/molbev/msy077>.

HUGHES, C.; EASTWOOD, R. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the andes. *Proceedings of the National Academy of Sciences*, v. 103, n. 27, p. 10334–10339, 2006. Available at: <https://www.pnas.org/doi/abs/10.1073/pnas.0601928103>.

HUNTLEY, B.; CRAMER, W.; MORGAN, A. V.; PRENTICE, H. C.; ALLEN, J. R. Predicting the response of terrestrial biota to future environmental changes. In: *Past and Future Rapid Environmental Changes*. [S.l.]: Springer, 1991. p. 487–504.

HWANG, S.; PARK, S.-C.; KRUG, J. Genotypic complexity of fisher's geometric model. *Genetics*, Oxford University Press, v. 206, n. 2, p. 1049–1079, 2017.

IWASA, Y.; MICHOR, F.; NOWAK, M. A. Evolutionary dynamics of invasion and escape. *Journal of Theoretical Biology*, v. 226, n. 2, p. 205–214, 2004. ISSN 0022-5193. Available at: <https://www.sciencedirect.com/science/article/pii/S0022519303003333>.

KAUFFMAN, S.; LEVIN, S. Towards a general theory of adaptive walks on rugged landscapes. *Journal of theoretical Biology*, Elsevier, v. 128, n. 1, p. 11–45, 1987.

KAUFFMAN, S. A.; WEINBERGER, E. D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, v. 141, n. 2, p. 211–245, 1989. ISSN 10958541.

KENDALL, D. G. On the Generalized "Birth-and-Death" Process. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 19, n. 1, p. 1 – 15, 1948. Available at: <https://doi.org/10.1214/aoms/1177730285>.

KINGMAN, J. On the properties of bilinear models for the balance between genetic mutation and selection. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society*. [S.l.], 1977. v. 81, n. 3, p. 443–453.

KLUG, A.; PARK, S.-C.; KRUG, J. Recombination and mutational robustness in neutral fitness landscapes. *PLoS computational biology*, Public Library of Science San Francisco, CA USA, v. 15, n. 8, p. e1006884, 2019.

KREINER, J. M.; STINCHCOMBE, J. R.; WRIGHT, S. I. Population genomics of herbicide resistance: Adaptation via evolutionary rescue. *Annual Review of Plant Biology*, Annual Reviews, v. 69, n. Volume 69, 2018, p. 611–635, 2018. ISSN 1545-2123. Available at: <https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-042817-040038>.

LALAND, K. N.; ULLER, T.; FELDMAN, M. W.; STERELNY, K.; MÜLLER, G. B.; MOCZEK, A.; JABLONKA, E.; ODLING-SMEE, J. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the royal society B: biological sciences*, The Royal Society, v. 282, n. 1813, p. 20151019, 2015.

LARSSON, D.; FLACH, C.-F. Antibiotic resistance in the environment. *Nature Reviews Microbiology*, Nature Publishing Group, v. 20, n. 5, p. 257–269, 2022.

LEMANT, J.; SUEUR, C. L.; MANOJLOVIć, V.; NOBLE, R. Robust, Universal Tree Balance Indices. *Systematic Biology*, v. 71, n. 5, p. 1210–1224, 04 2022. ISSN 1063-5157. Available at: <https://doi.org/10.1093/sysbio/syac027>.

LENSKI, R. E. Convergence and divergence in a long-term experiment with bacteria. *The American Naturalist*, v. 190, n. S1, p. S57–S68, 2017. PMID: 28731830.

LEVINS, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*, Oxford University Press, v. 15, n. 3, p. 237–240, 1969.

LIN, W.-H.; KUSSELL, E. Complex interplay of physiology and selection in the emergence of antibiotic resistance. *Current Biology*, Elsevier, v. 26, n. 11, p. 1486–1493, 2016.

LOBKOVSKY, A. E.; KOONIN, E. V. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers in genetics*, Frontiers Media SA, v. 3, p. 246, 2012.

LOBKOVSKY, A. E.; WOLF, Y. I.; KOONIN, E. V. Predictability of evolutionary trajectories in fitness landscapes. *PLoS computational biology*, Public Library of Science San Francisco, USA, v. 7, n. 12, p. e1002302, 2011.

LORENZEN, E. D.; NOGUÉS-BRAVO, D.; ORLANDO, L.; WEINSTOCK, J.; BINLADEN, J.; MARSKE, K. A.; UGAN, A.; BORREGAARD, M. K.; GILBERT, M. T. P.; NIELSEN, R. et al. Species-specific responses of late quaternary megafauna to climate and humans. *Nature*, Nature Publishing Group, v. 479, n. 7373, p. 359–364, 2011.

LYNCH, M.; ACKERMAN, M. S.; GOUT, J.-F.; LONG, H.; SUNG, W.; THOMAS, W. K.; FOSTER, P. L. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, Nature Publishing Group, v. 17, n. 11, p. 704–714, 2016.

MACARTHUR, R. H.; WILSON, E. O. *The Theory of Island Biogeography*. Rev - revised. Princeton University Press, 1967. ISBN 9780691088365. Available at: <http://www.jstor.org/stable/j.ctt19cc1t2>.

MACKEN, C. A.; PERELSON, A. S. Protein evolution on rugged landscapes. *Proceedings of the National Academy of Sciences*, v. 86, n. 16, p. 6191–6195, 1989. ISSN 0027-8424.

MACPHERSON, A.; NUISMER, S. L. The probability of parallel genetic evolution from standing genetic variation. *Journal of Evolutionary Biology*, v. 30, n. 2, p. 326–337, 02 2017. ISSN 1010-061X. Available at: <https://doi.org/10.1111/jeb.13006>.

MANHART, M.; MOROZOV, A. V. Statistical physics of evolutionary trajectories on fitness landscapes. In: *First-Passage Phenomena and Their Applications*. [S.l.]: World Scientific, 2014. p. 416–446.

MANZO, F.; PELITI, L. Geographic speciation in the derrida-higgs model of species formation. *Journal of Physics A: Mathematical and General*, v. 27, n. 21, p. 7079, nov 1994. Available at: <https://dx.doi.org/10.1088/0305-4470/27/21/022>.

MARQUITTI, F. M. D.; FERNANDES, L. D.; AGUIAR, M. A. M. de. Allopatry increases the balance of phylogenetic trees during radiation under neutral speciation. *Ecography*, v. 43, n. 10, p. 1487–1498, 2020. ISSN 16000587.

MARREC, L.; BITBOL, A.-F. Adapt or perish: Evolutionary rescue in a gradually deteriorating environment. *Genetics*, Oxford University Press, v. 216, n. 2, p. 573–583, 2020.

MARTIN, G.; ELENA, S. F.; LENORMAND, T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature genetics*, Nature Publishing Group, v. 39, n. 4, p. 555–560, 2007.

MATUSZEWSKI, S.; HERMISSON, J.; KOPP, M. Fisher's geometric model with a moving optimum. *Evolution*, Wiley Online Library, v. 68, n. 9, p. 2571–2588, 2014.

MAYR, E. *Animal species and evolution*. [S.l.]: Harvard University Press, 1963.

MCDONOUGH, Y.; CONNALLON, T. Effects of population size change on the genetics of adaptation following an abrupt change in environment. *Evolution*, Oxford University Press US, p. qpad103, 2023.

MCPEEK, M. The ecological dynamics of clade diversification and community assembly. *The American Naturalist*, v. 172, n. 6, p. E270–E284, 2008. PMID: 18851684. Available at: <https://doi.org/10.1086/593137>.

MOOERS, A. O.; HEARD, S. B. Inferring evolutionary process from phylogenetic tree shape. *The Quarterly Review of Biology*, University of Chicago Press, v. 72, n. 1, p. 31–54, 1997. ISSN 00335770, 15397718. Available at: <http://www.jstor.org/stable/3036810>.

MORLON, H. Phylogenetic approaches for studying diversification. *Ecology Letters*, v. 17, n. 4, p. 508–525, 2014. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12251>.

NEE, S.; MOOERS, A. O.; HARVEY, P. H. Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences*, v. 89, n. 17, p. 8322–8326, 1992. Available at: <https://www.pnas.org/doi/abs/10.1073/pnas.89.17.8322>.

NOSIL, P.; HARMON, L. J.; SEEHAUSEN, O. Ecological explanations for (incomplete) speciation. *Trends in ecology & evolution*, Elsevier, v. 24, n. 3, p. 145–156, 2009.

NOWAK, M. A. *Evolutionary dynamics: exploring the equations of life*. [S.l.]: Harvard university press, 2006.

OLIVEIRA, V. M. de; MENDES, B. B.; ROQUE, M.; CAMPOS, P. R. Extinction-colonization dynamics upon a survival-dispersal trade-off. *Ecological Complexity*, Elsevier, v. 43, p. 100856, 2020.

ORR, H. A. Adaptation and the cost of complexity. *Evolution*, Wiley Online Library, v. 54, n. 1, p. 13–20, 2000.

ORR, H. A. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, Nature Publishing Group, v. 6, n. 2, p. 119–127, 2005.

ORR, H. A. The probability of parallel evolution. *Evolution*, v. 59, n. 1, p. 216–220, 2005.

ORR, H. A.; UNCKLESS, R. L. Population extinction and the genetics of adaptation. *The American Naturalist*, The University of Chicago Press, v. 172, n. 2, p. 160–169, 2008.

ORR, H. A.; UNCKLESS, R. L. The population genetics of evolutionary rescue. *PLoS genetics*, Public Library of Science San Francisco, USA, v. 10, n. 8, p. e1004551, 2014.

OSGOOD, G. J.; WHITE, E. R.; BAUM, J. K. Effects of climate-change-driven gradual and acute temperature changes on shark and ray species. *Journal of Animal Ecology*, Wiley Online Library, 2021.

OSMOND, M. M.; MAZANCOURT, C. de. How competition affects evolutionary rescue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 368, n. 1610, p. 20120085, 2013. Available at: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2012.0085>.

OSMOND, M. M.; OTTO, S. P.; MARTIN, G. Genetic paths to evolutionary rescue and the distribution of fitness effects along them. *Genetics*, Oxford University Press, v. 214, n. 2, p. 493–510, 2020.

OVASKAINEN, O.; HANSKI, I. Extinction threshold in metapopulation models. In: JSTOR. *Annales Zoologici Fennici*. [S.l.], 2003. p. 81–97.

PAN, T.; WANG, H.; OROZCOTERWENGEL, P.; HU, C. C.; WU, G. Y.; QIAN, L. F.; SUN, Z. L.; SHI, W. B.; YAN, P.; WU, X. B.; ZHANG, B. W. Long-term sky islands generate highly divergent lineages of a narrowly distributed stream salamander (Pachyhynobius shangchengensis) in mid-latitude mountains of East Asia. *BMC Evolutionary Biology*, BMC Evolutionary Biology, v. 19, n. 1, p. 1–15, 2019. ISSN 14712148.

PARADIS, E.; SCHLIEP, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, v. 35, p. 526–528, 2019.

PATWA, Z.; WAHL, L. The fixation probability of beneficial mutations. *Journal of The Royal Society Interface*, v. 5, n. 28, p. 1279–1289, 2008. ISSN 1742-5689.

PAULS, S. U.; NOWAK, C.; BÁLINT, M.; PFENNINGER, M. The impact of global climate change on genetic diversity within populations and species. *Molecular ecology*, Wiley Online Library, v. 22, n. 4, p. 925–946, 2013.

PERES, E. A.; ROCHA, R. Pinto-da; LOHMANN, L. G.; MICHELANGELI, F. A.; MIYAKI, C. Y.; CARNAVAL, A. C. Patterns of species and lineage diversity in the atlantic rainforest of brazil. In: _____. *Neotropical Diversification: Patterns and Processes*. Cham: Springer International Publishing, 2020. p. 415–447. ISBN 978-3-030-31167-4. Available at: <https://doi.org/10.1007/978-3-030-31167-4_16>.

PIGOT, A. L.; PHILLIMORE, A. B.; OWENS, I. P.; ORME, C. D. L. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Systematic Biology*, v. 59, n. 6, p. 660–673, 2010. ISSN 10635157.

Posit team. *RStudio: Integrated Development Environment for R*. Boston, MA, 2022. Available at: <http://www.posit.co/>.

PRINCEPE, D.; CZARNOBAI, S.; CAETANO, R. A.; MARQUITTI, F. M. D.; AGUIAR, M. A. M. de; ARAUJO, S. B. L. Intermittent migration can induce pulses of speciation in a two-island system. *Evolution*, p. qpad210, 12 2023. ISSN 0014-3820. Available at: <https://doi.org/10.1093/evolut/qpad210>.

PRINCEPE, D.; CZARNOBAI, S.; PRADELLA, T. M.; CAETANO, R. A.; MARQUITTI, F. M. D.; AGUIAR, M. A. M. de; ARAUJO, S. B. L. Diversity patterns and speciation processes in a two-island system with continuous migration. *Evolution*, p. 1–12, 2022. ISSN 0014-3820.

PYBUS, O. G.; HARVEY, P. H. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings: Biological Sciences*, The Royal Society, v. 267, n. 1459, p. 2267–2272, 2000. ISSN 09628452. Available at: <http://www.jstor.org/stable/2665821>.

QVARNSTRÖM, A.; ÅLUND, M.; MCFARLANE, S. E.; SIRKIÄ, P. M. Climate adaptation and speciation: particular focus on reproductive barriers in ficedula flycatchers. *Evolutionary applications*, Wiley Online Library, v. 9, n. 1, p. 119–134, 2016.

RAM, Y.; HADANY, L. The probability of improvement in Fisher's geometric model: A probabilistic approach. *Theoretical population biology*, Elsevier, v. 99, p. 1–6, 2015.

REIA, S. M.; CAMPOS, P. R. Analysis of statistical correlations between properties of adaptive walks in fitness landscapes. *Royal Society open science*, The Royal Society, v. 7, n. 1, p. 192118, 2020.

REICHL, L. E. *A modern course in statistical physics*. [S.l.]: John Wiley & Sons, 2016.

RIDLEY, M. *Evolution*. [S.l.]: Wiley-Blackwell; 3rd ed., 2003.

RUNDLE, H. D.; NOSIL, P. Ecological speciation. *Ecology letters*, Wiley Online Library, v. 8, n. 3, p. 336–352, 2005.

SACKIN, M. J. "Good" and "Bad" Phenograms. *Systematic Biology*, v. 21, n. 2, p. 225–226, 07 1972. ISSN 1063-5157. Available at: <https://doi.org/10.1093/sysbio/21.2.225>.

SCHIFFERS, K.; BOURNE, E. C.; LAVERGNE, S.; THUILLER, W.; TRAVIS, J. M. J. Limited evolutionary rescue of locally adapted populations facing climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 368, n. 1610, p. 20120083, 2013. Available at: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2012.0083>.

SCHLUTER, D.; CLIFFORD, E. A.; NEMETHY, M.; MCKINNON, J. S. Parallel evolution and inheritance of quantitative traits. *The American Naturalist*, v. 163, n. 6, p. 809–822, 2004. PMID: 15266380.

SEARLE, C. L.; CHRISTIE, M. R. Evolutionary rescue in host-pathogen systems. *Evolution*, Blackwell Publishing Inc Malden, USA, v. 75, n. 11, p. 2948–2958, 2021.

SHAO, K. T.; SOKAL, R. R. Tree balance. *Systematic Zoology*, v. 39, n. 3, p. 266–276, 1990. ISSN 1063-5157.

SOUSA, J. A. M. de; ALPEDRINHA, J.; CAMPOS, P. R.; GORDO, I. Competition and fixation of cohorts of adaptive mutations under fisher geometrical model. *PeerJ*, PeerJ Inc., v. 4, p. e2256, 2016.

SOUZA, K. S.; FORTUNATO, D. S.; JARDIM, L.; TERRIBILE, L. C.; LIMA-RIBEIRO, M. S.; MARIANO, C. A.; PINTO-LEDEZMA, J. N.; LOYOLA, R.; DOBROVOLSKI, R.; RANGEL, T. F. et al. Evolutionary rescue and geographic range shifts under climate change for global amphibians. *Frontiers in Ecology and Evolution*, Frontiers, v. 11, p. 1038018, 2023.

SPRATT, R. M.; LISIECKI, L. E. A Late Pleistocene sea level stack. *Clim. Past*, Copernicus Publications, v. 12, n. 4, p. 1079–1092, apr 2016. ISSN 1814-9332.

SRIVASTAVA, M.; PAYNE, J. L. On the incongruence of genotype-phenotype and fitness landscapes. *PLoS Computational Biology*, Public Library of Science San Francisco, CA USA, v. 18, n. 9, p. e1010524, 2022.

STUART, Y. E.; VEEN, T.; WEBER, J. N.; HANSON, D.; RAVINET, M.; LOHMAN, B. K.; THOMPSON, C. J.; TASNEEM, T.; DOGGETT, A.; IZEN, R. et al. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature ecology & evolution*, Nature Publishing Group UK London, v. 1, n. 6, p. 0158, 2017.

SZENDRO, I. G.; FRANKE, J.; VISSER, J. A. G. M. de; KRUG, J. Predictability of evolution depends nonmonotonically on population size. *Proceedings of the National Academy of Sciences*, v. 110, n. 2, p. 571–576, 2013.

TANAKA, M. M.; WAHL, L. M. Surviving environmental change: when increasing population size can increase extinction risk. *Proceedings of the Royal Society B*, The Royal Society, v. 289, n. 1976, p. 20220439, 2022.

TENAILLON, O. The utility of Fisher's geometric model in evolutionary genetics. *Annual review of ecology, evolution, and systematics*, Annual Reviews, v. 45, p. 179–201, 2014.

THOMPSON, K. A.; OSMOND, M. M.; SCHLUTER, D. Parallel genetic evolution and speciation from standing variation. *Evolution Letters*, Oxford University Press, v. 3, n. 2, p. 129–141, 2019.

THOMPSON, K. A.; RIESEBERG, L. H.; SCHLUTER, D. Speciation and the city. *Trends in ecology & evolution*, Elsevier, v. 33, n. 11, p. 815–826, 2018.

UECKER, H.; OTTO, S. P.; HERMISSON, J. Evolutionary rescue in structured populations. *The American Naturalist*, University of Chicago Press Chicago, IL, v. 183, n. 1, p. E17–E35, 2014.

WAHL, L. M.; CAMPOS, P. R. Evolutionary rescue on genotypic fitness landscapes. *Journal of the Royal Society Interface*, Royal Society, v. 103, n. 20, p. 20230424, 2023.

WAXMAN, D. Fisher's geometrical model of evolutionary adaptation—beyond spherical geometry. *Journal of Theoretical Biology*, Elsevier, v. 241, n. 4, p. 887–895, 2006.

WEINREICH, D. M.; WATSON, R. A.; CHAO, L. Perspective: sign epistasis and genetic costraint on evolutionary trajectories. *Evolution*, Wiley Online Library, v. 59, n. 6, p. 1165–1174, 2005.

WICHMAN, H. A.; BADGETT, M. R.; SCOTT, L. A.; BOULIANNE, C. M.; BULL, J. J. Different trajectories of parallel evolution during viral adaptation. *Science*, v. 285, n. 5426, p. 422–424, 1999. ISSN 0036-8075.

WILLI, Y.; HOFFMANN, A. A. Demographic factors and genetic variation influence population persistence under environmental change. *Journal of evolutionary biology*, Blackwell Publishing Ltd Oxford, UK, v. 22, n. 1, p. 124–133, 2009.

WOESE, C. R. Phylogenetic trees: Whither microbiology? *Current Biology*, Elsevier, v. 6, n. 9, p. 1060–1063, 1996.

WOESE, C. R.; KANDLER, O.; WHEELIS, M. L. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences*, v. 87, n. 12, p. 4576–4579, 1990.

YAMAGUCHI, R.; WILEY, B.; OTTO, S. P. The phoenix hypothesis of speciation. *Proceedings of the Royal Society B*, The Royal Society, v. 289, n. 1987, p. 20221186, 2022.

YULE, G. U. A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f.r.s. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, The Royal Society, v. 213, p. 21–87, 1925. ISSN 02643960. Available at: <http://www.jstor.org/stable/92117>.

APPENDIX A – BRANCHING PROCESS

The branching process is a mathematical model of a population in which each individual in generation n produces some random number of individuals in generation $n+1$, accordingly to a probability distribution. The probability of extinction or survival of a lineage can be derived as follows.

Let $p_0, p_1, p_2, ..$ be the probabilities that the lineage will leave $0, 1, 2, ...$ descendants in the next generation, i.e. $\sum_k p_k = 1$. In particular, p_0 is the probability that given lineage will go extinct. Such assumption have a probability generating function:

$$f(x) = p_0 + p_1 x + p_2 x^2 + ... = \sum_{k=0}^{\infty} p_k x^k, \tag{A.1}$$

where the probability of leaving k mutant genes for the next generation is given as the coefficient of x^k. The mean number of descendants, i.e. the mean of this distribution is given by $\frac{df}{dx}\Big|_{x=1} = f'(1)$, since

$$f'(x) = p_1 + 2p_2 x + 3p_3 x^2 + ... = \sum_{k=0}^{\infty} k p_k x^k$$

$$f'(1) = \sum_{k=0}^{\infty} k p_k =< k >= \mu.$$

Assuming that the subsequent offspring distribution is independent of the previous generations, in the next generation

$$f\Big(f(x)\Big) = \sum_{k=0}^{\infty} p_k (f(x))^k \tag{A.2}$$

and analogously, after n generations $(n = 0, 1, 2, ...)$,

$$\underbrace{f(f(f(\ldots f}_{n\ \ times}(x)\ldots)) = f_{n-1}(f(x)) = f(f_{n-1}(x)) = f_n(x), \tag{A.3}$$

which one denotes by $f_n(x)$, with $f_1(x) = f(x)$ (CROW; KIMURA, 1970). Therefore, the statistics of the *n*th generation is the compound of its generating functions. Hence, by the chain rule, the average number of descendants at the nth generation is:

$$\begin{aligned}\mu_n &= f_n'(1)\\ &= f_{n-1}'\Big(f(1)\Big)\, f'(1)\\ &= f_{n-1}'(1)\, f'(1)\\ &= f_{n-1}'(1)\, \mu\\ &= f_{n-2}'(1)\, \mu^2 = \dots\\ &= \mu^n.\end{aligned}$$

So if each individual is expected to have more than one offspring, then the population will increase. If each individual is expected to have either one or no offspring, then the population will remain constant or decrease until eventually die out. On average, if the lineage leaves $\mu < 1$ descendants per generation it is bound to go extinct as n increases, in contrast to the case $\mu > 1$.

$$\begin{cases}\mu < 1 & \mu_n \to 0,\\ \mu = 1 & \mu_n \to 1,\\ \mu > 1 & \mu_n \to N.\end{cases}$$

Interestingly, for this stochastic model, the lineage might have an appreciable probability of extinction, despite its replication rate. We refine these qualitative results by investigating the probability of the lineage being extinct *by* generation n (stress on the italic). Let

$$\begin{aligned}\theta_n &= \text{Prob (nth generation has no individuals)}\\ &= \text{Prob (extinction occurs } \textit{by} \text{ nth generation)}\\ &= f_n(0)\\ &= f_{n-1}(f(0))\\ &= f(f...(0)...))\\ &= f(f_{n-1}(0))\\ \theta_n &= f(\theta_{n-1}).\end{aligned} \tag{A.4}$$

Since such an event could happen in any previous generation, we add the probabilities:

$$\text{Prob (extinct } \textit{by} \text{ nth generation)} = \text{Prob (extinct } \textit{by} \text{ (n - 1)th)} + \text{Prob (extinct } \textit{at} \text{ nth).}$$

$$\text{or } \theta_n = \theta_{n-1} + \text{Prob (extinct } \textit{at} \text{ nth)}$$

$$\Rightarrow\ \theta_n \geq \theta_{n-1},\ \forall \text{ n.}$$

Assuming the non-trivial case where $\theta_0 = 0$, we have that

$$0 = \theta_0 \leq \theta_1 \leq \theta_2 \leq \theta_3 \leq ... \leq 1, \tag{A.5}$$

thus θ_n is a non-decreasing sequence bounded by 1 (its a probability), hence, there exists a value θ such that θ_n converges to θ, as $n \to \infty$. We call θ the probability of ultimate extinction, and it is the solution of Equation A.4 when taking this limit:

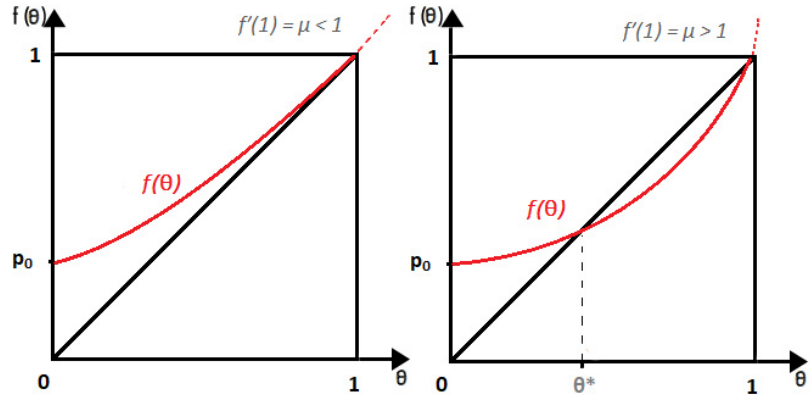
$$\begin{aligned} \lim_{n\to\infty} \theta_n &= \lim_{n\to\infty} f(\theta_{n-1}) \\ \theta &= f(\theta), \end{aligned} \tag{A.6}$$

with $\theta \in [0, 1]$. Remember that $f(\theta)$ is just the probability generating function in Equation A.1,

$$f(\theta) = \sum_{k=0}^{\infty} p_k \theta^k. \tag{A.7}$$

It is easy to see that $f(1) = 1$ is a trivial solution. To investigate the existence of other roots, i.e., if there exists other points in which the $f(\theta)$ curve intersects Equation A.6, we note that its derivative $f'(\theta)$ is strictly increasing for $0 < \theta < 1$ and $f''(\theta)$ is convex on that same interval; as shown in Figure 49. Thus, for $f(0) = p_0 > 0$, the existence of another solution depends on the slope of the curve at $\theta = 1$, which happens to be $f'(1) = \mu$, the average number of descendants.

Figure 49 – Depending on the slope at $\theta = 1$, the PGF $f(\theta)$ may have one additional solution at $\theta*$.



In the line of our previous discussion, we see that ultimate extinction is inevitable when the mean number of offspring is $\mu \leq 1$. On the other hand, despite the average increase in individuals' frequency with $\mu > 1$, there is still a probability $0 < \theta^* < 1$ of such lineage

disappears over time. This result evinces the role of random genetic drift[1], a stochastic effect inherent to the sampling process. It is known that when there are few copies of a lineage, the effect of genetic drift is larger (GILLESPIE, 1998). We can calculate the variance of the process to investigate this observation mathematically. Still, sadly, the branching process variance is rather cumbersome and makes no explicit reference to the size N of the population.

Poisson-distributed offspring

We now assume a particular form of distribution p_k. Let the number of offspring per birth event be drawn from a Poisson distribution with mean λ. Thus, if $X \sim Poisson(\lambda)$, so $Pr(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}$ with offspring number $x = 0, 1, 2...$. Therefore its generating function is

$$\begin{aligned} G_X(k) &= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} k^x \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda k)^x}{x!} k^x \\ &= e^{-\lambda} e^{\lambda k} \\ &= e^{\lambda(k-1)}. \end{aligned} \tag{A.8}$$

The probability of extinction of a branching process starting with a single individual is the smallest root of the equation $G_X(k) = k$ for $k \in (0, 1)$. Which admits a fixed point solution given by $k = e^{\lambda(k-1)}$. Rewriting $k = 1 - \pi$, where π is the probability that such mutation is fixed or established (i.e. does not go extinct), one have $G_X(1 - \pi) = e^{\lambda(1-\pi-1)}$. Solving the fixed point for π:

$$\pi = 1 - e^{-\lambda\pi}. \tag{A.9}$$

In our general case, we use a phenotypic FGM (i.e. infinite alleles or genoma) to establish the fitness of a wild-type population at a distance $\vec{d}$ from the optimum phenotype. Given a mutation displacement $\vec{u}$, the mean offspring number is

[1] The "genetic" part is referent to our particular problem.

$$\begin{aligned}\lambda &= W_{max} \exp\left[-\frac{1}{2}(\vec{d}-\vec{u})^2\right] \\ &= W_{max} \exp\left[-\frac{1}{2}(d^2-u^2+2\vec{d}.\vec{u})\right] \\ &= W_{max} \exp\left[-\frac{d^2}{2}-\frac{u^2}{2}+du\cos\theta\right] \\ &= W(u,\theta),\end{aligned} \tag{A.10}$$

where θ is the angle between vector $\vec{u}$ and $\vec{d}$, and W_{max} is the maximum growth rate. Substituting in the fixed point for $\pi(u,\theta)$, we finally obtain:

$$\pi(u,\theta) = 1 - e^{-W(u,\theta)\pi(u,\theta)}. \tag{A.11}$$

APPENDIX B – DERRIDA-HIGGS DYNAMICS

In the Derrida-Higgs model of speciation, the similarity between two individuals α and β is given by

$$q^{\alpha\beta} = \frac{1}{G}\sum_{i=1}^{G} S_i^{\alpha}\, S_i^{\beta}, \tag{B.1}$$

with genome size G and alleles S_i at each locus i. We can then calculate the similarity between two individuals at generation $t+1$, in terms of their parents at generation t. Let $P_1(\alpha)$ and $P_2(\alpha)$ be the parents of α and $P_1(\beta)$ and $P_2(\beta)$ the parents of β. On average, α gets the allele S_i from one of its parent with probability

$$E[S_i^{\alpha}] = \frac{e^{-2\mu}}{2}\,(S_i^{P_1(\alpha)} + S_i^{P_2(\alpha)}). \tag{B.2}$$

The same calculation can be performed for β, and thus the average similarity is given by

$$\begin{aligned} E[q^{\alpha\beta}] &= \frac{e^{-4\mu}}{4G}\sum_{i=1}^{G}(S_i^{P_1(\alpha)} + S_i^{P_2(\alpha)})(S_i^{P_1(\beta)} + S_i^{P_2(\beta)}) \\ &= \frac{e^{-4\mu}}{4}\left(q^{P_1(\alpha)P_1(\beta)} + q^{P_1(\alpha)P_2(\beta)} + q^{P_2(\alpha)P_1(\beta)} + q^{P_2(\alpha)P_2(\beta)}\right) \end{aligned} \tag{B.3}$$

which is exact for a very large genome $G \to \infty$, thus the expected values coincide with the real similarity $E[q^{\alpha\beta}] = q^{\alpha\beta}$. When the parents of an individual are equal, the similarity between them equals 1, and among the other potential parents is approximately q, the average similarity of the distribution. Following Eq. B.3, the average similarity between pairs in the next generations is:

$$q_{t+1} = \frac{e^{-4\mu}}{4}(1 + 3q_t). \tag{B.4}$$

On the other hand, when all progenitors are different, their average similarity will be q, and so

$$q_{t+1} = q_t\, e^{-4\mu}. \tag{B.5}$$

The probability of these events are simpley $4/M$ and $1-4/M$, respectively, thus, the average similarity among the descendants of the next generation is

$$q_{t+1} = \frac{4}{M}\frac{e^{-4\mu}}{4}(1 + 3q_t) + \left(1 - \frac{4}{M}\right)e^{-4\mu}q_t = \left[q_t\left(1 - \frac{1}{M}\right) + \frac{1}{M}\right]e^{-4\mu}. \tag{B.6}$$

For small mutation rate and large population size, an explicit expression can be found for the equilibrium similarity q_0, where $q' = q$:

$$q_0 = \frac{1}{1 + 4\mu M}. \tag{B.7}$$

However, if exists a minimum similarity $q_{min} > q_0$ necessary for reproduction to occur between mating pairs, before the similarity distribution reach an equilibrium at q_0, the population will split in groups of species where $q > q_{min}$ within a species, but $q < q_{min}$ among species. Therefore, in the Derrida-Higgs model, the condition $q_{min} > q_0$ is essential to promote speciation.

Two-islands

Now, with a population divided into two patches or islands, we have the similarity among individuals in the same patch q and between patches p. The evolution of q is given by Eq. B.6, while for p, the contribution to offspring similarity coming from parents that have come from a different island is $p_{t+1} = p_t\, e^{-4\mu}$. The dynamics of q and p after migration are, therefore, given by the following equations:

$$\begin{aligned} q_{t+1} &= [a(\varepsilon)\, q_t + b(\varepsilon)\, p_t]\; e^{-4\mu} \\ p_{t+1} &= [b(\varepsilon)\, q_t + a(\varepsilon)\, p_t]\; e^{-4\mu} \end{aligned} \tag{B.8}$$

where $a(\varepsilon) = (1-\varepsilon)^2 + \varepsilon^2$is the probability that both α and β did not migrate or that they both migrated, therefore keeping their original geographic relation, and $b(\varepsilon) = 2\varepsilon(1-\varepsilon)$ is the probability that one of them exchanged places, altering the geographic relative position of the pair. The equilibrium solutions are obtained by setting $q_{t+1} = q_t$ and $p_{t+1} = p_t$. For ε, μ and $1/M$ all much smaller than 1, we obtain

$$\begin{aligned} q_0(\nu, \sigma) &= \frac{\nu + 2\sigma}{2\sigma\,(2\nu+1) + \nu\,(\nu+1)} \\ p_0(\nu, \sigma) &= \frac{2\sigma}{2\sigma\,(2\nu+1) + \nu\,(\nu+1)} \end{aligned} \tag{B.9}$$

where $\nu = 4\mu M$ and $\sigma = M\varepsilon$ is the average number of exchanged migrants at each generation. Note that for $\sigma >> \nu$, we have $p_0(\nu, \sigma) = q_0(2\nu, 0)$, indicating that for a large migration intensity the two islands behave as a single island with twice the population. The expected number of species in each island N can be estimated through Eq. B.9, by replacing $q_0(\nu, \mu) \to q_{min}$ and $M \to m$ - the population size that equilibrates at q_{min}. Solving for m, one found the number of species in each island to be $N = M/m$:

$$N = \frac{\nu(4\sigma + \nu)}{(2\sigma + \nu)\,(q_{min}^{-1} - 1)}. \tag{B.10}$$

The total number of species in the islands, N_T, can also be estimated as

$$N_T = N\left(2 - \frac{p_0(\nu, \sigma)}{q_0(2\nu, 0)}\right). \tag{B.11}$$

When $\sigma = 0$ the islands are independent and $N_T = 2N$. When migration dominates over mutations, $\sigma >> \mu$, $N_T = N$.

APPENDIX C – CENTRAL LIMIT THEOREM

The Central Limit Theorem states that the probability density describing the distribution of outcomes of a large number of events universally approaches a Gaussian distribution. Let us consider a stochastic variable, Y_N, which is the deviation from the average of N statistically independent measurements of a stochastic variable X. We write Y_N as

$$Y_N = \frac{1}{N}(X_1 + ... + X_N) - \bar{x} = Z_1 = .. + Z_N. \tag{C.1}$$

The characteristic function, $f_Z(k;N)$, for the stochastic variable, $Z_i = (1/N)(X_i - \bar{x})$, can be written

$$f_Z(k;N) = \int_{-\infty}^{\infty} dx\, e^{i(k/N)(x-\bar{X})} P_X(x) = 1 - \frac{1}{2}\frac{k^2}{N^2}\sigma_X^2 + ..., \tag{C.2}$$

where $\sigma_X^2 = \overline{x^2} - \bar{x}^2$ is the variance. For large N and finite variance, higher-order terms in the expansion of the right-hand side can be neglected. The characteristic function for Y_N is then

$$f_{Y_N}(k) = \left(1 - \frac{1}{2}\frac{k^2}{N^2}\sigma_X^2\right)^N \to exp\left(-\frac{k^2\sigma_X^2}{2N}\right) \quad as\, N \to \infty. \tag{C.3}$$

In the last equation, we have used the identity $\lim_{N\to\infty}(1 + x/N)^N = e^x$. Thus, for large N,

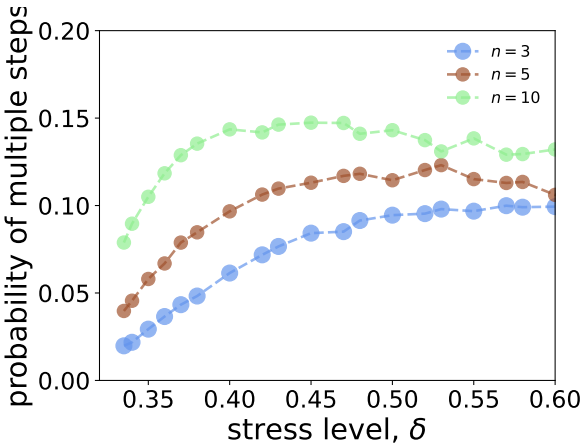
$$P_{Y_N}(y) = \frac{1}{2\pi}\int dk\, e^{iky}\, exp\left(-\frac{k^2\sigma_X^2}{2N}\right) = \sqrt{\frac{N}{2\pi\sigma_X^2}} exp\left(-\frac{Ny^2}{2\sigma_X^2}\right). \tag{C.4}$$

Regardless of the form of $P_X(x)$, if it has finite moments, the average of a large number of statically independent measurements of X will be a Gaussian centered at $\bar{x}$, with standard deviation which is $1/\sqrt{N}$ times the standard deviation of the probability density of X.

APPENDIX D – SUPPLEMENTARY RESULTS

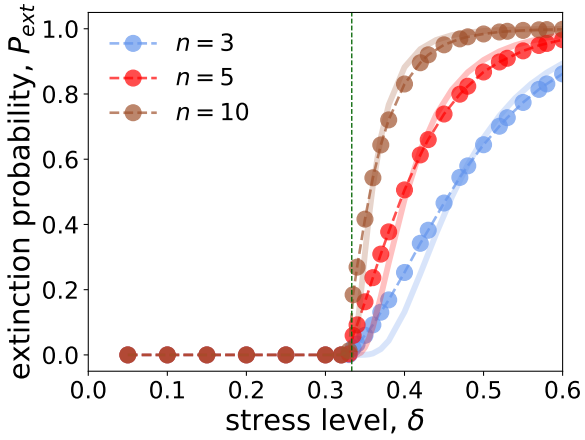
Chapter 2

Figure 50 – Proportion of cases in which evolutionary rescue is attained by multiple mutational steps as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points.



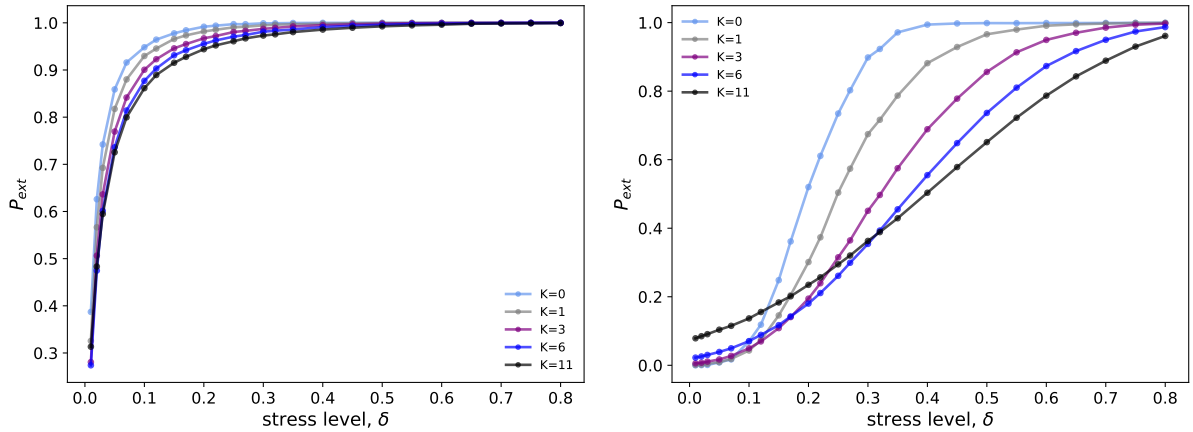
Source: The author (2024)

Figure 51 – Extinction probability as a function of the stress level δ for the genotypic FGM. The number of traits n is indicated in the legend. From the green vertical line onwards, the population is in the regime of rescue/extinction. The plot compares the simulation results (data points) with the analytical predictions (lines), given by Eq.(19). The parameter values are genome size $L = 12$, carrying capacity $K = 10000$, mutation probability $U = 0.005$, and mean value of phenotypic effects $\lambda = 0.4$. Error bars were omitted because their size is of the order or smaller than the symbols marking the data points.



Source: The author (2024)

Figure 52 – Probability of extinction as a function of the stress level for the genotypic model. Curves represent different degrees of epistasis K. Error bars are smaller than points. Parameters for both cases are $L = 12$, $N_0 = 10^5$, $W_{max} = 1.5$, and $b = 6$. Mutation rate are $\mu = 10^{-5}$ for the left panel and $\mu = 10^{-3}$ for the right panel.



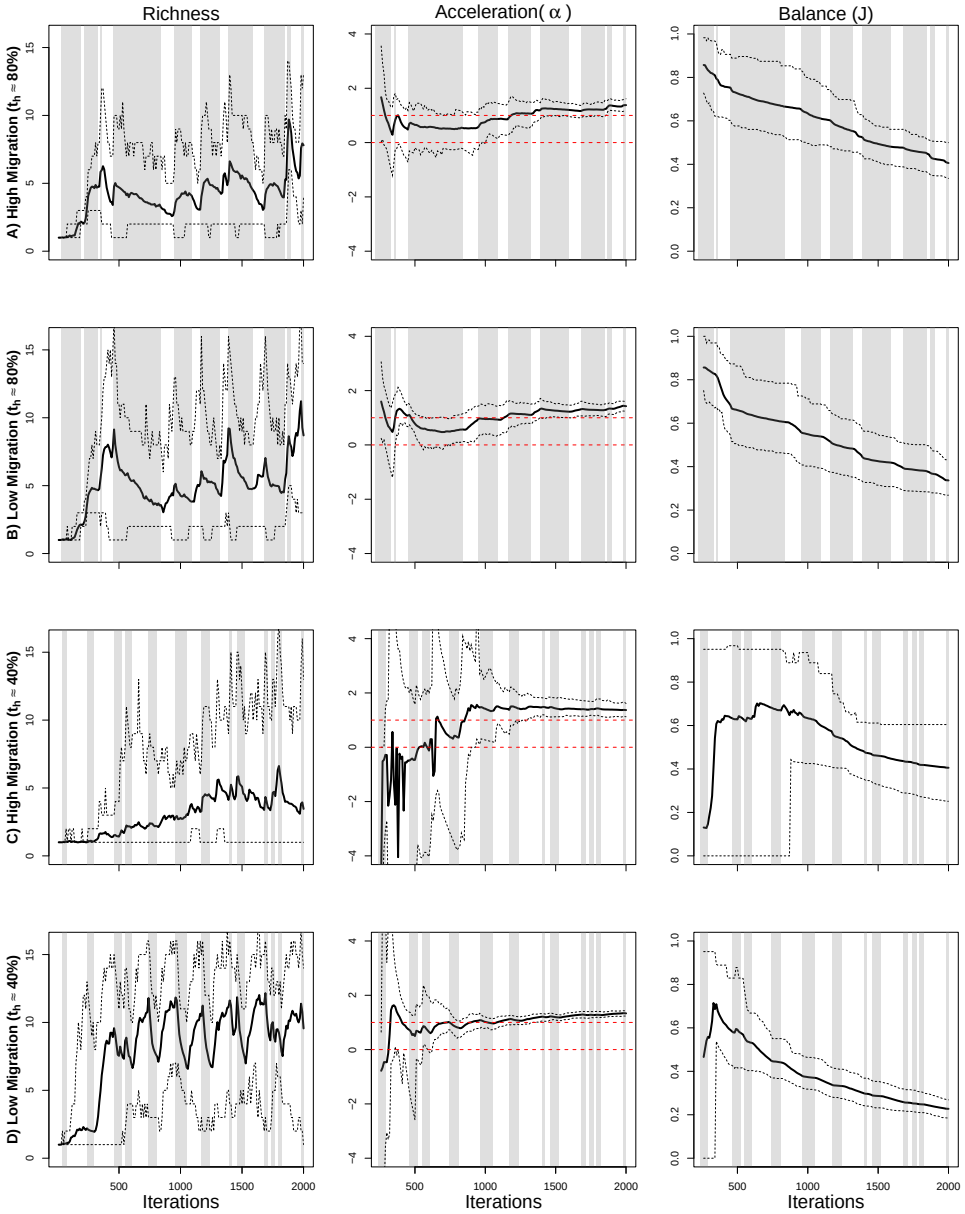
Source: The author (2024)

Chapter 3

Here we report, over time, the richness, phylogeny balance, and speciation acceleration. The two last metrics consider the complete phylogeny. We combine scenarios of high and low migration ($\bar{\varepsilon} = \{0.08, 0.02\}$), for when the population spends more time in isolation ($t_h \approx 80\%$) and in connection ($t_h \approx 40\%$) (Fig. 53).

The richness oscillatory pattern observed in all scenarios (Fig. 53)) reveals that the system is out of equilibrium (as pointed by (PRINCEPE et al., 2023)); however, acceleration of speciation tends to a positive stationary value while the phylogeny balance tends to decrease regardless of the scenarios. In fact, for a long enough time, the phylogeny balance should tend to zero since each extinct branch contributes with the decrease of J. Both α and J metrics depend on all historical events of extinctions and speciations, and they tend to converge as the recent events become only an insignificant fraction of the whole history.

Figure 53 – Evolution of richness, acceleration of speciation and phylogeny balance for combinations of isolation $t_h \approx \{80, 20\}$ and migration rate $\bar{\varepsilon} = \{0.02, 0.08\}$. Solid lines represent the average of 50 replications, while dotted ones delimit the confidence interval of 90%. Following the α definition, the dashed red lines indicate the interval in which the results do not differ from random branching models. Background gray bands indicate when patches are isolated for each t_h.



Source: The author (2024)