

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Programa de Pós-graduação em Física

Alexandre de Aquino Soares

**O PROCESSO ADAPTATIVO E O EFEITO DA
MUTABILIDADE**

Belo Horizonte

2023

Alexandre de Aquino Soares

O PROCESSO ADAPTATIVO E O EFEITO DA MUTABILIDADE

Tese apresentada ao Programa de Pós-graduação em Física do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Ciências.

Orientador: Dr. Ronald Dickman

Coorientador: Dr. Louis Bernard Klaczko e
Dr. Lucas Lages Wardil

Belo Horizonte

2023

Dados Internacionais de Catalogação na Publicação (CIP)

S676p Soares, Alexandre de Aquino.
O processo adaptativo e o efeito da mutabilidade / Alexandre de Aquino
Soares. – 2023.
137 f. : il.

Orientador: Ronald Dickman.
Coorientadores: Louis Bernard Klaczko e Lucas Lages Wardil.
Tese (doutorado) – Universidade Federal de Minas Gerais,
Departamento de Física.
Bibliografia: f. 111-119.

1. Evolução. 2. Adaptação. 3. Mutação. I. Título. II. Klaczko, Louis Bernard.
III. Wardil, Lucas Lages. IV. Universidade Federal de Minas Gerais,
Departamento de Física.

CDU – 573:53 (043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

FOLHA DE APROVAÇÃO

A presente tese, intitulada "**O processo adaptativo e o efeito da mutabilidade**" de autoria de **ALEXANDRE DE AQUINO SOARES** submetida à Comissão Examinadora, abaixo-assinada, foi aprovada para obtenção do grau de **DOUTOR EM CIÊNCIAS**, em vinte e quatro de fevereiro de 2023.

Belo Horizonte, 24 de fevereiro de 2023.

Prof. Ronald Dickman
Orientador do estudante
Departamento de Física/UFMG

Prof. Lucas Lages Wardil
Coorientador do estudante
Departamento de Física/UFMG

Prof. Armando Gil Magalhães Neves
Departamento de Matemática/UFMG

Profa. Flávia Maria Darcie Marquitti
Instituto de Física Gleb Wataghin/UNICAMP

Prof. José Fernando Fontanari
Instituto de Física de São Carlos/USP

Prof. Paulo Roberto de Araujo Campos
Departamento de Física/UFPE



Documento assinado eletronicamente por **Lucas Lages Wardil, Professor do Magistério Superior**, em 28/02/2023, às 11:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Paulo Roberto de Araujo Campos, Usuário Externo**, em 28/02/2023, às 11:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Armando Gil Magalhaes Neves, Professor do Magistério Superior**, em 28/02/2023, às 15:09, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Flávia Maria Darcie Marquitti, Usuária Externa**, em 28/02/2023, às 17:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **José Fernando Fontanari, Usuário Externo**, em 01/03/2023, às 09:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

A todos os meus professores, do beabá ao [Â, Ê].

Agradecimentos

Agradeço principalmente a Deus, que me permitiu esse caminho de muitas aprendizagens e possibilitou a conclusão deste trabalho. Meu muito obrigado aos meus orientadores Dr. Ronald Dickman, Dr. Lucas Wardil e Dr. Louis Bernard Klaczko. Muito obrigado à banca avaliadora, em particular ao professor Dr. Armando Neves por seu exame minucioso e pelas muitas sugestões. As eventuais falhas que permanecem são de minha responsabilidade. Agradeço também aos meus pais pela presença importante durante a pandemia e por todo o apoio durante o doutorado. Agradeço à Sheilla, seu Judson e dona Rosângela pela amizade e pelo apoio ao longo do programa. Agradeço também à ABC² por permitir uma integração importante da minha fé e do meu trabalho científico. Obrigado também ao Departamento de Física por todo o conhecimento proporcionado em forma de disciplinas e seminários, e à CAPES e ao CNPq pelos financiamentos das bolsas e treinamentos.

*“A sorte é lançada no colo,
mas a decisão vem do Senhor.”
(Bíblia Sagrada, Provérbios 16, 33)*

Resumo

Estudamos o processo evolutivo de adaptação em populações assexuadas infinitas e finitas analiticamente e via simulação computacional. Verificamos que, na presença de certas relações entre seleção e taxa de mutação para um conjunto de mutações, a mutabilidade pode alterar tanto a ordem com que as adaptações acontecem como o destino evolutivo (i.e. o conjunto final das adaptações adquiridas) na linhagem vencedora. Estabelecemos os regimes de mutabilidade e seleção em que o tipo menos apto prevalece, em um sistema de um único passo mutacional e duas alternativas. Na presença de mais de duas mutações, a repetibilidade do processo é máxima para mutabilidades intermediárias e redes mutacionais suficientemente pequenas, onde a interferência clonal entre diferentes trajetórias e destinos é maior e poucas trajetórias estão disponíveis. Esses resultados mudam nossa forma de entender o que a mutabilidade faz de fato, e podem ter impacto sobre nossa compreensão atual da evolução da mutabilidade.

Palavras-chave: evolução, adaptação, mutabilidade.

Abstract

Here we study the adaptive process in infinite and finite asexual populations, analytically and via simulation. We verify that, in the presence of certain relationships between selection and mutation rate for a set of mutations, mutability may alter both the order which adaptations happen, and the evolutionary fate (i.e. the final set of acquired adaptations) of the winning lineage. We establish the regimes of mutability and selection in which the least fit mutant prevails, in a system of a sole mutation step and two alternatives. In the presence of more than two mutations, the repeatability of the adaptive process is maximum for intermediate mutabilities and relatively small mutation networks, in which clonal interference between distinct trajectories is greater, and few trajectories are available. These results change our way of understanding what indeed mutability really does, and they may impact our current comprehension of the evolution of mutability.

Keywords: evolution. adaptation. mutability.

Lista de ilustrações

- Figura 1 – Modos de adaptação. Cada cor representa um mutante distinto. A cor branca representa uma população inicial selvagem. Acima: Mutações sequenciais. Cada nova mutação adaptativa se acumula a outras anteriores. Este é o regime do modelo origem-fixação. Abaixo: Interferência clonal. As mutações coexistem e competem entre si por fixação. 33
- Figura 2 – Caminhada evolutiva. Esquerda: as mutações possíveis (setas) entre os genótipos (círculos), e a caminhada evolutiva (setas pretas) realizada de um genótipo inicial (círculo branco) até um genótipo final (círculo preto). Direita: as possíveis caminhadas evolutivas em termos de dinâmica de aptidão (vertical) da população ao longo do tempo (medido em número de eventos adaptativos, da esquerda para a direita), e a caminhada evolutiva realizada (linha preta). 34
- Figura 3 – O modelo simplificado, sem as remoções por mutações deletérias ou letais. A variável n é o número de alelos favoráveis, e r é a taxa de mutações favoráveis por *loco*. De cima para baixo: um exemplo para $G = 8$ locos; as taxas de mutação do genótipo com n alelos favoráveis quando o genótipo é selvagem, $n = 0$; idem, quando o genótipo tem $1 \leq n \leq G - 1$ alelos favoráveis; e quando o genótipo só possui alelos favoráveis, $n = G$ 38
- Figura 4 – Uma via adaptativa. Comparação entre as soluções numérica em tempo discreto (esquerda) e analítica (direita) para as distribuições genotípicas de um único conjunto de genes, com $G = 16$, $s = 0,005$ and $r = 10^{-8}$ em diferentes gerações t . O eixo x corresponde ao número de mutações acumulado. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante. 46
- Figura 5 – Verificação da solução para uma via adaptativa. Comparação entre as soluções numérica em tempo discreto (linha contínua) e analítica (círculo) para as distribuições genotípicas de um único conjunto de genes, com $G = 16$, $s = 0,005$ and $r = 10^{-8}$ em diferentes gerações t . Acima: média da distribuição de alelos adaptativos no tempo; abaixo: desvio padrão da distribuição de alelos adaptativos no tempo. 47

- Figura 6 – Uma via deletéria. Comparação entre a distribuição numérica em tempo discreto (esquerda) e analítica (direita) para tempos longos, de um único conjunto de genes em $t = 10000$, para $G = 8$, $s = -0,0005$ e r indicado na primeira coluna. O eixo x corresponde ao número de mutações acumulado. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante. 47
- Figura 7 – Verificação da solução de quase-espécies para uma via deletéria. Comparação entre as soluções numérica em tempo discreto (linha contínua) e analítica (círculo) para as médias das distribuições genotípicas de um único conjunto de genes, com $s = -0,0005$ com diferentes taxas de mutação r 48
- Figura 8 – Duas vias adaptativas. Comparação entre a dinâmica da distribuição numérica em tempo discreto (esquerda) e analítica (direita) de dois conjuntos de genes, com $s_1 = 0,005$, $r_1 = 10^{-8}$, $G_1 = 2$, $s_2 = 0,00595$, $r_2 = 10^{-9}$ e $G_2 = 4$. O eixo y corresponde ao número de mutações do tipo 1 acumulado, e o eixo x o do tipo 2. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante. 50
- Figura 9 – Número médio de alelos adaptativos com evolução em duas vias adaptativas. Comparação entre a dinâmica da distribuição numérica em tempo discreto (linha contínua) e analítica (círculo) de dois conjuntos de genes, com $s_1 = 0,005$, $r_1 = 10^{-8}$, $G_1 = 2$, $s_2 = 0,00595$, $r_2 = 10^{-9}$ e $G_2 = 4$. Acima: número médio de alelos do tipo 1; abaixo: número médio de alelos do tipo 2. 51
- Figura 10 – Quase-espécies com duas vias deletérias. Comparação entre o estado estacionário da distribuição numérica em tempo discreto (esquerda) e analítica (direita) de dois conjuntos de genes, com $r_1 = r_2 = 10^{-3}$, s_1 mostrado na primeira coluna, $s_2 = s_1 - 0,0009$, $G_1 = 2$, $G_2 = 4$ e tempo $t = 10000$. O eixo y corresponde ao número de mutações do tipo 1 acumulado, e o eixo x o do tipo 2. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante. 52

- Figura 11 – Número médio de alelos com evolução em duas vias deletérias. Comparação entre o número médio de alelos deletérios da distribuição numérica em tempo discreto (linha contínua) e analítica (círculo) de dois conjuntos de genes, variando-se o coeficiente de seleção da primeira via s_1 , enquanto $s_2 = s_1 - 0,0009$. Os demais parâmetros são $r_1 = 10^{-8}$, $G_1 = 2$, $s_2 = 0,00595$, $r_2 = 10^{-9}$ e $G_2 = 4$. Acima: número médio de alelos do tipo 1; abaixo: número médio de alelos do tipo 2. 54
- Figura 12 – Quase-espécies com duas vias opostas. Comparação entre o estado estacionário da distribuição numérica em tempo discreto (esquerda) e analítica (direita) de dois conjuntos de genes, com $r_1 = r_2 = 10^{-3}$, $s_1 = 0,005$ (adaptativo), s_2 (deletério) mostrado à esquerda dos gráficos, $G_1 = 2$, $G_2 = 4$ e tempo $t = 10000$. O eixo y corresponde ao número de mutações do tipo 1 acumulado, e o eixo x o do tipo 2. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante. 55
- Figura 13 – Número médio de alelos de cada tipo no estado estacionário com evolução em uma via adaptativa e outra deletéria. Comparação entre o número médio de alelos deletérios da distribuição numérica em tempo discreto (linha contínua) e analítica (círculo) de dois conjuntos de genes, variando-se o coeficiente de seleção da segunda via s_2 . Os demais parâmetros são $s_1 = 0,005$, $r_1 = 10^{-8}$, $G_1 = 2$, $r_2 = 10^{-9}$ e $G_2 = 4$. Acima: número médio de alelos do tipo 1; abaixo: número médio de alelos do tipo 2. 56
- Figura 14 – Modelo para mutações em um único loco. O tipo selvagem muta para o tipo A , que possui aptidão $w_A = 1 + s_A$ menor que a aptidão $w_B = 1 + s_B$ do mutante alternativo B . No entanto, a taxa de mutação do tipo selvagem para o tipo B é ρ vezes menor que a taxa para o tipo A , com $\rho > 1$. A taxa r é um fator comum a ambas as mutações, e reflete a mutabilidade da população. 60
- Figura 15 – Abundância relativa de cada tipo dentre os mutantes em uma população infinita. Os parâmetros são $r = 1 \times 10^{-12}$, $s_A = 0,01$, $s_B = 0,02$ e $\rho = 10$. 62

- Figura 16 – Dependência da prevalência do tipo A da mutabilidade r , o tamanho da população N , a aptidão (em termos do coeficiente de seleção s_A) e o fluxo gênico f . Os discos e os círculos representam resultados das simulações. Os discos sólidos (\bullet) representam uma aptidão infinita para o tipo B e os círculos vazados (\circ) representam um coeficiente de seleção de B quatro vezes mais alto que o do tipo A . A linha contínua (—) representa a aproximação analítica, Eq. 3.10. No gráfico (d), “Frequência relativa” significa a frequência de A em relação a ambos os mutantes no estado estacionário. Os coeficientes de seleção s_A e s_B são dados por $s_A = w_A - 1$ e $s_B = w_B - 1$. Os parâmetros são (a) $N = 10^6$, $s_A = 0,01$, $\rho = 1000$; (b) $r = 1 \times 10^{-6}$, $s_A = 0,01$, $\rho = 10$; (c) $N = 10^6$, $r = 1 \times 10^{-12}$, $\rho = 10$; e (d) $N = 100$, $\rho = 1,2$, $w_A = 1,01$, $w_B = 1,04$ 63
- Figura 17 – Probabilidade de fixação de cada tipo (A) e o tempo de fixação (B) como função da mutabilidade. A queda abrupta do tempo de fixação do tipo A e seu desvio padrão é devido ao tamanho finito da amostra. Os parâmetros são $N = 10^7$, $s_A = 0,01$, $s_B = 0,056$ e $\rho = 100$ 66
- Figura 18 – Tamanhos efetivos de cada população mutante e coeficientes de seleção efetivos. Os dois gráficos, (A) e (B), devem ser lidos em pares: para cada r , os valores de N_e e s_e correspondem à solução de um sistema de equações composto com o tipo selvagem e N_e mutantes de um único tipo mutante que conduz às mesmas probabilidades de fixação e tempos de fixação. O dado original vem de simulação. Os parâmetros são $N = 10^7$, $s_A = 0,01$, $s_B = 0,056$ e $\rho = 100$ 68
- Figura 19 – Histórias simuladas típicas da composição da população, encerradas na absorção de A ou B . Os gráficos da esquerda (A–D) mostram a evolução no simplex $N_W + N_A + N_B = N$, com os vértices correspondendo a estados de único tipo e os lados a estados com dois tipos. Os gráficos da direita (1–4) mostram a diferença $N_A - N_B$ (independentemente do valor de N_W) ao longo do tempo. Histórias que acabam na absorção de A (B) são mostradas em azul (vermelho). As linhas correspondem às mutabilidades (de cima para baixo) $r = 10^{-11}$, $r = 10^{-7}$, $r = 10^{-3}$ e $r = 10/11$, que é o maior valor possível de r com $\rho = 10$ em que $r_a + r_B = (1 + 1/\rho)r \leq 1$, condição do modelo de Wright-Fisher. Os parâmetros são $N = 10^6$, $s_A = 0,01$, $s_B = 0,02$ e $\rho = 10$ 69

Figura 20 – Limiar para ρ acima do qual o tipo A prevalece no regime de mutações raras da Eq. (3.16) para $N = 10$ (azul), $N = 100$ (amarelo), e $N = 1000$ (verde). O limite $N \rightarrow \infty$, Eq. (3.17) (preto, tracejado) e a aproximação da Eq. (3.18) (vermelho) são também mostrados. O eixo ω corresponde a ω_A/ω_B . Quanto maior ω , maior a vantagem seletiva de B sobre A . O coeficiente $s_A = 0,01$, o que implica $\omega_A \approx 0.98$. Note que a aproximação $N \rightarrow \infty$ tem um bom ajuste para $N = 1000$. Quanto maior a vantagem de B sobre A ou quanto maior o tamanho da população, maior é a razão ρ necessária para A prevalecer.	71
Figura 21 – Sistema com três mutantes adaptativos, com aptidões crescentes e taxas de mutação decrescentes. Os parâmetros são $N = 10^6$, $s_A = 0,01$, $s_B = 0,02$, $s_C = 0,04$, $\rho = 10$ de A para B e de B para C	73
Figura 22 – Há seis trajetórias possíveis entre 000 e 111. O caminho destacado em marrom é bA.	77
Figura 23 – Probabilidade de cada trajetória em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.	80
Figura 24 – Diferença da probabilidade de cada trajetória em função da mutabilidade para a probabilidade esperada no regime SSWM. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.	80
Figura 25 – Média das entropias S por realização, em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico. . . .	81
Figura 26 – Entropia S_T da distribuição média, em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico. . . .	82
Figura 27 – Correlação entre as distribuições das trajetórias indicadas acima dos gráficos e das legendas, em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.	83

- Figura 28 – Distribuições de probabilidade das trajetórias em diferentes experimentos. Da esquerda para a direita, e de cima para baixo: condições padrão; condições padrão exceto por aptidões neutras ($s = 0$); condições padrão com cada mutação contribuindo com o mesmo coeficiente de seleção ($s = 0,01$); as taxas de mutação são iguais e a evolução é neutra ($s = 0$); as taxas de mutação são iguais, mas os coeficientes de seleção dependem do locus, como originalmente; e as taxas de mutação são iguais, e cada mutação contribui com o mesmo coeficiente de seleção ($s = 0,01$). Nas condições padrão, o tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico, exceto para simulações neutras onde foram realizadas 2048 simulações. 84
- Figura 29 – Média das entropias por realização S no experimento padrão e nos controles. A legenda representa: "Standard" são condições padrão; "StandardMutNeutral" corresponde a condições padrão exceto por aptidões neutras ($s = 0$); "StandardMutUniformSelStep" corresponde a condições padrão com cada mutação contribuindo com o mesmo coeficiente de seleção ($s = 0,01$); "UniformMutNeutral" corresponde a taxas de mutação são iguais e a evolução neutra ($s = 0$); no "UniformMutStandardSel" as taxas de mutação são iguais, mas os coeficientes de seleção dependem do locus, como originalmente; e no "UniformMutUniformSelStep" as taxas de mutação são iguais, e cada mutação contribui com o mesmo coeficiente de seleção ($s = 0,01$). No experimento padrão, o tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico, exceto para simulações neutras onde foram realizadas 2048 simulações. 86
- Figura 30 – Entropia da média das distribuições de cada realização em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico. 87
- Figura 31 – Distribuições de probabilidade das trajetórias com diferentes tamanhos de população para redes completas com três locos. De cima para baixo: $N = 10^3$, $N = 10^6$ e $N = 10^9$. A rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico. 88
- Figura 32 – Média das entropias das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico. 89

Figura 33 – Entropias da média das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.	90
Figura 34 – Média das entropias das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 5$ loci, $\sigma = 0,01$ e $\beta = 1$	91
Figura 35 – Entropia da média das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 5$ loci, $\sigma = 0,01$ e $\beta = 1$	91
Figura 36 – Média da entropia das distribuições de cada realização S em função da mutabilidade para diferentes coeficientes de seleção médios. A rede tem $G = 5$ loci, $\beta = 1$ e o tamanho da população é $N = 10^6$	92
Figura 37 – Média da entropia das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de rede G . O tamanho da população é $N = 10^6$, $\beta = 1$ e $\sigma = 0,01$	93
Figura 38 – Entropia da média das distribuições de cada realização S_T em função da mutabilidade para diferentes tamanhos de rede. O tamanho da população é $N = 10^6$, $\beta = 1$ e $\sigma = 0,01$	94
Figura 39 – Rede completa a partir da qual serão construídas as redes com epistasia aleatória. As setas correspondem a transições entre os nós. Quanto mais escuras e mais largas as setas, maior é a taxa de transição entre os respectivos nós. Quanto mais escuros os nós, maior é a aptidão. Note uma relação inversa entre a tonalidade da seta que chega ao nó e a do próprio nó.	97
Figura 40 – Exemplos de redes com epistasia aleatória. Todos os nós da rede completa são mostrados como referência. Os nós que pertencem à rede estão ligados por setas a partir do nó raiz. Quanto mais escuras as setas, maior é a taxa de transição entre os respectivos nós. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$	100
Figura 41 – Representação da frequência de nós (acima) e folhas (abaixo) da rede com epistasia aleatória. Quanto maior o tamanho do nó, maior é a sua frequência relativa. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$	102
Figura 42 – Histograma da probabilidade de um nó ser de um dado genótipo. Os genótipos são representados por números binários que codificam a representação das mutações em cada sítio (veja Métodos), em ordem crescente de aptidão aditiva. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$	103

- Figura 43 – Histogramas da probabilidade de uma folha ser de um dado genótipo. Os genótipos são representados por números binários que codificam a representação das mutações em cada sítio (veja Métodos), em ordem crescente de aptidão aditiva. . A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$ 104
- Figura 44 – Histograma da probabilidade de uma folha ter n mutações. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$ 104
- Figura 45 – Frequências dos genótipos de acordo com o número de mutações que carrega (acima), e número médio de mutações por genoma, como função da mutabilidade. O tamanho da população é $N = 10^6$ e o coeficiente de seleção unitário é $s = 0,01$. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$. Para cada mutabilidade foram construídas 16384 novas redes, com uma simulação por rede. A barra de erro é mostrada em ambos os gráficos, mas não é grande o suficiente no segundo gráfico. 106
- Figura 46 – Média das entropias de cada realização, e entropia da média das distribuições de cada realização, como função da mutabilidade. O tamanho da população é $N = 10^6$ e o coeficiente de seleção unitário é $s = 0,01$. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$ 107

Lista de tabelas

Tabela 1 – Parâmetros evolutivos. Os valores correspondem a medidas experimentais para <i>E. coli</i> (TADDEI et al., 1997), exceto para o número de locos.	45
---	----

Lista de abreviaturas e siglas

CAS	Sistema de álgebra computacional
CDF	Função distribuição acumulada
COVID	Doença do coronavírus
DEA	Distribuição de efeitos de aptidão
DNA	Ácido desoxirribonucleico
IC	Interferência clonal
MM	Mutações múltiplas
PDF	Função densidade de probabilidade
QAGS	Algoritmo de integração adaptativa por quadratura de integrando geral e que lida com singularidades
RNA	Ácido ribonucleico
SARS-CoV-2	Coronavírus 2 da síndrome respiratória aguda grave
SSWM	Seleção forte, mutação fraca

Lista de símbolos

a	Primeira constante da função hipergeométrica
a_m	Identificação de a com constante correspondente da equação para Y
A	Alelo mutante do tipo A
b	Segunda constante da função hipergeométrica
b_m	Identificação de b com constante correspondente da equação para Y
B	Alelo mutante do tipo B
c	Terceira constante da função hipergeométrica
c_m	Identificação de c com constante correspondente da equação para Y
C_1, C_2	Constantes de integração
C_m	m -ésimo coeficiente da expansão de F
d	Taxa total de mutação deletéria
E	Esperança
$E_{\Delta p}$	Esperança sobre todos os saltos de probabilidade
F	Função geratriz
f	Função densidade de probabilidade derivada de p , ou fração da população que imigra ou emigra, ou função massa de probabilidade
f_c	Aptidão da classe atual
f_c	Aptidão da classe de origem
$f(x; p, t)$	Probabilidade da frequência mudar de p para x até o tempo t
${}_2F_1$	Função hipergeométrica ordinária
$g(\delta p; p)$	Função densidade probabilidade para a mudança δp a partir de p
G	Número total de sítios
h	Fator integrante
G_i	Número total de sítios do loco do tipo i

H	Função de Heaviside discreta
m, n	Número total de sítios que sofreram mutação
$m(p)$	Salto médio em frequência a partir da frequência do alelo p
n_i	Número total de sítios do tipo i que sofreram mutação
N	Tamanho da população
N_A^t	Tamanho da população do tipo A no tempo t
N_B^t	Tamanho da população do tipo B no tempo t
N_c^t	Tamanho da população da classe atual no tempo t
N_e	Tamanho da população efetivo
N_s^t	Tamanho da população da classe de origem no tempo t
N_W^t	Tamanho da população do tipo selvagem no tempo t
p	Função distribuição acumulada do alelo selvagem se tornar mutante ou frequência do alelo na população
p_c^t	Probabilidade de sorteio da classe c no tempo t
p_i	Frequência de indivíduos na classe i
p_{ij}	Probabilidade de transição de $X = i$ para $X = j$ em uma geração
p_n	Frequência do genótipo com n mutações iguais
$p'(x)$	Frequência esperada ao fim da seleção
$P_0(p, t)$	Probabilidade de que a absorção em zero a partir da frequência p aconteceu até o tempo t
$P_1(p, t)$	Probabilidade de que a absorção em um a partir da frequência p aconteceu até o tempo t
P_A	Tamanho da população de mutantes do tipo A
P_B	Tamanho da população de mutantes do tipo B
P_A^t	Peso de amostragem do mutante do tipo A no tempo t
P_B^t	Peso de amostragem do mutante do tipo B no tempo t
P_C^t	Peso de amostragem do mutante do tipo C no tempo t

P_i	Probabilidade do passo i em uma rede. É igual ao peso W_i normalizado pelos pesos de todas as alelos mutantes disponíveis no mesmo passo.
P_n	Tamanho da população com n mutações iguais
P_W	Tamanho da população de indivíduos selvagens
P_W^t	Peso de amostragem do mutante do tipo W no tempo t
$\text{Prob}(\Delta p)$	Probabilidade do salto em frequência Δp
q	Frequência inicial do tipo mutante
$(q)_n$	Símbolo de Pochhammer
r	Taxa de mutação por sítio ou mutabilidade
r_i	Taxa de mutação por sítio do loco do tipo i
r_A	Taxa de mutação do alelo selvagem para o alelo A
r_B	Taxa de mutação do alelo selvagem para o alelo B
r_{cs}	Taxa de mutação do genótipo da classe original para o genótipo da classe atual
r_{dc}	Taxa de mutação do genótipo da classe atual para o genótipo da classe de destino imediato d
r_T	Taxa total de mutação sobre todos os sítios
s	Coefficiente de seleção
s_i	Coefficiente de seleção de um sítio do tipo i
s_A	Coefficiente de seleção do alelo mutante do tipo A
s_B	Coefficiente de seleção do alelo mutante do tipo B
s_e	Coefficiente de seleção efetivo
t	Tempo
$\bar{t}(p)$	Tempo médio até que alguma absorção aconteça, dada a frequência inicial p
$t^*(x; p)$	Tempo de permanência (<i>soujorn time</i>) da frequência inicial p até a frequência x
T	Função de t que é fator multiplicativo de F

u	Taxa de mutação do alelo A_1 para o alelo A_2
v	Taxa de mutação do alelo A_2 para o alelo A_1
Var	Variância
$v(p)$	Variância do salto em frequência a partir da frequência do alelo p
w	Aptidão ou solução geral da equação hipergeométrica
W_i	Peso do passo i em uma rede. É o produto da taxa de mutação para gerar um único indivíduo mutante do tipo i com a probabilidade de fixação deste indivíduo.
x	Frequência do alelo mutante na sopa de gametas
X	Número de alelos do tipo A_1
\bar{w}	Aptidão média
w_m	Média da frequência dos alelos após a seleção, ponderada pela aptidão
w_n	Aptidão dos genótipos com n mutações iguais
W	Alelo selvagem
Y	Função hipergeométrica de z que é fator multiplicativo de Z
z	Variável da função geratriz ou ploidia
Z	Função de z que é fator multiplicativo de F
α	Constante igual tanto a T como a Z , ou igual a $2Ns$
α_m	Constante α dependente do número de mutações m
β	Parâmetro da distribuição de Gumbel
δ_{ij}	Delta de Kronecker
δp	Mudança de p no intervalo de tempo $(0, \delta t)$
δt	Duração de um intervalo de tempo
Δt	Intervalo de uma geração
Δp	Salto em frequência
θ	Suprimento mutacional
π_f^A	Probabilidade de fixação do mutante do tipo A

π_f^B	Probabilidade de fixação do mutante do tipo B
π_f^W	Probabilidade de fixação do tipo selvagem
π_a^A	Probabilidade de absorção do mutante do tipo A
π_a^B	Probabilidade de absorção do mutante do tipo B
π_a^W	Probabilidade de absorção do tipo selvagem
$\pi(p)$	Probabilidade de fixação do alelo com frequência p
Π_f^A	Probabilidade de fixação do mutante do tipo A considerando apenas o efeito da seleção, sem considerar efeitos da mutação
Π_f^B	Probabilidade de fixação do mutante do tipo B considerando apenas o efeito da seleção, sem considerar efeitos da mutação
Π_f^i	Probabilidade de fixação do tipo i (que pode ser igual a A ou B) considerando apenas o efeito da seleção, sem considerar efeitos de mutação
ρ	Fator de decaimento de certas taxas de mutação
σ	Média da distribuição de efeitos de aptidão
σ^2	Variância do ruído gaussiano sobre as aptidões
\sum_d	Soma sobre todos os destinos d da classe atual
τ	Tempo de fixação
$\phi(t; p)$	Função densidade do tempo t até a absorção a partir da frequência p
ψ_i	Fator que generaliza o modelo de Wright-Fisher quando existem i indivíduos do alelo A_1 sob seleção e/ou mutação
ω_A	Exponencial de $-2s_A$
ω_B	Exponencial de $-2s_B$

Sumário

1	INTRODUÇÃO	26
1.1	Contexto	26
1.2	Biologia molecular	27
1.3	Mutabilidade	28
1.4	Caminhadas adaptativas	30
1.5	Simulação	33
1.6	Problema	34
1.7	Objetivo	35
1.8	Implicações	35
2	ADAPTAÇÃO DE POPULAÇÃO INFINITA	36
2.1	Introdução	36
2.2	Métodos	37
2.3	Resultados	44
2.4	Discussão	56
2.5	Conclusão	58
3	MUTABILIDADE E PASSO EVOLUTIVO	59
3.1	Introdução	59
3.2	Métodos	59
3.3	Resultados	62
3.4	Discussão	72
3.5	Conclusão	73
4	ORDEM DAS MUTAÇÕES	75
4.1	Introdução	75
4.2	Métodos	75
4.3	Resultados	78
4.4	Conclusão	93
5	DESTINO EVOLUTIVO	95
5.1	Introdução	95
5.2	Métodos	95
5.3	Resultados	99
5.4	Conclusão	105
6	CONCLUSÃO	109

REFERÊNCIAS	111
-----------------------	-----

APÊNDICES	120
------------------	------------

APÊNDICE A – PLOIDIA E PROPRIEDADES DE FIXAÇÃO . . .	121
---	------------

A.1	Probabilidade de fixação	121
------------	---	------------

A.2	Tempo de fixação	123
------------	-----------------------------------	------------

A.3	Independência da ploidia	125
------------	---	------------

ANEXOS	127
---------------	------------

1 Introdução

1.1 Contexto

O estudo da evolução biológica traz implicações imediatas para diversas áreas. De um ponto de vista mais fundamental, compreendemos a história das formas de vida e um panorama mais completo de quem somos e de nossa estrutura, nos situando melhor no universo.

Essa história também traz lições sobre adaptação, relações ecológicas e emergências em seres vivos. As aplicações vão desde a engenharia genética, passando por controle de epidemias, melhoramento de espécies, estratégias de preservação ambiental, predições do futuro da biosfera, terapias de desordens genéticas, identificação de genes, até inspiração para algoritmos genéticos, usados inclusive em estudos de superfícies em nosso departamento.

Do ponto de vista da Física, a evolução é tanto interessante como oportunidade de avanços em p. ex. Física Estatística e Sistemas Complexos, como pelo potencial de formalização de processos evolutivos em uma linguagem matemática e dos próprios fenômenos físicos.

A abordagem quantitativa da evolução, baseada na teoria de herança de Mendel, tem importantes fundamentos lançados no começo do século XX com o teorema de Hardy-Weinberg e os trabalhos monumentais de Ronald Fisher, J. B. S. Haldane e Sewall Wright. Algumas de suas contribuições são a inalterância da frequência de alelos quando da ausência de fatores evolutivos como a seleção, a deriva genética, a migração e a mutação; a conciliação da observação de variações contínuas com a natureza particulada dos genes; as bases matemáticas para a mudança de frequência de alelos de um único gene; o estudo de combinações de genes e o efeito da deriva genética em pequenas populações em endogamia.

A essa abordagem quantitativa a Física pode contribuir significativamente. Um forte motivo para isso é que ela apresenta muitos pontos de contato com a Genética de Populações, área que responde por contribuições importantes à compreensão da evolução em bases matemáticas, como as apresentadas acima. Barbara Drossel (DROSSEL, 2001) já apontava: “médias de ensemble, equações de difusão, panorama energético, autômatos celulares e simulações computacionais de redes de unidades interagentes pertencem à caixa de ferramentas da teoria evolutiva”.

Vladar e Barton (VLADAR; BARTON, 2011) acreditam que as teorias físicas podem contribuir para o avanço da compreensão do processo evolutivo através de técnicas matemáticas em comum, e pelo desenvolvimento de analogias que incluem novos conceitos,

como entropia e informação. Como uma ilustração da intimidade entre as áreas, recomendamos o trabalho de Paul G. Higgs, que apresenta (HIGGS, 1995) dois exemplos em que problemas idênticos surgiram independentemente na Física Estatística e na Genética de Populações.

Dentro da Física, talvez a Física Estatística compartilhe mais semelhanças fundamentais com a Genética de Populações (DROSSEL, 2001; VLADAR; BARTON, 2011). Ambas modelam populações ao invés de indivíduos (em uma descrição de macroestados com poucas variáveis), e esses macroestados são suficientes para prever propriedades independentemente dos microestados.

Algumas analogias matemáticas podem ser sugeridas (BARTON; COE, 2009), como entre a fórmula de Wright para a distribuição estacionária de frequência de alelos e a distribuição de Boltzmann; a temperatura inversa e o tamanho de uma população; aptidão e energia; e a possibilidade de definição de entropia em Genética de Populações. Neste trabalho a Física Estatística está a serviço da Genética de Populações.

Abordamos uma pergunta evolutiva importante através da análise de uma equação mestra e de simulações. A ênfase é a pergunta de cunho biológico, o que difere de muitas abordagens de físicos à evolução (VLADAR; BARTON, 2011).

1.2 Biologia molecular

Para se falar nos conceitos de Genética de Populações, é importante mencionar suas origens moleculares. Por exemplo, a seleção natural se dá sobre o perfil de caracteres dos indivíduos da população, e esse perfil está associado à multitude de funções essenciais realizados pelas proteínas, que são moléculas formadas por uma ou mais cadeias de aminoácidos.

As proteínas são codificadas no material genético, que pode ser DNA ou RNA. Ambos são compostos por unidades de quatro tipos, chamadas de nucleotídeos, e sujeitas a mudanças pontuais. Note que neste trabalho, o código genético que usamos nas simulações é binário, sem perda de generalização.

Quando o material genético é o DNA, ele é primeiramente transcrito em RNA, que depois vai ser traduzido em uma cadeia de aminoácidos. À sequência de aminoácidos de uma proteína chamamos estrutura primária. Essa cadeia irá se enovelar, formando a estrutura tridimensional da cadeia, ou estrutura terciária, que é fundamental para a eficácia e eficiência das funções das proteínas.

As cadeias de aminoácidos estão codificadas no DNA em genes, que estão posicionados em locos. Aqui trataremos locos ou sítios de nucleotídeos indistintamente, como os locais discretos em que as mutações acontecem.

Portanto, modificações em sítios no DNA ou no RNA podem trazer alterações no funcionamento das proteínas, e em última instância, na capacidade de sobrevivência do indivíduo em questão. Dizemos que o genótipo (a composição do material genético) está associado ao fenótipo (conjunto de caracteres do organismo). Para nossos propósitos, mapeamos os fenótipos nos genótipos por uma relação biunívoca.

Podemos também mapear os genótipos diretamente às vantagens adaptativas associadas com eles. Tal mapa mais as relações de mutação entre os genótipos define uma paisagem adaptativa ou paisagem genética (SVENSSON; CALSBEEK, 2012). Muitos trabalhos recentes têm sido feitos para mapearem tais paisagens e de suas variantes (PISTONO et al., 2022; GABZI; PILPEL; FRIEDLANDER, 2022; SINAI et al., 2022; CHEN; FOWLER; TOKURIKI, 2022; WU et al., 2022; SONG; ZHANG, 2021; LALANNE; PARKER; LI, 2021; SOO et al., 2021; COSSIO-DIAZ; UGUZZONI; PAGNANI, 2021; BOLTON et al., 2020; MARTIN; GOULD, 2020; TIAN; BEST, 2020; YANG et al., 2019; YANG et al., 2020; GRECH et al., 2019; TAMER et al., 2019; POKUSAEVA et al., 2019). Uma paisagem adaptativa biofísica, em que se mede a afinidade de uma proteína ao invés da aptidão, foi realizada para a variante delta do SARS-Cov-2, o causador da pandemia da COVID (PATRICK et al., 2022).

Os genes estão em sequências contíguas de DNA chamados cromossomos. Ao número de cópias dos cromossomos presentes no material genético denominamos ploidia. Quando há uma só cópia, a população é haplóide; se duas, diplóide; se três, triplóide, e assim por diante. À exceção da discussão de fórmulas relacionadas à ploidia no Apêndice A, por simplicidade estaremos lidando com populações haplóides.

Em organismos poliplóides, o efeito de um variante de um gene, ou alelo, pode prevalecer sobre os demais, constituindo o efeito de dominância. Em organismos diplóides, se para um caráter surgir basta que apenas um dos alelos corresponda a ele, dizemos que o alelo é dominante. Se duas cópias são necessárias, ele é recessivo. Note que a taxa de mutação necessária para gerar este último caráter em uma geração é o quadrado da metade da taxa para gerar o caráter dominante, dado que a taxa de mutação entre o alelo original e o responsável pelo caráter é fixa entre os sistemas de dominância.

1.3 Mutabilidade

Existe em seres vivos uma maquinaria complexa para tornar as mutações raras. No entanto taxas de mutação mais altas são selecionadas em muitas populações experimentais e naturais (DENAMUR; MATIC, 2006; BAER; MIYAMOTO; DENVER, 2007). As taxas de mutação observadas resultam de forças evolutivas diferentes. A compreensão dessas forças ainda é obscura.

A propagação de genes que promovem mutações (ou simplesmente mutadores) por

carona com mutações adaptativas (JOHNSON, 1999a; JOHNSON, 1999b; TENAILLON et al., 2000; SNIEGOWSKI et al., 2000) parece ser muito mais provável de ocorrer em organismos onde não há troca de pedaços de material genético (recombinação) de forma que a associação entre o mutador e as mutações não seja quebrada.

Os alelos mutadores são encontrados em muitas populações naturais (GROSS; SIEGEL, 1981; LECLERC et al., 1996; LECLERC et al., 1998; MATIC et al., 1997; OLIVER et al., 2000; BJORKHOLM et al., 2001; DENAMUR et al., 2002; GERRISH; LENSKI, 1998; RICHARDSON et al., 2002; PRUNIER et al., 2003; WATSON; BURNS; SMITH, 2004; CAMPO et al., 2005; LABAT et al., 2005) e experimentais (COX; GIBSON, 1974; TROBNER; PIECHOCKI, 1981; CHAO; COX, 1983; CHAO et al., 1983; MAO et al., 1997; SNIEGOWSKI; GERRISH; LENSKI, 1997; GIRAUD et al., 2001; NOTLEY-MCROBB; SEETO; FERENCI, 2002; SHAVER et al., 2002; THOMPSON; DESAI; MURRAY, 2006; PAL et al., 2007). As mutações por perda de função dos sistemas de reparo do DNA implicam em um aumento das taxas de mutações de 10 a 10.000 vezes maiores que em bactérias selvagens (DENAMUR; MATIC, 2006; BAER; MIYAMOTO; DENVER, 2007). Isso tem permitido o estudo da dinâmica evolutiva de genótipos mutadores.

Experimentos iniciais de competição em quimiostatos com misturas de tipos selvagem e mutador de *Escherichia coli* revelaram uma preponderância frequente de genótipos mutadores (COX; GIBSON, 1974; CHAO; COX, 1983). O aparecimento e a fixação subsequente de tipos mutadores em três das doze linhagens selecionadas por um longo prazo do agora extinto grupo de Lenski (SNIEGOWSKI; GERRISH; LENSKI, 1997) mostrou que tais genes podem se espalhar também quando aparecem naturalmente.

Estes resultados mostram que mutadores podem se espalhar em uma população, e em uma taxa muito mais elevada que a esperada por deriva genética. Esse espalhamento se deveria a maior possibilidade de se adaptar (adaptabilidade) ou a uma seleção direta positiva? Experimentos adicionais levantam dúvidas em uma ligação necessária entre uma taxa de mutação elevada e adaptabilidade aumentada.

Quando a taxa de adaptação de *E. coli* selvagem e mutadora foi comparada diretamente, a taxa elevada de mutação teve um efeito mínimo na taxa de adaptação exceto em populações pouco adaptadas com tamanhos populacionais efetivos pequenos (VISSER et al., 1999).

A taxa de adaptação de populações maiores foi limitada pela eficiência da seleção ao invés de suprimento de mutações, com pouco efeito das elevadas taxas de mutação (GERRISH; LENSKI, 1998; VISSER et al., 1999). Assim, o suprimento mutacional elevado não garantiria um aumento da adaptabilidade.

Por outro lado, em duas das três linhagens de mutadores do grupo de Lenski acima mencionadas, a propagação do alelo mutador foi acompanhada de um aumento da taxa

de adaptação em relação a populações que não portavam um mutador. Mais ainda, não houve evidência de que o mutador fosse diretamente selecionado. No entanto os benefícios do fornecimento aumentado de mutações foi de curta duração.

A comparação da aptidão de populações de mutadores com selvagens após muitos milhares de gerações não revelou diferenças nas aptidões finais (SHAVER et al., 2002). Isso é esperado porque em condições em que a população está bem adaptada não é esperado um efeito significativo da adaptabilidade (VISSER et al., 1999). Isso quer dizer que se por um lado a adaptabilidade pode explicar a fixação dos mutadores, ela não explica a sua manutenção na população. Esta é uma questão em aberto (SHAVER et al., 2002).

Tais estudos dão apoio à predição teórica que genes com elevada taxa de mutação podem se espalhar devido a sua adaptabilidade aumentada. No entanto elas também mostram quão restritivos e transientes tais benefícios podem ser. Tal seleção é improvável de ser forte em populações naturais, ao menos que uma mudança ambiental ocorra com alta frequência.

Isso se alinha com o fato de que enquanto em populações naturais os mutadores geralmente se mantêm em baixa frequência, eles podem ser comumente encontrados com altas frequências em bactérias patogênicas (LECLERC et al., 1996; OLIVER et al., 2000), onde experimentam gargalos frequentes e forte seleção direcional.

Uma variação sobre os mutadores são as bactérias que modulam a mutagênese em ambientes específicos (ROSENBERG; THULIN; HARRIS, 1998; METZGAR; WILLS, 2000; BJEDOV et al., 2003). Por exemplo, quando *E. coli* estão famintas, suas taxas de mutação aumentam, e o aumento de mutações benéficas permite um prolongamento de sua sobrevivência nesta situação. A maquinaria celular envolvida é altamente complexa, que possibilita um ajuste fino na taxa de mutação.

Os mutadores em organismos assexuados servem com um excelente modelo para o estudo da adaptabilidade, por sua simplicidade e praticidade. Eles podem ser criados e monitorados em laboratório, permitindo testes experimentais de predições teóricas e simulacionais.

1.4 Caminhadas adaptativas

O pensamento populacional da evolução darwiniana, juntamente com a noção de herança particulada (em que os caracteres dos organismos, ou fenótipos, são transmitidos por elementos discretos hoje conhecidos por genes) são a base da Genética de Populações. Ela investiga processos como a adaptação e especiação. Dentro desta área de estudos, cinco conceitos fundamentais emergiram, a saber:

- *Evolução* ocorre sempre que acontecem mudanças nas frequências das variedades

(alelos) de um ou mais genes. Ao perfil de alelos de um indivíduo chamamos genótipo;

- *Seleção natural* um processo de sobrevivência diferencial de sobrevivência e reprodução associada ao conjunto de caracteres (fenótipos) de um indivíduo. Essa probabilidade é proporcional à medida chamada aptidão;
- *Mutação* é o processo que gera novos alelos. Ela “alimenta” a seleção natural com variabilidade, a matéria-prima da evolução. A capacidade de um indivíduo gerar novas mutações é chamada de mutabilidade;
- *Deriva genética* é a estocasticidade associada à amostragem de indivíduos de uma geração para outra, na presença ou ausência de seleção. Em geral, é tanto maior quanto menor for o tamanho de uma população;
- *Fluxo gênico* é a troca de alelos entre duas populações, tipicamente por migração de indivíduos entre elas.

A história evolutiva é um aspecto muito importante nos estudos evolutivos. Ela aponta para as origens e os percursos das diferentes espécies. Dentro dela, uma pergunta muito relevante é, se rodássemos de novo a fita da evolução, se ela ocorreria da mesma maneira. A repetibilidade ou previsibilidade de processos evolutivos pode ser medida pela entropia de Shannon, uma quantidade bem conhecida em teoria da informação.

Consideraremos apenas processos adaptativos, em que adaptações se acumulam nas populações, e três destinos possíveis para os genótipos. Um é a fixação, isto é, toda a população ser dominada por um único tipo, mesmo que ela ainda esteja sujeita a mutações. Outro é a absorção, quando o genótipo corresponde a um dos possíveis estados finais ou absorventes do processo. E finalmente pode haver a prevalência, que é sua manutenção em frequência superior a qualquer um dos outros genótipos envolvidos. Essa manutenção pode ser dinâmica, na presença de fluxos populacionais como a migração. Quando os fluxos são constantes, o estado final é denominado estacionário.

A seleção, a mutação, a deriva e o fluxo gênico são considerados as principais causas de mudança da frequência gênica em um dado momento, e portanto são os mecanismos evolutivos mais importantes.

Já falamos sobre a origem da mutação nos genes mutadores. Dependendo da mutação de interesse, podemos falar que ela é direta ou reversa. Neste trabalho as mutações diretas sempre acontecem de um genótipo associado a menor aptidão a outro de maior aptidão. Dada a ineficácia em geral da população reduzir sua aptidão por mutação, consideraremos apenas mutações diretas. Os indivíduos mutantes são apenas aqueles portadores de uma ou mais mutações em relação ao genótipo típico, ou selvagem, independente de quão radicais sejam as mudanças trazidas ao seu fenótipo. Mesmo no caso de mutações imperceptíveis, o portador ainda será mutante.

A vantagem adaptativa é medida pela aptidão, que neste trabalho é um valor proporcional à chance de sobrevivência de um indivíduo, dado seu fenótipo. De acordo com a aptidão, as mutações podem ser deletérias (diminuem a aptidão), neutras (não modificam a aptidão, mesmo que modifiquem o fenótipo) e adaptativas (aumentam a aptidão). Existe uma distribuição de efeitos de aptidão (DEA) de cada mutação, que relaciona a frequência da mutação com seu coeficiente de seleção s (que neste trabalho é definido por $w = 1 + s$, onde w é a aptidão de um indivíduo, e $w = 1$ é a aptidão do indivíduo selvagem).

Em microorganismos como a *E. coli*, cujos tamanhos de população podem chegar a mais de $N = 10^9$ indivíduos, as taxas de mutação adaptativas são de cerca de 10^{-10} por nucleotídeo por indivíduo por geração, e que podem ser flexivelmente modificadas pela ação conjunta de mutadores e antimutadores. O suprimento mutacional θ é a taxa de mutação de interesse vezes o tamanho da população. Se a população é inteiramente selvagem, θ mede na prática a chance de surgir ao menos um mutante na população. Quando admitimos que mais de um mesmo evento mutacional pode acontecer na população ao longo do tempo, dizemos que a mutação é recorrente.

Ao contrário do senso comum, a evolução não implica necessariamente no aumento da aptidão de uma população. Por exemplo, ela pode envolver a mudança da frequência de alelos neutros (sem seleção natural) por deriva genética. O aumento em aptidão de uma população devido à seleção natural chama-se adaptação.

A adaptação pode se dar através de mutações sequenciais em uma mesma linhagem (conjunto de indivíduos que compartilham um ancestral comum) ou de mutações simultâneas em linhagens diferentes (Figura 1), tal qual a melodia e a harmonia em uma música. A esses modos chamamos mutações múltiplas (MM) e interferência clonal (IC) (FOGLE; NAGLE; DESAI, 2008), respectivamente, que podem acontecer concomitantemente em uma dinâmica. Ambas são prevalentes em experimentos evolutivos (LANG et al., 2013).

Independentemente dos modos de adaptação, podemos imaginar uma população sendo descrita como uma série de alelos adaptativos que se fixam nela, em uma espécie de caminhada evolutiva (Figura 2). Quando as taxas de mutação são baixas e a seleção natural é suficientemente forte a adaptação acontece praticamente apenas por MM, no que chamamos fixações sequenciais. Definimos a fixação apenas na presença de seleção à fixação que desconsidera efeitos em particular da mutação gerando os mutantes iniciais ou outros ao longo de seu processo. A probabilidade de um passo adaptativo, i.e. a taxa de transição de um estado para outro, torna-se o produto da taxa de mutação e da probabilidade de fixação de um alelo na presença de seleção apenas. Modelos que adotam essa taxa de transição evolutiva são chamados de origem-fixação (MCCANDLISH; STOLTZFUS, 2014) e possuem grande poder de formular previsões. Cadeias de Markov resultantes desta



Figura 1 – Modos de adaptação. Cada cor representa um mutante distinto. A cor branca representa uma população inicial selvagem. Acima: Mutações sequenciais. Cada nova mutação adaptativa se acumula a outras anteriores. Este é o regime do modelo origem-fixação. Abaixo: Interferência clonal. As mutações coexistem e competem entre si por fixação.

modelagem, quando obedecem o balanço detalhado, são prontamente analisáveis usando-se ferramentas da Física Estatística.

Mas o que acontece se nos voltarmos para além de modelos origem-fixação, como quando a mutabilidade varia desde valores muito baixos (HERR; WILLIAMS; PRESTON, 2011) para valores muito altos (GENTILE et al., 2011)? As caminhadas evolutivas seriam alteradas por um eventual aumento de IC com mutabilidades altas? Novos fenômenos seriam possíveis?

1.5 Simulação

Existem diferentes esquemas para sorteio dos tipos a cada geração, ao longo de uma simulação de um processo evolutivo. Um dos mais comuns é o modelo de Wright-Fisher (EWENS, 2004) e suas variantes.

O modelo de Wright-Fisher é um modelo de geração discreta, e em sua forma original, admite que os alelos na geração seguinte são amostrados com reposição da geração anterior. Em um sistema com um loco e dois alelos neutros apenas, o número $X(t)$ do alelo A_1 será uma variável aleatória binomial. A probabilidade de que, dado que



Figura 2 – Caminhada evolutiva. Esquerda: as mutações possíveis (setas) entre os genótipos (círculos), e a caminhada evolutiva (setas pretas) realizada de um genótipo inicial (círculo branco) até um genótipo final (círculo preto). Direita: as possíveis caminhadas evolutivas em termos de dinâmica de aptidão (vertical) da população ao longo do tempo (medido em número de eventos adaptativos, da esquerda para a direita), e a caminhada evolutiva realizada (linha preta).

$X(t) = i, X(t + 1) = j$ é dada por

$$p_{ij} = \binom{N}{j} (i/N)^j [1 - (i/N)]^{N-j}, \quad (1.1)$$

para $i, j = 0, 1, 2, \dots, N$.

Em variantes do modelo de Wright-Fisher que consideram mutação e/ou seleção, o termo i/N é substituído por uma expressão que reflita a nova probabilidade de cada alelo A_1 . Por exemplo, se existe uma taxa de mutação u de A_1 para A_2 e outra taxa reversa v de A_2 para A_1 , cada termo i/N na Equação 1.1 é substituído por

$$\psi_i = (1 - u)(i/N) + [1 - (i/N)]v. \quad (1.2)$$

Para mais alelos, a distribuição binomial é substituída pela multinomial, como veremos caso a caso nos capítulos 3 a 5.

1.6 Problema

Na formulação de nossa questão consideramos um fato adicional. Espera-se que a cauda da DEA decaia monotonicamente (ORR, 2002), especialmente quando a aptidão do tipo selvagem é alta (MACLEAN; BUCKLING, 2009). Comparando-se duas mutações distintas quaisquer com tal DEA (expressa como histograma), uma será necessariamente mais vantajosa mas com uma taxa de mutação menor, e outra terá taxa de mutação maior, mas uma vantagem menor.

Nesse contexto, torna-se incerto qual destes alelos seria mais propenso a se fixar. Para uma mutabilidade baixa, o modelo origem-fixação seria mais adequado. O alelo

menos apto poderia ter maiores chances de se fixar, por ter a maior taxa de mutação, ao mesmo tempo em que as probabilidades de fixação não se alteram tanto quanto estas taxas. Mas para mutabilidades altas (que podem ocorrer mesmo para linhagens selvagens (PERFEITO et al., 2007)), a competição entre alelos torna-se provável, e o modelo origem-fixação estaria fora de seu regime de validade. Se o alelo mais apto surgir em número suficiente, a caminhada evolutiva passa a “escolher” o passo de maior aumento de aptidão para a população.

1.7 Objetivo

Demonstrar que a mutabilidade altera a caminhada evolutiva de populações, favorecendo a fixação de alelos em cada passo de acordo com os balanços entre a taxa de mutação e a vantagem adaptativa.

1.8 Implicações

A tese tem implicações importantes para se compreender a evolução das taxas de mutação, a repetibilidade do processo adaptativo (LENORMAND; ROZE; ROUSSET, 2009), restringindo os caminhos evolutivos adicionalmente à epistasia (WEINREICH et al., 2006), a evolução da diploidia (OTTO; GERSTEIN, 2008), a evolução da dominância (BOURGUET, 1999) e a ordem com que as mutações adaptativas acontecem.

2 Adaptação de população infinita

2.1 Introdução

O estudo de populações infinitas é uma investigação preliminar, mas fundamental, realizada desde os primórdios da Genética de Populações (e.g. a lei de Hardy-Weinberg (EWENS, 2004)) até os dias atuais (DONNELLY; KURTZ, 1999; GILLESPIE, 2000; MATSUDA; ISHII, 2001; SZOLLOSI; DERENYI, 2008; PARK; SIMON; KRUG, 2010; SAAKIAN; KIRAKOSYAN; HU, 2012; EVANS; STEINSALTZ; WACHTER, 2013; SAAKIAN; GHAZARYAN; HU, 2014; RIO; STEPHENS; ROSENBLUETH, 2015; YAKUSHKINA; SAAKIAN, 2017; SAAKIAN, 2017; SAAKIAN, 2018; YAKUSHKINA; SAAKIAN, 2018). Apesar de por vezes perder o realismo da estocasticidade, ele permite a obtenção de comportamentos limites e expressões fechadas. Tais resultados orientam os aspectos a serem investigados em populações finitas. Eles também funcionam como experimentos-controle para a investigação de efeitos estocásticos, em particular a deriva genética.

Por vezes, resultados em populações infinitas conservam aspectos estocásticos. Um exemplo é a probabilidade de fixação de um indivíduo de maior aptidão em uma população infinita selvagem. De maneira ingênua, poderíamos esperar que a dinâmica populacional fosse determinística, obedecendo a equações diferenciais usuais, e assim a população sempre seria dominada pelo indivíduo mais apto.

Todos os nossos modelos são do tipo Wright-Fisher, com N sendo o tamanho da população, independente do tempo por hipótese. Podemos tomar o limite $N \rightarrow \infty$ da probabilidade de fixação de um indivíduo mais apto na presença de seleção apenas de um modelo de população finita. Neste caso tal probabilidade será aproximadamente $2s$, onde s é o coeficiente de seleção do alelo mais apto (EWENS, 2004). Esse resultado foi obtido para diplóides, mas veremos no Apêndice que se aplica a haplóides também, que é o foco de nosso interesse.

Estudaremos neste capítulo prioritariamente populações haplóides infinitas que estão sujeitas a G mutações sucessivas, onde cada uma delas sempre contribui com o mesmo $s > 0$. Infelizmente para este processo não pudemos conservar aspectos estocásticos na solução da dinâmica deste modelo, mas obtivemos resultados fáceis de serem generalizados tanto para o número de tipos de alelos como para incluir mutações deletérias.

Nós obtivemos com o auxílio de um sistema de computação algébrico (CAS) uma solução exata, inédita para aptidão aditiva, e uma aproximação útil para um regime em que a taxa de mutação r é muito menor do que o coeficiente de seleção s em módulo, e este é muito menor que a unidade, diretamente a partir do desenvolvimento da equação mestra.

Uma solução mais ampla mas aparentemente não exata, baseada em um *ansatz* ao invés de uma dedução rigorosa, foi obtida por John e Jain (JOHN; JAIN, 2015). Aqui mostramos todo o passo-a-passo para a solução exata e para uma solução simplificada. Esperamos que nossa estratégia possa inspirar novas soluções para outros modelos evolutivos.

2.2 Métodos

Inicialmente, proporemos a equação mestra para o modelo em questão. Dela obteremos diretamente soluções numéricas, normalizadas passo-a-passo, a serem comparadas com as soluções analíticas (ver a seção Resultados). Em seguida, deduziremos a equação diferencial para a função geratriz. Após esta etapa, esboçaremos a solução geral. Finalmente, obteremos as distribuições das classes de genótipos, onde uma classe é definida pelo número de mutações adaptativas iguais em um genótipo, e a expressão para a aptidão média. O modelo corresponde ao do Monte Fuji (AITA; HUSIMI, 1996).

A solução da equação diferencial para a função geratriz foi primeiramente obtida através do CAS Mathematica. A partir daí, reconhecemos a solução hipergeométrica de uma equação-chave, e obtivemos as probabilidades de cada um dos estados. Relataremos aqui os resultados como se eles tivessem sido obtidos quase exclusivamente por álgebra.

Assim, nosso modelo consiste em um genoma com G loci adaptativos e cada genótipo é definido pelo número n de alelos favoráveis. Todas as mutações são favoráveis, o que implica que as mutações do alelo favorável para o selvagem serão desprezadas. A Figura 3 ilustra o modelo e apresenta os fluxos de cada genótipo. A taxa de transição para sair de um genótipo é dada pelo número de alelos selvagens ($G - n$) vezes a taxa de mutação por locus $r = r_T/G$, onde r_T é a taxa total de mutações favoráveis.

A dinâmica em cada geração se dá em duas etapas. Na primeira, cada genótipo cresce de acordo com sua aptidão. Na segunda, acontecem os fluxos entre as populações. A Figura 3, na ausência de seleção, poderia representar o processo de Decaimento de Poisson para os sítios de mutação de um único indivíduo. Portanto, nos referiremos provisoriamente ao nosso processo como Decaimento de Poisson Evolutivo.

Definindo como $P_n(t)$ a população do genótipo n no tempo t , e usando a aptidão aditiva $w_n = 1 + ns$ para o genótipo n , as etapas da dinâmica dentro de uma geração são dadas por

$$P_n(t + 1/2) = w_n P_n(t) \tag{2.1}$$

$$P_n(t + 1) = P_n(t + 1/2) + [G - (n - 1)]r P_{n-1}(t + 1/2) - (G - n)r P_n(t + 1/2). \tag{2.2}$$

Na equação 2.2 o primeiro termo é a população do genótipo n no tempo anterior, o segundo é o quanto esta população recebe do genótipo $n - 1$, e o último termo, o quanto

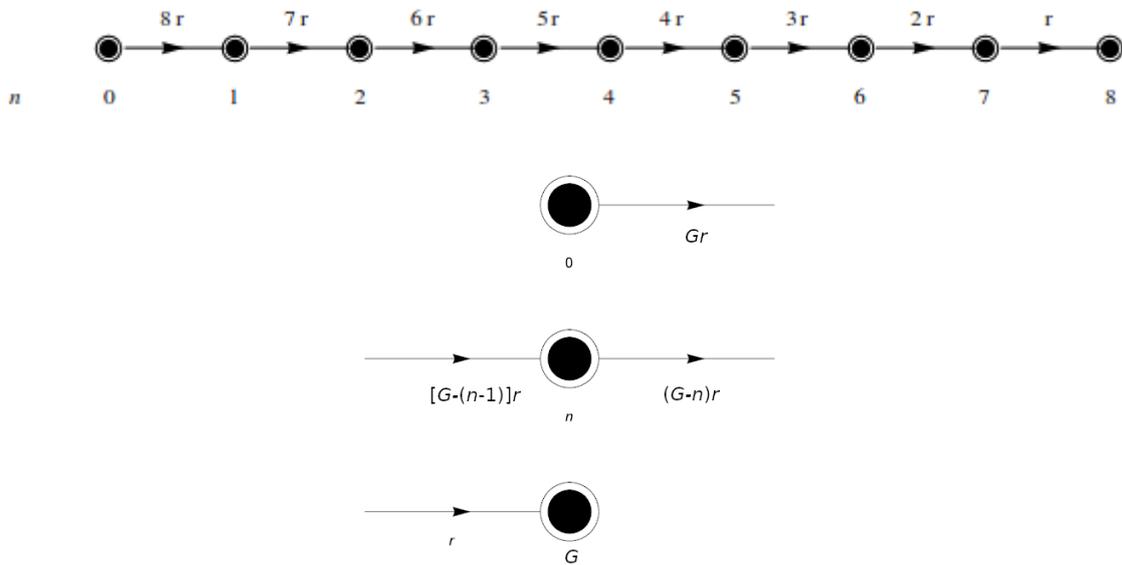


Figura 3 – O modelo simplificado, sem as remoções por mutações deletérias ou letais. A variável n é o número de alelos favoráveis, e r é a taxa de mutações favoráveis por *loco*. De cima para baixo: um exemplo para $G = 8$ locos; as taxas de mutação do genótipo com n alelos favoráveis quando o genótipo é selvagem, $n = 0$; idem, quando o genótipo tem $1 \leq n \leq G - 1$ alelos favoráveis; e quando o genótipo só possui alelos favoráveis, $n = G$.

ela transfere para o genótipo $n + 1$. Note que esta equação vale para $n = G$, que é um estado absorvente, pois o termo de transferência para $n = G + 1$ se anula. Não custa lembrar que $P_n(t)$ não é uma quantidade normalizada. Tal quantidade será obtida ao final da análise.

A forma da aptidão é aditiva, isto é, somam-se os coeficientes de seleção de interesse para obter-se um coeficiente final. No nosso caso, quando há n mutações de coeficiente de seleção s o coeficiente de seleção final será ns .

Juntando as equações 2.1 e 2.2 em uma expressão para uma só geração, obtemos

$$\begin{aligned}
 P_n(t+1) &= (1 + ns)P_n(t) + [1 + (n-1)s][G - (n-1)]rP_{n-1}(t) + \\
 &\quad - (1 + ns)(G - n)rP_n(t).
 \end{aligned}
 \tag{2.3}$$

Para $n = 0$, as equações têm a forma

$$\begin{aligned}
 P_0(t + 1/2) &= P_0(t) \\
 P_0(t + 1) &= P_0(t + 1/2) - GrP_0(t + 1/2) \\
 &= (1 - Gr)P_0(t).
 \end{aligned}
 \tag{2.4}$$

Por indução, considerando que a distribuição inicial de genótipos está concentrada no estado selvagem ($P_0(0) = 1$), podemos mostrar a partir da equação 2.4 que

$$P_0(t) = (1 - Gr)^t.
 \tag{2.5}$$

Vamos agora expandir os termos da Equação 2.3 e rearranjá-los para obtermos a aproximação para tempo contínuo. Vamos então expandir os termos.

$$P_n(t+1) = P_n(t) + nsP_n(t) + GrP_{n-1}(t) + (n-1)GrsP_{n-1}(t) - (n-1)rP_{n-1}(t) - (n-1)^2rsP_{n-1}(t) - GrP_n(t) + nrP_n(t) - nGrsP_n(t) + n^2rsP_n(t). \quad (2.6)$$

Passando o primeiro termo do lado esquerdo para o lado direito, e agrupando as potências de n e $n-1$, temos

$$P_n(t+1) - P_n(t) = GrP_{n-1}(t) + (n-1)(Gs-1)rP_{n-1}(t) - (n-1)^2rsP_{n-1}(t) - GrP_n(t) + [r+s-Grs]nP_n(t) + n^2rsP_n(t). \quad (2.7)$$

Tomando o intervalo de uma geração $\Delta t = 1$ como dt , e $P_n(t+1) - P_n(t)$ como dP_n , temos

$$\begin{aligned} \dot{P}_n(t) &= -GrP_n(t) + (s+r-Grs)nP_n(t) + rsn^2P_n(t) + \\ &+ GrP_{n-1}(t) - (r-Gsr)(n-1)P_{n-1}(t) + \\ &- rs(n-1)^2P_{n-1}(t). \end{aligned} \quad (2.8)$$

A equação 2.6 que representa a dinâmica de tempo discreto é linear e pode ser escrita na forma $P(t+1) = AP(t)$, onde A é uma matriz e $P(t)$ é uma matriz coluna com $P_0(t), P_1(t), \dots, P_n(t)$ como componentes. A solução exata é $P(t) = A^tP(0)$.

Por outro lado, se B é uma outra matriz, a solução exata da dinâmica de tempo contínuo $\dot{P}(t) = BP(t)$ é $P(t) = e^{tB}P(0)$. A solução da dinâmica irá coincidir exatamente com a dinâmica contínua quando t assume os valores $1, 2, \dots$ se $e^B = A$, ou $B = \log A = \log[I + (A - I)]$.

Caso $A - I$ seja uma matriz pequena, $\log[I + (A - I)] \approx A - I$. Sabemos portanto, a priori, que a troca da dinâmica discreta (equação 2.6) pela dinâmica contínua (equação 2.8) será uma boa aproximação se os elementos da matriz $A - I$ forem pequenos.

Infelizmente não conseguimos estabelecer isto. Vamos prosseguir daqui como se a aproximação feita para a equação 2.8 fosse válida.

Para resolvermos a equação mestra 2.8, definimos a função geratriz

$$F(z, t) = \sum_{n=0}^{\infty} z^n P_n(t), \quad (2.9)$$

com derivada temporal

$$\frac{\partial F(z, t)}{\partial t} = \sum_{n=0}^{\infty} z^n \dot{P}_n(t). \quad (2.10)$$

Substituindo-se a equação mestra 2.8 na equação 2.10, obtemos

$$\begin{aligned}
 \frac{\partial F(z, t)}{\partial t} = & -Gr \sum_{n=0}^{\infty} z^n P_n(t) + (s + r - srG) \sum_{n=0}^{\infty} n z^n P_n(t) + \\
 & + rs \sum_{n=0}^{\infty} n^2 z^n P_n(t) + Gr \sum_{n=1}^{\infty} z^n P_{n-1}(t) + \\
 & - (r - Gsr) \sum_{n=1}^{\infty} (n-1) z^n P_{n-1}(t) + \\
 & - rs \sum_{n=1}^{\infty} (n-1)^2 z^n P_{n-1}(t).
 \end{aligned} \tag{2.11}$$

Expressando os somatórios do lado direito da equação 2.11 como funções de F , obtemos

$$\begin{aligned}
 \frac{\partial F(z, t)}{\partial t} = & rsz^2(1-z) \frac{\partial^2 F}{\partial z^2} + [s + r(1+s-Gs)(1-z)] z \frac{\partial F}{\partial z} + \\
 & -Gr(1-z)F.
 \end{aligned} \tag{2.12}$$

A condição inicial é

$$P_n(0) = \delta_{n0},$$

onde δ_{ij} é o delta de Kronecker.

A Equação 2.12 é a equação diferencial para a função geratriz que pretendemos resolver.

Vamos tentar solucionar a equação diferencial parcial 2.12 por separação de variáveis. Assumamos que $F(z, t)$ tenha a forma $T(t)Z(z)$. Substituindo-a na equação 2.12 e dividindo toda a expressão por $T(t)Z(z)$, obtemos

$$\begin{aligned}
 \frac{\dot{T}}{T} = & rsz^2(1-z) \frac{\ddot{Z}}{Z} + [s + r(1+s-Gs)(1-z)] z \frac{\dot{Z}}{Z} + \\
 & -Gr(1-z).
 \end{aligned} \tag{2.13}$$

Esta equação só tem solução se ambos os lados forem iguais a uma constante α . A função $T(t)$ pode ser facilmente obtida agora:

$$T(t) = C \exp(\alpha t). \tag{2.14}$$

A equação para $Z(z)$ torna-se

$$\begin{aligned}
 rsz^2(1-z) \ddot{Z} + [s + r(1+s-Gs)(1-z)] z \dot{Z} + \\
 -[\alpha + Gr(1-z)] Z = 0.
 \end{aligned} \tag{2.15}$$

Vamos aplicar o método de Frobenius à equação 2.15. Façamos

$$Z(z) = \sum_{l=0}^{\infty} a_l z^{l+m}. \tag{2.16}$$

Para identificarmos $P_n(t)$ na solução em série para $F(z, t)$, a potência de z na função geratriz precisa ser maior ou igual a zero. Portanto, na Equação 2.16, m precisa ser um inteiro não-negativo para que essa condição seja sempre atendida. Substituindo-se o somatório da Equação 2.16 e suas derivadas na equação 2.15 obtemos a seguinte fórmula envolvendo potências de z e seus coeficientes:

$$\begin{aligned} & \{rsm(m-1) + [s + r(1 + s - Gs)]m - (Gr + \alpha)\}a_0z^m + \\ & + \sum_{l=0}^{\infty} \{rs(l+m+1)(l+m)a_{l+1} - rs(l+m)(l+m-1)a_l + \\ & + [s + r(1 + s - Gs)](l+m)a_{l+1} - r(1 + s - Gs)(l+m)a_l + \\ & - (Gr + \alpha)a_{l+1} + Gra_l\}z^{l+m} = 0. \end{aligned}$$

A equação indicial é portanto

$$rsm(m-1) + [s + r(1 + s - Gs)]m - (Gr + \alpha) = 0, \quad (2.17)$$

que resulta para α

$$\begin{aligned} \alpha &= \alpha_m = rsm(m-1) + [s + r(1 + s - Gs)]m - Gr \\ \alpha_m &= m[r + s - (G - m)rs] - Gr, \end{aligned} \quad (2.18)$$

e para m ,

$$m = \frac{\pm \sqrt{4rs(\alpha_m + Gr) + (Grs - r - s)^2 + Grs - r - s}}{2rs}. \quad (2.19)$$

O argumento da raiz quadrada na equação 2.19, após a substituição de α_m de 2.18, torna-se

$$4rs(\alpha_m + Gr) + (Grs - r - s)^2 = (r + s - Grs + 2mrs)^2. \quad (2.20)$$

A única raiz quadrada que nos interessa para que a equação 2.19 seja verdadeira é $r + s - Grs + 2mrs$ antecedida do sinal positivo, porque se tivesse um valor negativo, m tornaria-se função de G , r e s . Além de ser uma variável independente, m precisa ser um inteiro não-negativo. Esta é uma condição para F ser a função geratriz para $P_n(t)$.

Vamos tentar encontrar uma função que representa a somatória da equação 2.16 da seguinte forma:

$$Z(z) = z^m Y(z) \quad (2.21)$$

Vamos substituir α_m na equação 2.15.

$$\begin{aligned} & rs(1-z)z^2 \ddot{Z}(z) + [-rz(-Gs + s + 1) + r(-Gs + s + 1) + \\ & + s]z \dot{Z}(z) - [m(-Grs + mrs + r + s) - Grz]Z(z) = 0. \end{aligned} \quad (2.22)$$

Substituindo agora a expressão para $Z(z)$ da Equação 2.21 e suas derivadas na equação 2.22, após manipulação algébrica simples, temos

$$\begin{aligned} & z(1-z)\ddot{Y}(z) + \left\{-G + 2m + 1 + \frac{1}{s} + \frac{1}{r}\right\} \\ & - \left[\left(m - G\right) + \left(m + \frac{1}{s}\right) + 1\right]z\dot{Y}(z) - (m - G)\left(m + \frac{1}{s}\right)Y(z) = 0. \end{aligned} \quad (2.23)$$

A equação 2.23 é um caso particular da equação hipergeométrica

$$\{z(1-z)\ddot{w}(z) + [c - (a+b+1)z]\dot{w}(z) - abw(z)\} = 0, \quad (2.24)$$

com soluções

$$Y(z) = {}_2F_1(a, b; c; z), \text{ ou} \quad (2.25)$$

$$Y(z) = z^{1-c} {}_2F_1(1+a-c, 1+b-c; 2-c; z) \quad (2.26)$$

para c não-inteiro, que é o caso mais geral. Sendo assim, o termo multiplicativo z^{1-c} não conduz a uma solução da equação 2.12 na forma da equação 2.9, e a segunda solução precisa ser descartada.

Se identificarmos $a = a_m = m - G$, $b = b_m = m + \frac{1}{s}$ e $c = c_m = -G + 2m + 1 + \frac{1}{s} + \frac{1}{r}$ entre as equações 2.24 e 2.23, temos que a solução é

$$Y(z) = {}_2F_1(a_m, b_m; c_m; z). \quad (2.27)$$

O parâmetro a_m em ${}_2F_1(a_m, b_m; c_m; z)$ é igual a $m - G$, que é um inteiro. Para ser não-positivo e truncar a série em z de $Y(z)$, $m \leq G$. A restrição então é que $0 \leq m \leq G$. A solução $Z(z)$ é portanto

$$Z_m(z) = z^m {}_2F_1(a_m, b_m; c_m; z), \quad (2.28)$$

com $a_m = m - G$, $b_m = m + \frac{1}{s}$ e $c_m = 2m + \frac{1}{s} + \frac{1}{r} - (G - 1)$. A solução geral é

$$\begin{aligned} F(z, t) &= \sum_{m=0}^G T_m(t) Z_m(z) \\ &= \sum_{m=0}^G C_m \exp(\alpha_m t) z^m {}_2F_1(a_m, b_m; c_m; z) \end{aligned} \quad (2.29)$$

onde C_m é um coeficiente, e a_m , b_m , c_m e α_m como definidos antes.

Para encontrarmos a distribuição dos genótipos precisamos antes identificar $P_n(t)$ na equação 2.50 com os coeficientes de z ao expandirmos a equação 2.29 em série.

A expansão da função hipergeométrica é dada por

$${}_2F_1(a, b; c; z) = \sum_{l=0}^{\infty} \frac{(a)_l (b)_l}{(c)_l} \frac{z^l}{l!}, \quad (2.30)$$

com $(q)_n$ sendo o símbolo de Pochhammer definido por

$$\begin{cases} (q)_0 = 1 \\ (q)_l = q(q+1) \dots (q+l-1), l > 0. \end{cases}$$

Na equação 2.30, $(a)_l$ pode ser escrita como

$$\begin{aligned} (a)_l &= (m - G)_l \\ &= (-1)^l \frac{(G - m)!}{(G - m - l)!}. \end{aligned} \quad (2.31)$$

Note que para $l > G - m$, $(a)_l = 0$. A expressão para $F(z, t)$ torna-se

$$F(z, t) = \sum_{m=0}^G C_m \exp(\alpha_m t) \sum_{l=0}^{G-m} \frac{(G-m)!}{(G-m-l)! l!} \frac{(b)_l}{(c)_l} z^{l+m}. \quad (2.32)$$

Fazendo $n = l + m$, obtemos

$$F(z, t) = \sum_{m=0}^G \sum_{n=0}^G C_m \exp(\alpha_m t) \binom{G-m}{n-m} \frac{(b)_{n-m}}{(c)_{n-m}} H[n-m] z^n, \quad (2.33)$$

onde $H[n-m]$ é a função de Heaviside discreta

$$H[n-m] \begin{cases} 0, m > n \\ 1, m \leq n. \end{cases}$$

Assim, a equação 2.33 pode ser escrita como

$$F(z, t) = \sum_{n=0}^G \sum_{m=0}^n C_m \exp(\alpha_m t) \binom{G-m}{n-m} \frac{(b)_{n-m}}{(c)_{n-m}} z^n. \quad (2.34)$$

Identificando os termos da equação 2.34 com a definição da função geratriz 2.50, obtemos $P_n(t)$ como

$$P_n(t) = \sum_{m=0}^n C_m \exp(\alpha_m t) \binom{G-m}{n-m} \frac{(b)_{n-m}}{(c)_{n-m}}. \quad (2.35)$$

A expressão para as populações de genótipos é portanto encontrada. Como verificação, vamos examinar a expressão para $P_0(t)$.

$$\begin{aligned} P_0(t) &= C_0 \exp[(\alpha_0)t] \binom{G}{0} \frac{(b)_0}{(c)_0} \\ &= C_0 \exp[-Grt] \end{aligned} \quad (2.36)$$

Os coeficientes C_m são determinados da distribuição da população inicial $P_n(0) = \delta_{n0}$. Usando o fato que $P_0(0) = 1$, encontramos $C_0 = 1$. Logo, $P_0(t) = e^{-Grt}$. Esta expressão é a solução para a versão contínua da equação 2.4. Cada um dos outros coeficientes foi obtido adicionando-se sucessivamente 1 a $n = 0$, o que nos levou a discernir o seguinte padrão subjacente para os C_m :

$$C_m = \frac{G!}{m!(G-m)!} r^m \frac{\prod_{i=0}^{m-1} (1+is)}{\prod_{i=m}^{2m-1} \{r[(G-i)s-1]-s\}} \quad (2.37)$$

$$= \frac{(-1)^m}{s} \binom{G}{m} \frac{\left(1 + \frac{1}{s}\right)_{m-1}}{\left(m - G + \frac{1}{r} + \frac{1}{s}\right)_m}. \quad (2.38)$$

A partir da proposta de fórmula para C_m acima, a verificamos analiticamente e com o auxílio do CAS.

A distribuição exata da população no tempo $P_n(t)$ é dada por

$$P_n(t) = \sum_{m=0}^n \frac{(-1)^m}{s} \binom{G}{m} \binom{G-m}{n-m} \exp(\alpha_m t) \times \frac{\left(1 + \frac{1}{s}\right)_{m-1} \left(m + \frac{1}{s}\right)_{n-m}}{\left(m - G + \frac{1}{r} + \frac{1}{s}\right)_m \left(2m + 1 - G + \frac{1}{r} + \frac{1}{s}\right)_{n-m}}. \quad (2.39)$$

Ao usarmos a fórmula de Stirling no regime $\frac{1}{r} \gg \frac{1}{s} \gg G \geq 1$, obtemos

$$P_n(t) = e^{-Grt+1} \left(\frac{r}{s} + 1\right) \left(\frac{r}{s}\right)^n \binom{G}{n} \left(e^{(r+s)t} - 1\right)^n \quad (2.40)$$

$$\sim e^{[(G-n)r+ns]t}. \quad (2.41)$$

Os $P_n(t)$ são populações, não frequências. Para obtermos a frequência de cada um dos genótipos da população, é preciso normalizar os $P_n(t)$ pela soma deles no tempo t em todo o $0 \leq n \leq G$. Esta soma é o tamanho total da população $N(t)$, que é dado por

$$N(t) = e^{-Grt+1} \left(\frac{r}{s} + 1\right) \left[\frac{r}{s} \left(e^{(r+s)t} - 1\right) + 1\right]^G. \quad (2.42)$$

A frequência final é, portanto,

$$p_n(t) = \frac{\binom{G}{n} \left(\frac{r}{s}\right)^n \left(e^{(r+s)t} - 1\right)^n}{\left[\frac{r}{s} \left(e^{(r+s)t} - 1\right) + 1\right]^G}. \quad (2.43)$$

Esta frequência não é uma probabilidade, porque o modelo é determinístico, mas para fins de interpretação podemos entendê-la como se ela seguisse uma distribuição binomial $B(n, p)$ com probabilidade

$$p = \frac{\frac{r}{s} \left(e^{(r+s)t} - 1\right)}{\frac{r}{s} \left(e^{(r+s)t} - 1\right) + 1} \quad (2.44)$$

dependente do tempo. Isso implica que o comportamento de cada loco é independente. Note que p é a probabilidade do alelo selvagem se tornar mutante até o tempo t , portanto trata-se na realidade de uma função distribuição acumulada (CDF). A função densidade de probabilidade (PDF) f é a derivada do CDF.

$$f = \frac{r e^{(r+s)t}}{\left[\frac{r}{s} \left(e^{(r+s)t} - 1\right) + 1\right]^2}. \quad (2.45)$$

2.3 Resultados

Um tipo de alelo adaptativo

O modelo determinístico para adaptação de uma só linhagem com apenas um tipo de alelo adaptativo é o mais simples possível, mas é o ponto de partida para a solução

Tabela 1 – Parâmetros evolutivos. Os valores correspondem a medidas experimentais para *E. coli* (TADDEI et al., 1997), exceto para o número de locos.

Parâmetro	Variável	Valor
Taxa de mutação deletéria	d	1×10^{-4}
Taxa total de mutação adaptativa	r_T	1×10^{-8}
Vantagem do alelo adaptativo	s	0,005
Número de locos	G	8

do modelo com K tipos de alelos adaptativos e para modelagens posteriores. Ele serve como referência para avaliação do efeito da presença de mais de um alelo adaptativo. Do ponto de vista prático, a escolha de um coeficiente de seleção efetivo e de uma taxa de mutação adaptativa efetiva se mostraram adequados para a modelagem para descrever a evolução adaptativa de uma população de *Escherichia coli* (HEGRENESS et al., 2006). Esta modelagem pode ser adequada para a descrição da evolução de populações assexuadas em geral.

A solução obtida modela mutações adaptativas nos genes selvagens. As mutações deletérias não são levadas em conta porque elas subtraem uma fração igual de indivíduos de cada um dos genótipos. Elas não correspondem a mutações reversas às adaptativas: as mutações deletérias acontecem em todo o genoma e conjuntamente levam a uma redução da aptidão igual à sua taxa de mutação total. Entendemos assim porque o coeficiente de seleção de uma mutação deletéria é baixo o suficiente para levar seus portadores à extinção em poucas gerações. É como se tomássemos as mutações deletérias como letais. Apesar das taxas de mutações deletérias serem cerca de apenas cem vezes menores que o coeficiente de seleção, violando o regime de validade da aproximação, verificaremos se o modelo também pode de alguma maneira descrever este tipo de mutação, limitada a um certo número de sítios e com um coeficiente de seleção próprio. Compararemos nossa solução com a solução numérica para estes dois casos, a fim de verificarmos se a fórmula para a distribuição de genótipos está adequada.

A Figura 4 mostra a comparação da dinâmica das frequências genotípicas quando as mutações são adaptativas. Podemos perceber que para os parâmetros usados (ver Tabela 1), próximos às condições de evolução experimental de *E. coli* (TADDEI et al., 1997), as distribuições genotípicas apresentam grande semelhança quantitativa entre si. Note que o estado estacionário é desinteressante: praticamente toda a população encontra-se totalmente adaptada. Estas soluções não incluem as taxas de mutação deletérias porque elas removem cada genótipo de uma mesma fração, e portanto o resultado seria o mesmo que o atual após a normalização.

Na Figura 5 comparamos quantitativamente a solução analítica da dinâmica com uma via adaptativa com a solução numérica. Pelo mesmo motivo que da Figura anterior,

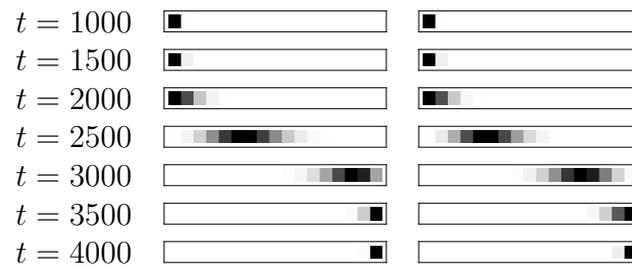


Figura 4 – Uma via adaptativa. Comparação entre as soluções numérica em tempo discreto (esquerda) e analítica (direita) para as distribuições genotípicas de um único conjunto de genes, com $G = 16$, $s = 0,005$ and $r = 10^{-8}$ em diferentes gerações t . O eixo x corresponde ao número de mutações acumulado. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante.

não incluímos mutações deletérias. Percebemos uma ótima concordância tanto em relação à média do número de alelos adaptativos quanto em relação à sua variância. Concluimos que a solução analítica para esta dinâmica é adequada quantitativamente.

No caso de mutação deletéria, esperamos que para mutabilidades suficientemente baixas a linhagem forme quase-espécies, ou seja, consista em uma distribuição de genótipos em torno do genótipo mais adaptado. Para mutabilidades suficientemente altas, a distribuição fica centrada em torno de genótipos mutantes. E para mutabilidade muito alta, a população fica centrada no genótipo menos apto, sugerindo a catástrofe de erros, isto é, a dispersão da população no espaço genotípico. Por isso, ao invés de estudarmos a dinâmica comparamos os estados estacionários oriundos desta evolução obtidos de nossa solução e da iteração numérica. A Figura 6 mostra concordância qualitativa entre os métodos.

Vamos agora examinar a validade da fórmula para o número médio de alelos deletérios ($s < 0$). Na Figura 7 vemos claramente o regime de validade para nossa solução analítica neste caso: para $r \ll s$, tal qual imposta na aplicação da aproximação de Stirling. A concordância quantitativa nesta região é boa, mas à medida que $r \rightarrow |s|$, $r + s \rightarrow 0$ e o termo da exponencial menos um vai a zero. Nesse regime a solução obtida parece ter uma descontinuidade. Isso acontece porque a exponencial troca de sinal ao r superar s , o que leva a comportamentos distintos se r se aproxima de s pela direita ou pela esquerda. De qualquer forma, a solução ainda tem um regime de validade considerável, e surpreendentemente, o valor assintótico do número médio de alelos quando $r \rightarrow \infty$ é o correto.

Esperamos destas verificações que, para regimes adequados, nossa solução gere tanto a dinâmica genotípica como o estado estacionário corretos. Passaremos agora a estudar uma quantidade evolutiva importante obtida a partir das distribuições genotípicas: a aptidão ao longo do tempo.

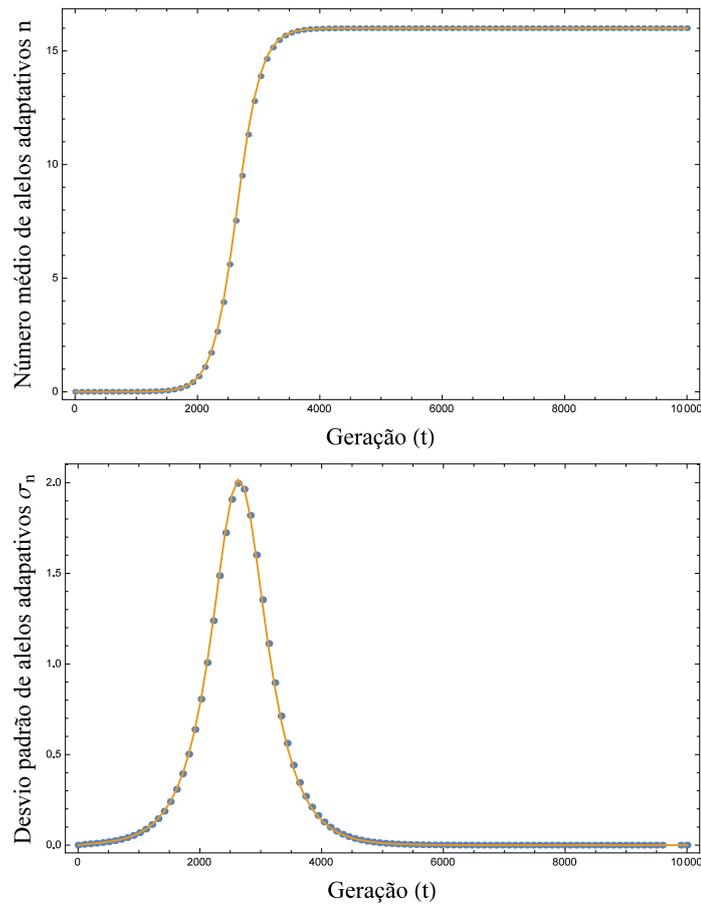


Figura 5 – Verificação da solução para uma via adaptativa. Comparação entre as soluções numérica em tempo discreto (linha contínua) e analítica (círculo) para as distribuições genóticas de um único conjunto de genes, com $G = 16$, $s = 0,005$ and $r = 10^{-8}$ em diferentes gerações t . Acima: média da distribuição de alelos adaptativos no tempo; abaixo: desvio padrão da distribuição de alelos adaptativos no tempo.

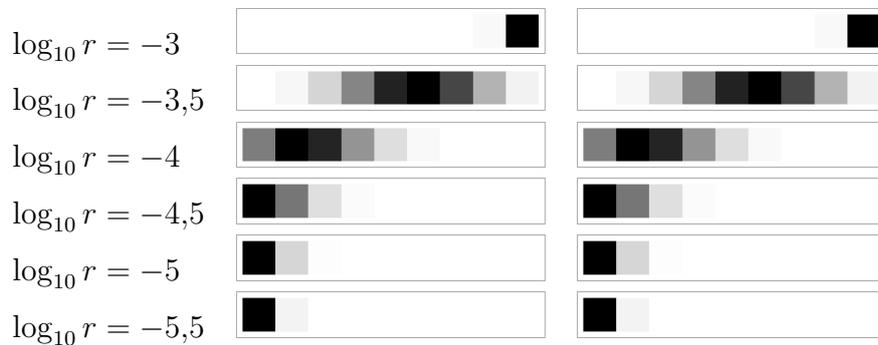


Figura 6 – Uma via deletéria. Comparação entre a distribuição numérica em tempo discreto (esquerda) e analítica (direita) para tempos longos, de um único conjunto de genes em $t = 10000$, para $G = 8$, $s = -0,0005$ e r indicado na primeira coluna. O eixo x corresponde ao número de mutações acumulado. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante.

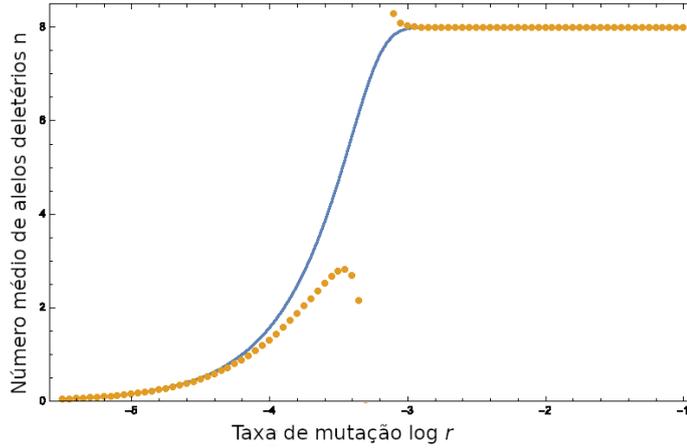


Figura 7 – Verificação da solução de quase-espécies para uma via deletéria. Comparação entre as soluções numérica em tempo discreto (linha contínua) e analítica (círculo) para as médias das distribuições genótípicas de um único conjunto de genes, com $s = -0,0005$ com diferentes taxas de mutação r .

Aptidão

A dinâmica da aptidão pode ser calculada em termos do coeficiente de seleção médio $\langle s \rangle$ como a média de $s_n = ns$ sobre todos os genótipos n ponderada pela frequência do genótipo, p_n . O resultado para uma linhagem mutadora genérica é

$$\begin{aligned} \langle s \rangle &= Gs \left(1 - \frac{s}{r(e^{(r+s)t} - 1) + s} \right) \\ &= Gsp, \end{aligned} \quad (2.46)$$

usando a condição $r \ll s$.

O tempo para a derivada máxima para mutações adaptativas ($s > 0$) é dado por $t_{\max} = \ln(s/r)/s$. Se aproximarmos os platôs como retas horizontais, e a zona de crescimento por uma reta que tem inclinação máxima da curva no ponto em que isto acontece, vemos que o intervalo de tempo para o crescimento da aptidão, que pode ser tomado aproximadamente como início da adaptação até o tempo para a derivada máxima é $2/s$, o mesmo tempo entre a derivada máxima e o fim da adaptação. O tempo de adaptação para população infinita neste caso é $\approx 4/s$.

K alelos adaptativos

Agora vamos modelar a evolução em que mais de um alelo adaptativo é permitido, mas apenas um mutante por loco. Além de tornar a modelagem mais realista, esta abordagem permite explorar de forma alternativa se existe uma dependência entre taxa de mutação e a ordem com que as mutações adaptativas ocorrem.

O modelo

O modelo para K alelos é análogo ao de um alelo. Primeiramente, admitimos que cada geração é composta por um passo de crescimento e outro de transições. Definimos uma classe como o conjunto de genótipos que têm o mesmo número de mutações para cada tipo de mutação. As equações para estes passos no caso de K alelos é

$$P_{\mathbf{n}}(t + 1/2) = w_{\mathbf{n}}P_{\mathbf{n}}(t) \quad (2.47)$$

$$P_{\mathbf{n}}(t + 1) = P_{\mathbf{n}}(t + 1/2) + \sum_{j=1}^K [G_j - (n_j - 1)]r_j P_{\mathbf{n}-\mathbf{e}_j}(t + 1/2) - (G_j - n_j)r_j P_{\mathbf{n}}(t + 1/2). \quad (2.48)$$

onde $P_{\mathbf{n}}$ é abundância da classe \mathbf{n} , $w_{\mathbf{n}}$ é a aptidão desta classe, e $P_{\mathbf{n}-\mathbf{e}_j}$ corresponde à classe \mathbf{n} com $n_j - 1$ alelos do tipo j (\mathbf{e}_j é o j -ésimo vetor unitário da base canônica). Esses dois passos podem ser agrupados em um único

$$P_{\mathbf{n}}(t + 1) = P_{\mathbf{n}}(t) + \sum_{i=1}^K n_i s_i P_{\mathbf{n}}(t) + \sum_{j=1}^K [1 + (n_j - 1)s_j] \times [G_j - (n_j - 1)]r_j P_{\mathbf{n},n_j-1}(t) - \sum_{i=1}^K (1 + n_i s_i)(G_i - n_i)r_i P_{\mathbf{n}}(t). \quad (2.49)$$

Após realizarmos a aproximação $P_{\mathbf{n}}(t+1) - P_{\mathbf{n}}(t) \approx \dot{P}_{\mathbf{n}}(t)$ a equação 2.49 pode ser resolvida com a ajuda da respectiva função geratriz, mostrada abaixo.

$$F(\mathbf{z}, t) = \prod_{i=1}^K \sum_{n_i=0}^{G_i} z_i^{n_i} P_{\mathbf{n}}(t). \quad (2.50)$$

A equação diferencial resultante é resolvida por separação de variáveis fazendo $F(\mathbf{z}, t) = T(t) \prod_{i=1}^K Z_i(z)$. Cada solução Z_i equivale a uma solução do tipo $Z(z)$ para um único alelo adaptativo. A solução final para a abundância do genótipo $P_{\mathbf{n}}$ é

$$P_{\mathbf{n}}(t) = \prod_{i=1}^K e^{-G_i r_i t + 1} \left(\frac{r_i}{s_i} + 1 \right) \left(\frac{r_i}{s_i} \right)^{n_i} \binom{G_i}{n_i} (e^{s_i t} - 1)_i^{n_i}.$$

e para a frequência genotípica é

$$p_{\mathbf{n}}(t) = \prod_{i=1}^K \frac{\left(\frac{r_i}{s_i} \right)^{n_i} \binom{G_i}{n_i} (e^{s_i t} - 1)^{n_i}}{\left[\frac{r_i}{s_i} (e^{s_i t} - 1) + 1 \right]^{G_i}}.$$

A solução é o produto das curvas dos coeficientes de seleção médios de cada classe de locos.

Verificação

Nos limitaremos a examinar aqui o caso $K = 2$. Primeiramente comparamos a dinâmica da distribuição genotípica quando os dois alelos são adaptativos obtida de nossa

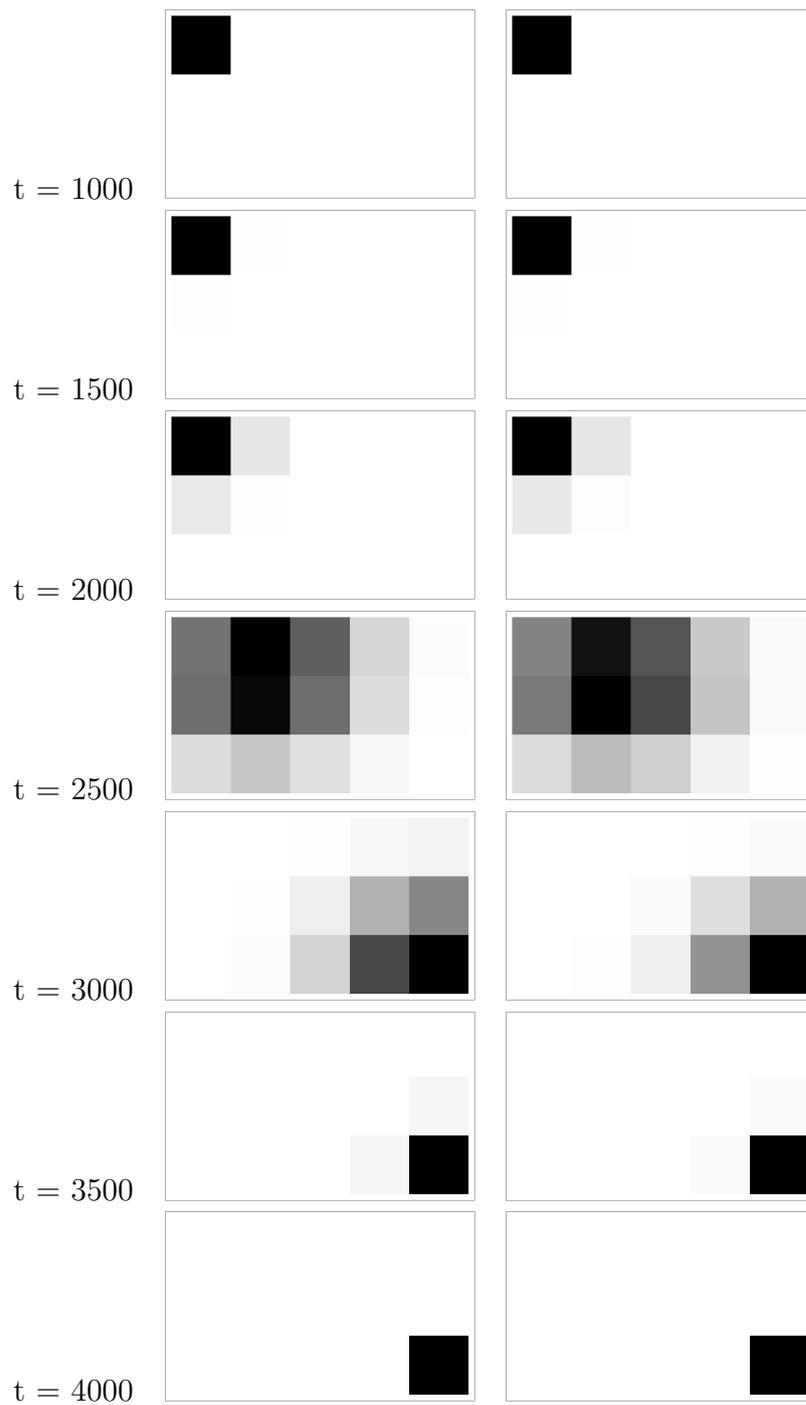


Figura 8 – Duas vias adaptativas. Comparação entre a dinâmica da distribuição numérica em tempo discreto (esquerda) e analítica (direita) de dois conjuntos de genes, com $s_1 = 0,005$, $r_1 = 10^{-8}$, $G_1 = 2$, $s_2 = 0,00595$, $r_2 = 10^{-9}$ e $G_2 = 4$. O eixo y corresponde ao número de mutações do tipo 1 acumulado, e o eixo x o do tipo 2. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante.

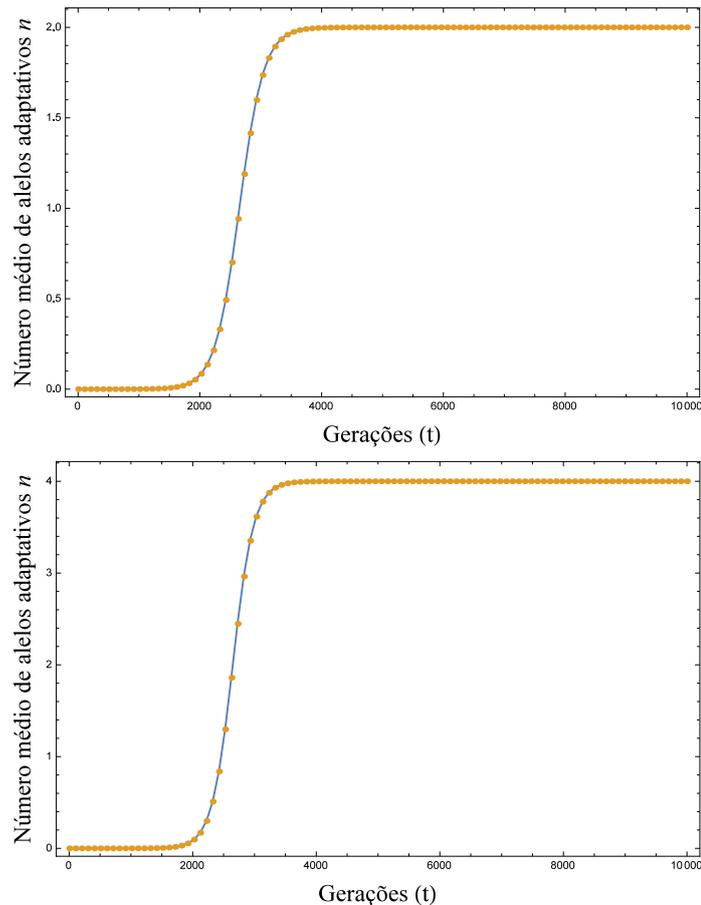


Figura 9 – Número médio de alelos adaptativos com evolução em duas vias adaptativas. Comparação entre a dinâmica da distribuição numérica em tempo discreto (linha contínua) e analítica (círculo) de dois conjuntos de genes, com $s_1 = 0,005$, $r_1 = 10^{-8}$, $G_1 = 2$, $s_2 = 0,00595$, $r_2 = 10^{-9}$ e $G_2 = 4$. Acima: número médio de alelos do tipo 1; abaixo: número médio de alelos do tipo 2.

solução e da iteração numérica. O resultado é mostrado na Figura 8, onde percebemos uma semelhança qualitativa entre nossa solução e a numérica.

Quantitativamente falando, como pode ser visto na Figura 9 a concordância do número médio de cada tipo de alelo adaptativo ao longo do tempo é muito boa. Concluimos portanto que a solução é adequada para descrever a dinâmica de adaptação.

Nas próximas comparações desta subseção verificaremos se há concordância entre as quase-espécies obtidas da distribuição estacionária de genótipos pelos dois métodos quando pelo menos um dos alelos é deletério.

Na Figura 10, examinamos a formação de quase-espécies quando ambos os alelos são deletérios. Como esperado, quanto menor o coeficiente de seleção, mais concentrada torna-se a distribuição em torno do estado selvagem, porque genótipos mais deletérios são mais rapidamente expurgados. Vemos também aqui que ambas as distribuições convergem para estados estacionários muito semelhantes, o que valida nossa solução.

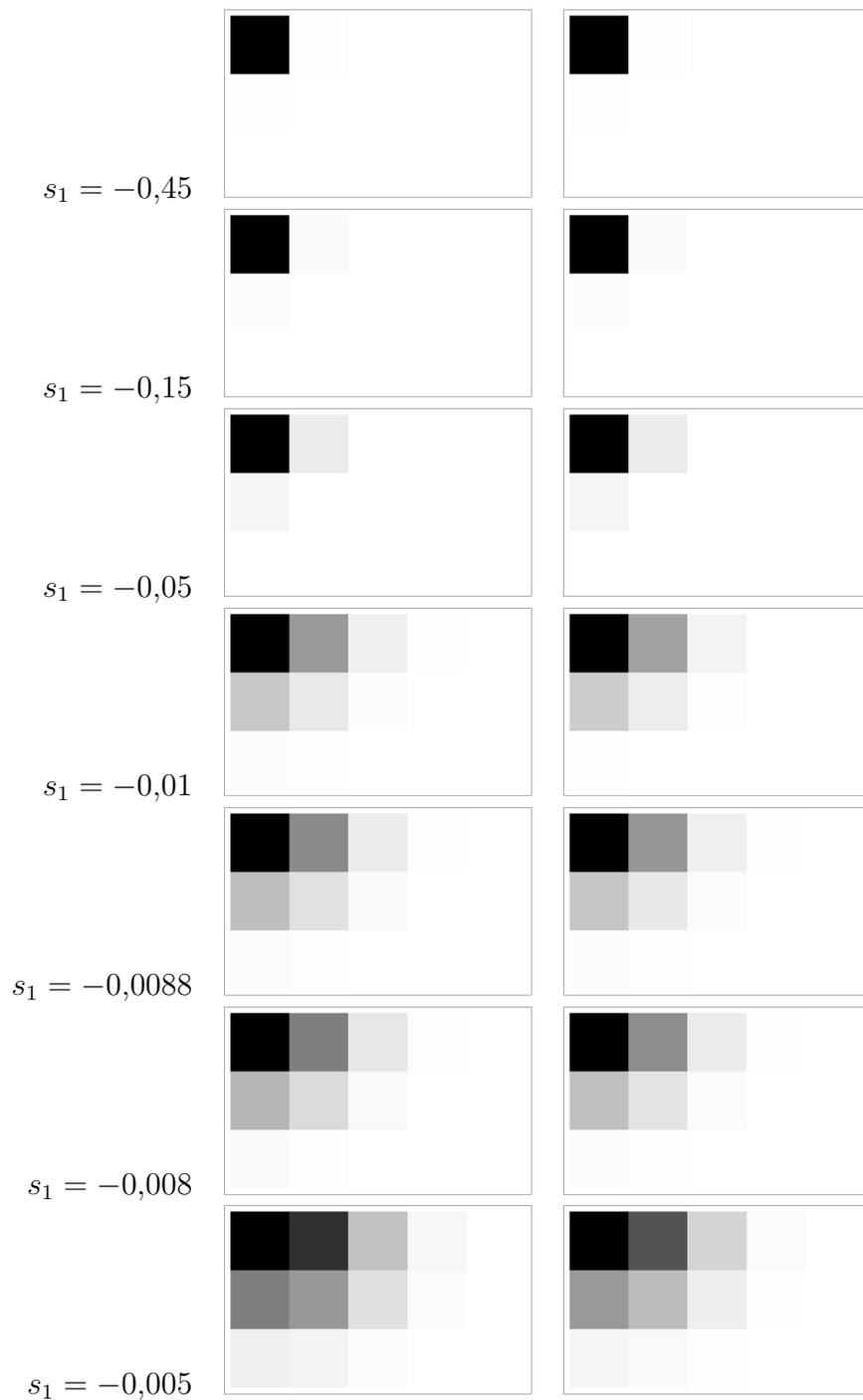


Figura 10 – Quase-espécies com duas vias deletérias. Comparação entre o estado estacionário da distribuição numérica em tempo discreto (esquerda) e analítica (direita) de dois conjuntos de genes, com $r_1 = r_2 = 10^{-3}$, s_1 mostrado na primeira coluna, $s_2 = s_1 - 0,0009$, $G_1 = 2$, $G_2 = 4$ e tempo $t = 10000$. O eixo y corresponde ao número de mutações do tipo 1 acumulado, e o eixo x o do tipo 2. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante.

A concordância qualitativa entre a solução analítica e a numérica para o processo deletério com duas vias deletérias é muito boa, como pode se ver na Figura 11. Isso ocorre porque o regime de validade da solução é respeitado aqui. Assim concluímos que a solução obtida é boa tanto para o processo adaptativo como o deletério puros desde que respeitados os limites de validade da solução analítica.

Quando uma das vias é adaptativa e a outra é deletéria (Figura 12) a distribuição se concentra em torno do estado mais adaptado quanto menor o coeficiente de seleção do alelo deletério, conforme esperado. Em particular, $s = -0,05$ pode ser considerado um coeficiente de seleção típico para alelos deletérios em *E. coli*, e para este valor uma fração pequena dos genótipos possui um alelo deletério.

Mais ainda, a concordância qualitativa da distribuição estacionária de nossa solução com a solução numérica atesta a qualidade da primeira.

De uma perspectiva mais quantitativa, a concordância entre a teoria e a solução numérica é satisfatória. A Figura 13 nos informa que quando combinamos uma via adaptativa com outra deletéria, o número médio de alelos-alvo de cada tipo ao longo do tempo é descrito corretamente por nossa solução analítica. Isso sugere que quaisquer combinações de processos adaptativos podem ser quantitativamente descritas por nossa solução analítica.

Aptidão

Como a solução da distribuição genotípica para K alelos é o produto de soluções para um alelo, a dinâmica do coeficiente de seleção médio será a soma da expressão da mesma dinâmica para apenas um alelo adaptativo, isto é,

$$\begin{aligned} \langle s \rangle &= \sum_{i=1}^K G_i s_i \left(1 - \frac{s_i}{r_i (e^{(r_i+s_i)t} - 1) + s_i} \right) \\ &= \sum_{i=1}^K G_i s_i p_i. \end{aligned} \quad (2.51)$$

Naturalmente, a curva da dinâmica deste coeficiente de seleção será a sobreposição das curvas de cada coeficiente de seleção médio.

O tempo de fixação de cada conjunto de loci é obtido da expressão $\log(s_i/r_i)/s_i$ para $s > 0$ e $r \ll s$. Esta expressão nos sugeriu que, uma vez estabelecida uma relação entre a taxa de mutação e o coeficiente de seleção, a mutabilidade (um fator comum que multiplica todas as r_i) poderia interferir na ordem dos tempos de fixação de diferentes tipos de alelos. Isso será investigado daqui a poucos capítulos.

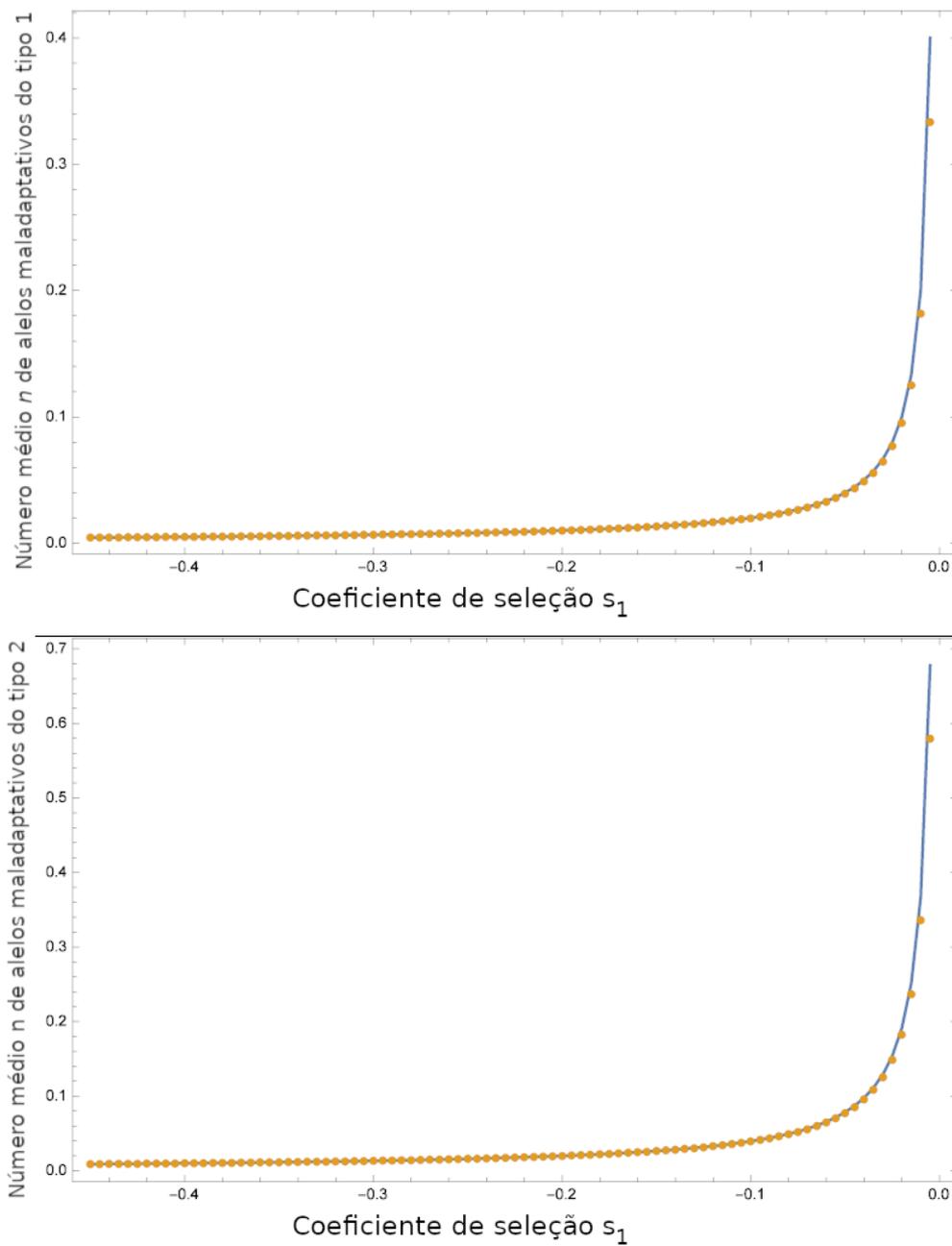


Figura 11 – Número médio de alelos com evolução em duas vias deletérias. Comparação entre o número médio de alelos deletérios da distribuição numérica em tempo discreto (linha contínua) e analítica (círculo) de dois conjuntos de genes, variando-se o coeficiente de seleção da primeira via s_1 , enquanto $s_2 = s_1 - 0,0009$. Os demais parâmetros são $r_1 = 10^{-8}$, $G_1 = 2$, $s_2 = 0,00595$, $r_2 = 10^{-9}$ e $G_2 = 4$. Acima: número médio de alelos do tipo 1; abaixo: número médio de alelos do tipo 2.

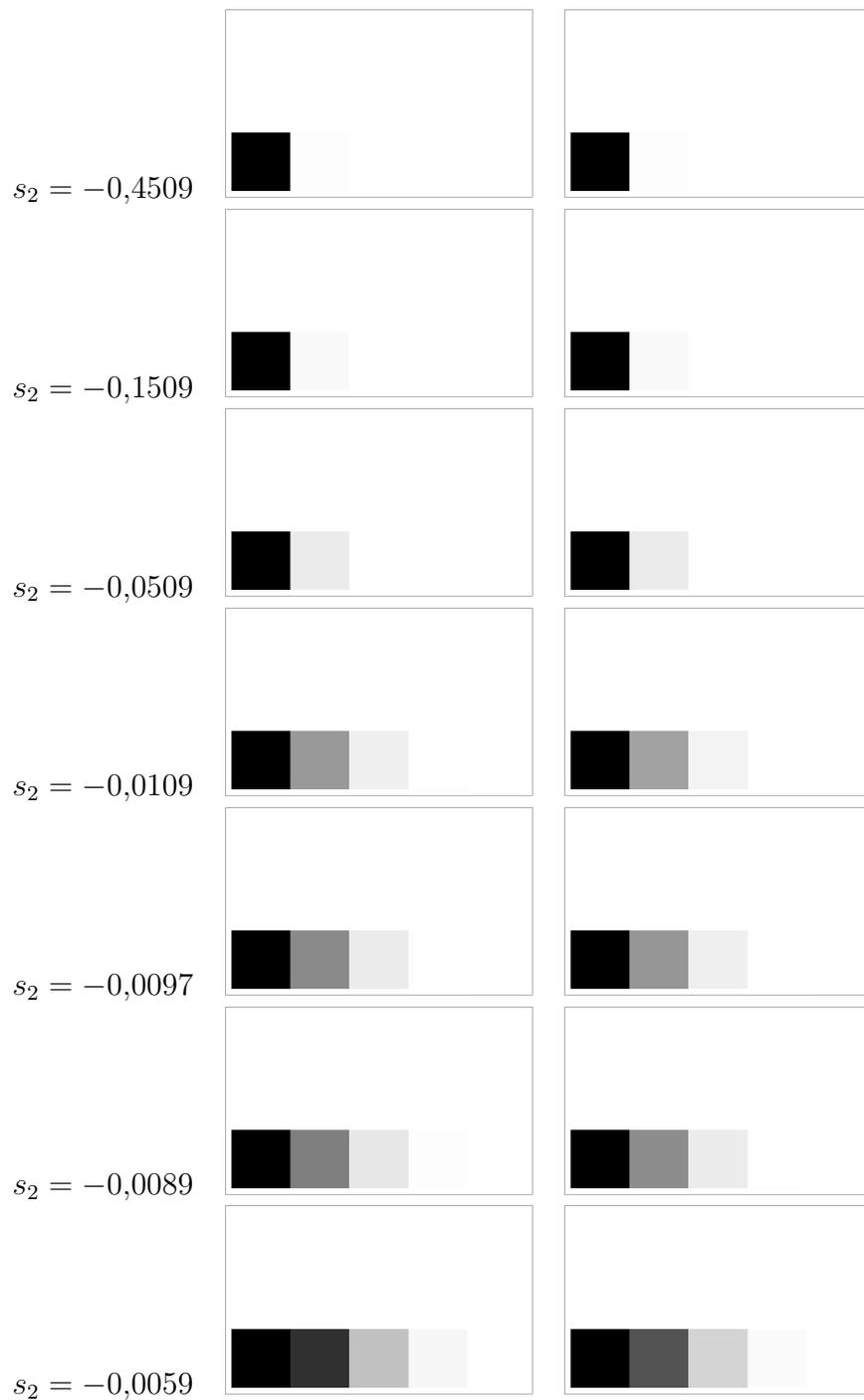


Figura 12 – Quase-espécies com duas vias opostas. Comparação entre o estado estacionário da distribuição numérica em tempo discreto (esquerda) e analítica (direita) de dois conjuntos de genes, com $r_1 = r_2 = 10^{-3}$, $s_1 = 0,005$ (adaptativo), s_2 (deletério) mostrado à esquerda dos gráficos, $G_1 = 2$, $G_2 = 4$ e tempo $t = 10000$. O eixo y corresponde ao número de mutações do tipo 1 acumulado, e o eixo x o do tipo 2. Na escala de cinza dos diagramas, o branco representa frequência zero, e o preto representa a frequência máxima no respectivo instante.

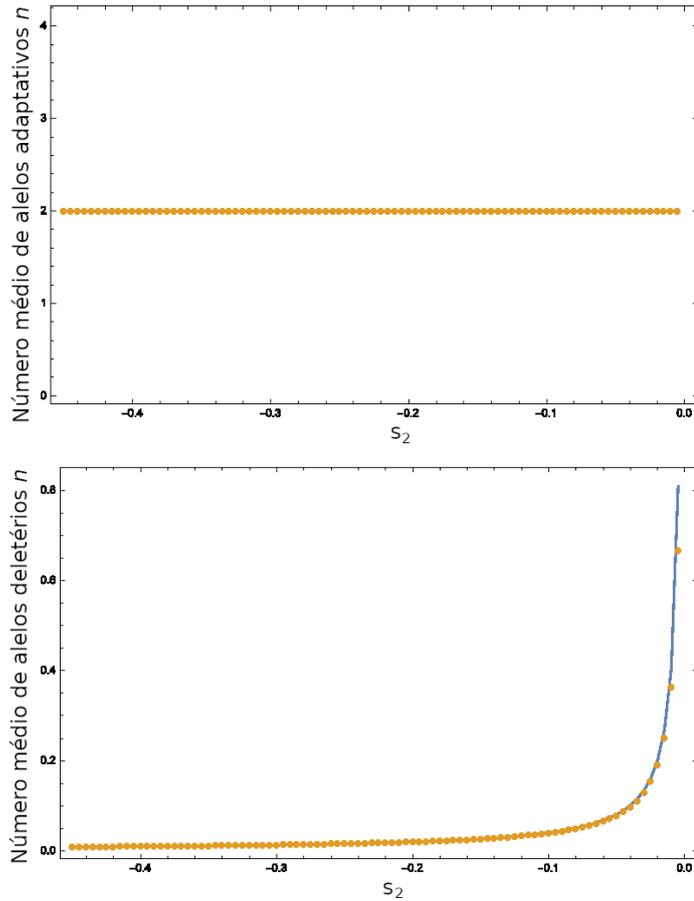


Figura 13 – Número médio de alelos de cada tipo no estado estacionário com evolução em uma via adaptativa e outra deletéria. Comparação entre o número médio de alelos deletérios da distribuição numérica em tempo discreto (linha contínua) e analítica (círculo) de dois conjuntos de genes, variando-se o coeficiente de seleção da segunda via s_2 . Os demais parâmetros são $s_1 = 0,005$, $r_1 = 10^{-8}$, $G_1 = 2$, $r_2 = 10^{-9}$ e $G_2 = 4$. Acima: número médio de alelos do tipo 1; abaixo: número médio de alelos do tipo 2.

2.4 Discussão

Vimos em Métodos que a frequência de um alelo com taxa r e coeficiente de seleção s é proporcional a

$$p \propto \frac{r}{s} \left(e^{(r+s)t} - 1 \right), \quad (2.52)$$

se a condição inicial da população é ser toda selvagem. Para mutações adaptativas ($s > 0$) vemos aqui que naturalmente a frequência do alelo mutante quando não normalizada cresce com o fator $e^{(r+s)t}$ que corresponde a um crescimento em tempo contínuo. Em uma geração $t = 1$ e o crescimento correspondente passa a ser de $e^{r+s} \approx 1 + s$ para $r \ll s$. Substituindo a exponencial pela aproximação, vemos que o resultado para a primeira geração é r , isto é, a única contribuição para um alelo mutante na primeira geração é a da taxa de mutação.

Pudemos generalizar a expressão de mutações adaptativas para deletérias porque

não nos utilizamos de fato da condição $s > 0$ na dedução. Se olharmos a Equação 2.44, vemos que ela também funciona para $s < 0$, uma vez que $r/s < 0$ e $e^{st} - 1 < 0$, tendo como consequência um produto e uma frequência positiva. No limite de $t \rightarrow \infty$, temos que a Equação 2.43 torna-se

$$p_n = \frac{\binom{G}{n} \left(\frac{r}{|s|}\right)^n}{\left(\frac{r}{|s|} + 1\right)^G}. \quad (2.53)$$

Esta é a distribuição estacionária para alelos deletérios. Note que, como $r \ll |s|$ (para mutações deletérias um r típico é de 10^{-4} e $s = -0,05$), espera-se que p_n caia rapidamente com o número de alelos deletérios. Nota-se aqui uma diferença para o caso estocástico, que tem o estado menos apto como estado absorvente da dinâmica. A solução estacionária para população infinita pode refletir um transiente da dinâmica estocástica para tempos longos, mas não infinitamente longos. A população selvagem no regime $N \rightarrow \infty$ tem probabilidade

$$p_0 = \frac{1}{\left(1 + \frac{r}{|s|}\right)^G}, \quad (2.54)$$

que após uma expansão binomial, obtemos

$$p_0 \approx 1 - \frac{Gr}{|s|}. \quad (2.55)$$

Quanto maior $|s|$, e quanto menor o produto Gr , menos a população selvagem perde indivíduos para estados deletérios. O produto Gr é a taxa de mutação que depleta o estado selvagem, e quanto maior $|s|$, menor é a contribuição dos estados deletérios para a população, e maior a frequência do estado selvagem. Para mutações adaptativas, o coeficiente de seleção médio para tempos infinitos é $\langle s \rangle = Gs$. Para mutações deletérias, o negativo deste mesmo coeficiente equivale à carga genética L , e é dada por

$$L = Gr \frac{s}{s - r}. \quad (2.56)$$

Se considerarmos $|s| \gg r$, $L \approx Gr$, a taxa de mutação para fora do estado selvagem (o mais apto), o que é o resultado clássico para haplóides sob mutação deletéria.

A comparação de $N \rightarrow \infty$ com realizações estocásticas aponta que, para $s > 0$, mesmo com taxas de mutação adaptativas por sítio altas e população de 10^9 existem diferenças substanciais (gráficos não mostrados). Uma das diferenças é que a dinâmica estocástica exhibe platôs intermediários, e outra é que uma medida do tempo para a população se tornar totalmente adaptada é muito mais longa que o caso determinístico.

Para uma aproximação razoável entre os dois casos, talvez mesmo populações de tamanhos irrazoavelmente grandes não sejam suficientes. Para $N = 10^6$ e $N = 10^9$ e $G = 1$ as curvas da dinâmica do coeficiente de seleção coincidem bastante e a fixação continua ocorrendo bem longe do tempo previsto para N infinito.

Isso deve ser causado pela fase inicial onde a estocasticidade demográfica alonga bastante os tempos inicial e final da fixação. Isso se agrava para G maiores e populações onde a chance de dupla mutação em um indivíduo na mesma geração é baixa, porque a fixação é sequencial e existem várias etapas para a estocasticidade demográfica atuar. Quando a população é grande o suficiente, a fixação deixa de depender de N , dependendo apenas do coeficiente de seleção.

John e Jain (JOHN; JAIN, 2015) chegaram a uma solução em que há mutações diretas e reversas, admitindo um *ansatz* de Woodcock e Higgs (WOODCOCK; HIGGS, 1996) de que os locos são independentes. Elas não resolveram diretamente a solução mestra para todos os locos. Bastou então para John e Jain solucionar as equações para as frequências de dois alelos em um único loco que sofrem mutação de um para outro com taxas distintas. Por admitir de antemão a independência dos locos, sua solução pode não ser exata. Verificamos que sua solução se reduz à nossa solução simplificada no regime $r \ll |s| \ll 1$ com a taxa de mutação reversa nula.

Esse pode ser o regime em que a independência entre os locos pode ser admitida. Tal possibilidade, observada independentemente por nós, abre caminho para que resolvam equações mais complexas para cada um dos locos envolvendo dois ou mais alelos ou sequências de alelos, podendo-se combinar estas soluções em uma expressão ainda mais geral.

2.5 Conclusão

Neste capítulo resolvemos um processo adaptativo simples com mutações iguais, e o generalizamos para quando há diferentes conjuntos de mutações, adaptativas ou deletérias. Para mutações deletérias apenas, obtivemos a distribuição de quase-espécies e calculamos a carga genética deste processo em um dos casos.

Da solução aproximada da dinâmica da aptidão no tempo, obtivemos uma expressão para o tempo de fixação de cada tipo de mutação, que nos sugere que, com o estabelecimento de uma relação entre aptidão e taxa de uma mutação, a mutabilidade pode alterar a ordem em que as mutações acontecem no ancestral que se fixa. Não apenas isto, mas dependendo da rede de trajetórias mutacionais, ela pode alterar a probabilidade dos destinos finais da evolução.

Os resultados de população infinita neste e em outros capítulos, que em geral baseiam-se no conteúdo agora apresentado, servem como um ponto de partida para se entender fenômenos em populações finitas, e como um controle para a compreensão dos fenômenos estocásticos.

3 Mutabilidade e passo evolutivo

3.1 Introdução

Estudaremos aqui um dos sistemas candidatos mais simples de mutação, seleção e deriva em que a mutabilidade pode influenciar o destino evolutivo. O sistema é uma população selvagem que pode sofrer uma mutação mais adaptativa e mais rara, ou outra, menos vantajosa e com maior aptidão.

Aqui aparece a tensão entre taxa de mutação e coeficiente de seleção que antecipamos da fórmula $t_{\max} = \log(s/r)/s$ para o tempo de adaptação em população infinita, onde s é o coeficiente de seleção de uma mutação adaptativa, e r é a sua taxa de mutação. Previmos nesta fórmula que uma relação entre r e s pode levar a uma mutação a se fixar antes de outra, dependendo da mutabilidade.

Inicialmente resolveremos o sistema de três tipos para população infinita, e concluiremos que embora para tempos infinitos o tipo mais apto sempre vence, para tempos iniciais se vê uma inversão da prevalência do tipo A para o tipo B .

Em seguida estudamos a fundo o mesmo sistema mas com população finita e evoluindo sob um esquema de Wright-Fisher. Vários aspectos dinâmicos e finais desta evolução são compreendidas, especialmente à luz de dois regimes simples: onde a mutação é extremamente rara, e onde a mutação é extremamente comum.

Assim, nos tornamos capazes de entender as condições e o porquê a mutabilidade é capaz de inverter o tipo para o qual a população é mais absorvida no processo.

3.2 Métodos

Estudamos o modelo diagramado na Figura 14. Note que não representamos o processo de imigração, a ser brevemente estudado com populações finitas. A mutabilidade é medida como a quantidade r , que é um fator multiplicativo comum entre todas as taxas de mutação.

O sistema de equações que descreve o modelo apresentado acima, para população infinita, é:

$$\begin{aligned}\dot{P}_W &= -(r_A + r_B)P_W, \\ \dot{P}_A &= s_A P_A + r_A P_W, \\ \dot{P}_B &= s_B P_B + r_B P_W.\end{aligned}\tag{3.1}$$

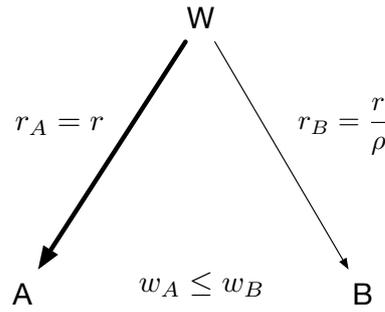


Figura 14 – Modelo para mutações em um único loco. O tipo selvagem muta para o tipo A , que possui aptidão $w_A = 1 + s_A$ menor que a aptidão $w_B = 1 + s_B$ do mutante alternativo B . No entanto, a taxa de mutação do tipo selvagem para o tipo B é ρ vezes menor que a taxa para o tipo A , com $\rho > 1$. A taxa r é um fator comum a ambas as mutações, e reflete a mutabilidade da população.

Os pesos dos tipos W , A e B , a serem normalizados, são P_W , P_A e P_B , respectivamente. Os coeficientes de seleção s_A e s_B correspondem aos mutantes A e B , com $s_B > s_A > 0$. As taxas de mutação do tipo selvagem para os tipos A e B são, respectivamente, r_A e r_B , com $r_A > r_B > 0$. As condições iniciais são $P_W(0) = 1$, $P_A(0) = 0$ e $P_B(0) = 0$.

A primeira equação é desacoplada das demais, e sua solução, facilmente obtida, é:

$$P_W(t) = e^{-(r_A+r_B)t}. \quad (3.2)$$

Substituindo-se esta solução no sistema de equações, obtemos duas equações diferenciais lineares de primeira ordem não-homogêneas desacopladas, cuja solução é facilmente obtida determinando-se as soluções homogênea e particular, somando-as e aplicando as condições de contorno:

$$\begin{aligned} P_A(t) &= \frac{r_A}{r_A + r_B + s_A} e^{s_A t} \left(1 - e^{-(r_A+r_B+s_A)t}\right), \\ P_B(t) &= \frac{r_B}{r_A + r_B + s_B} e^{s_B t} \left(1 - e^{-(r_A+r_B+s_B)t}\right). \end{aligned} \quad (3.3)$$

Usando a aproximação $r_A, r_B \ll s_A, s_B$, obtemos

$$P_A(t) = \frac{r_A}{s_A} \left(e^{s_A t} - 1\right), \quad (3.4)$$

$$P_B(t) = \frac{r_B}{s_B} \left(e^{s_B t} - 1\right). \quad (3.5)$$

Se tomarmos o limite $t \rightarrow \infty$ para as equações 3.2, 3.4, e 3.5, vemos que, do ponto de vista da frequência na população, o tipo B é o único que resta. Em relação à dinâmica, vemos que este tipo tem um coeficiente menor, o que lhe deve dar um início desvantajoso, mas seu crescimento exponencial com maior taxa enfim se impõe. Isso será melhor apresentado em Resultados. Na investigação estocástica, incluímos o efeito da

emigração de mutantes e imigração de indivíduos selvagens para uma das análises. Usamos um esquema de Wright-Fisher generalizado, onde o peso de amostragem de cada tipo é dada por

$$\begin{aligned} P_W^{t+1} &= N_W^t(1 - r_A - r_B) + (w_A N_A^t + w_B N_B^t)f, \\ P_A^{t+1} &= (1 - f)w_A N_A^t + N_W^t r_A \\ P_B^{t+1} &= (1 - f)w_B N_B^t + N_W^t r_B. \end{aligned} \quad (3.6)$$

O número de indivíduos dos tipos W , A e B , são respectivamente N_W , N_A e N_B , onde o tamanho total da população $N = N_W + N_A + N_B$ é constante. Os parâmetros $w_A = 1 + s_A$ e $w_B = 1 + s_B$ são as aptidões dos tipos A e B . A aptidão do tipo W é $w_W = 1$. E o parâmetro f é a fração de indivíduos mutantes que emigram, e que é completada por indivíduos selvagens de uma outra população implícita. Na maior parte dos casos estudados não há migração, portanto $f = 0$.

Na ausência de migração, a Figura 14 nos indica que os tipos A e B são estados absorventes da população. Mas neste processo é possível que a população se torne inteiramente selvagem muitas vezes antes que ocorra absorção em um dos estados com os mutantes.

Neste trabalho, nos referimos como fixação a primeira vez que a população se torna homogênea depois do primeiro mutante surgir, com ou sem mais mutantes ao mesmo tempo. Os tipos W , A e B podem se fixar com probabilidade π_f^W , π_f^A e π_f^B , respectivamente, e a probabilidade de absorção destes mesmos tipos é π_a^W , π_a^A e π_a^B , respectivamente. Note que $\pi_a^W = 0$, e $\pi_a^A + \pi_a^B = 1$.

Além das equações diferenciais, utilizamos também simulações do processo markoviano correspondente (N_W, N_A, N_B) de populações finitas. A amostragem foi feita conforme o processo de Wright-Fisher com mutação e seleção, para o sistema a ser estudado.

Cada realização começa com $N_W = N$ indivíduos, com a probabilidade de cada tipo aparecer na geração seguinte proporcional aos pesos não-normalizados da Equação 3.6. A amostragem é feita com reposição, seguindo uma distribuição multinomial. A simulação acaba quando um dos estados absorventes é alcançado (todo- A ou todo- B). Foram realizadas 2^{21} simulações por condição, de onde foram calculados a média e o desvio padrão. O tamanho da população variou de poucos indivíduos até 10^9 indivíduos. O algoritmo de amostragem é o “gsl_ran_multinomial” da biblioteca *GNU Scientific Library*.

Podemos mostrar a partir da dedução da fórmula de fixação de mutantes diplóides na presença de seleção apenas que ela se aplica exatamente na mesma forma para indivíduos haplóides, usando-se a transformação $2N \rightarrow N$ e $s \rightarrow 2s$ (veja o Apêndice A).

3.3 Resultados

Primeiramente mostraremos resultados de população infinita, e analisaremos a probabilidade de absorção, e em seguida as probabilidades e os tempos de fixação de populações finitas. Depois desenvolvemos aproximações analíticas para regimes de mutação rara e frequente, e encontramos as condições em que o mutante menos apto prevalece. Finalmente mostramos resultados generalizando o modelo dado pela Equação 3.6 para três tipos mutantes.

A partir da normalização apenas entre as abundâncias dos mutantes em uma população infinita dadas pelas Equações 3.4 e 3.5, fizemos o gráfico destas abundâncias relativas no tempo, mostrado na Figura 15. Nele vemos que durante a dinâmica ocorre uma inversão do tipo que prevalece mais. Isto nos sugere que um tipo pode prevalecer mais que outro dependendo dos tempos de fixação. Além disso, existe uma dependência do tempo de inversão com a mutabilidade (não-mostrada). Tudo isso nos sugere que a mutabilidade pode influenciar qual mutante irá se fixar.

Iremos agora estudar o efeito de quatro “forças” evolutivas sobre a escolha de um passo mutacional de um tipo selvagem para dois mutantes, a saber: a mutação (a rigor, a sua taxa), a deriva genética (em termos do tamanho da população), a seleção natural e o fluxo gênico.

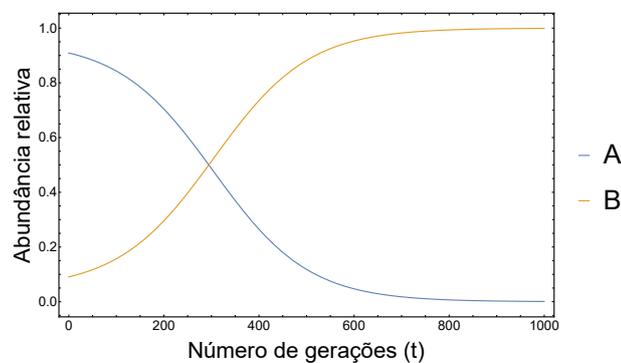


Figura 15 – Abundância relativa de cada tipo dentre os mutantes em uma população infinita. Os parâmetros são $r = 1 \times 10^{-12}$, $s_A = 0,01$, $s_B = 0,02$ e $\rho = 10$.

Os resultados simulacionais que mostram o efeito da mutabilidade, do tamanho da população, da aptidão e do fluxo gênico sobre a prevalência do tipo menos apto no estado estacionário são mostrados a seguir na Figura 16.

Na Figura 16(a) vemos que a probabilidade de que o estado absorvente seja o todo-A é próximo da unidade para mutabilidades bem baixas, e torna-se praticamente nula para mutabilidades altas. A razão para isso é que para mutabilidades altas a chance de coexistência do tipo A com o tipo mais apto B é maior, e com isso acontece uma competição direta em que B tem uma grande vantagem, dependendo da diferença das aptidões dos mutantes.

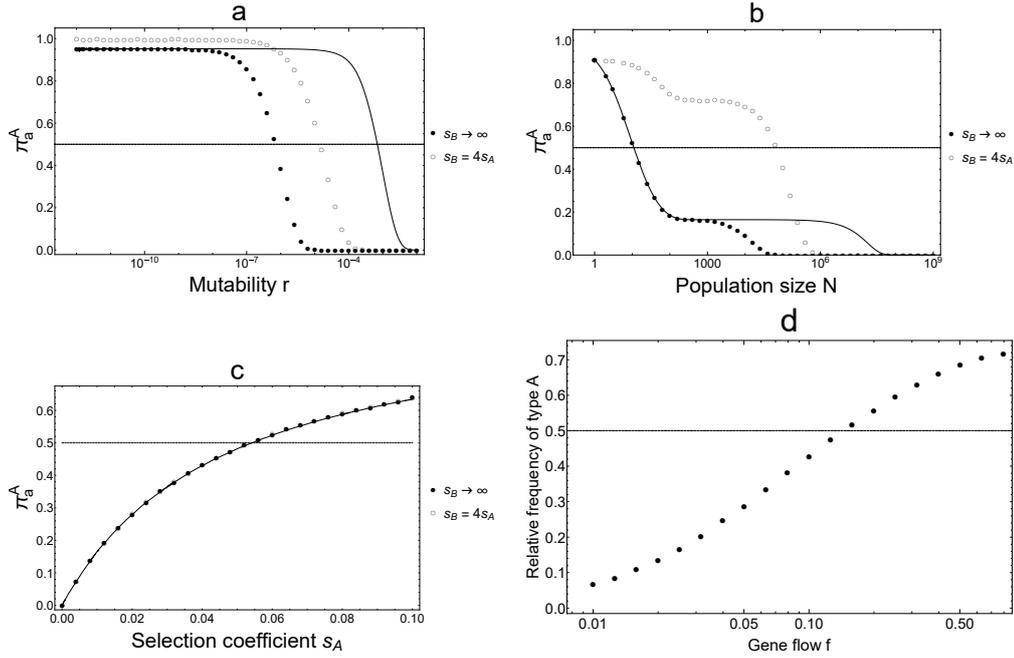


Figura 16 – Dependência da prevalência do tipo A da mutabilidade r , o tamanho da população N , a aptidão (em termos do coeficiente de seleção s_A) e o fluxo gênico f . Os discos e os círculos representam resultados das simulações. Os discos sólidos (●) representam uma aptidão infinita para o tipo B e os círculos vazados (○) representam um coeficiente de seleção de B quatro vezes mais alto que o do tipo A . A linha contínua (—) representa a aproximação analítica, Eq. 3.10. No gráfico (d), “Frequência relativa” significa a frequência de A em relação a ambos os mutantes no estado estacionário. Os coeficientes de seleção s_A e s_B são dados por $s_A = w_A - 1$ e $s_B = w_B - 1$. Os parâmetros são (a) $N = 10^6$, $s_A = 0,01$, $\rho = 1000$; (b) $r = 1 \times 10^{-6}$, $s_A = 0,01$, $\rho = 10$; (c) $N = 10^6$, $r = 1 \times 10^{-12}$, $\rho = 10$; e (d) $N = 100$, $\rho = 1,2$, $w_A = 1,01$, $w_B = 1,04$.

É interessante que mesmo que B tenha uma aptidão infinita, isto é, a presença de um único indivíduo deste tipo garante sua absorção, a probabilidade de absorção do tipo A para mutabilidades baixas ainda é próxima de um. Isso acontece porque como as taxas de mutação são extremamente baixas, a fixação do tipo A acontece sem que surja um competidor do tipo B . Considerando-se ainda que $r_A > r_B$, a chance de surgir um mutante do tipo A é maior que surgir um tipo B .

Por outro lado, o tipo A não possui mais esta vantagem se a população é grande. Neste caso, o suprimento mutacional do tipo B favorece a coexistência dos tipos mutantes, e o B sempre vencerá. Isto pode ser visto na Figura 16(b). A Figura 16(c) mostra que o aumento da aptidão do tipo A no regime de mutações raras aumenta sua probabilidade de absorção π_A^A , como esperado.

Na presença de migração, existe um fluxo contínuo de tipos W e portanto nenhum estado absorvente. O mutante que prevalece pode ser definido como o mais populoso, mais especificamente, como aquele que alcança primeiro mais da metade de toda a população.

A Figura 16(d) mostra que o tipo menos apto é beneficiado pelo aumento da taxa de migração. A remoção frequente de mutantes por fluxo gênico é compensado pela maior taxa de mutação deste tipo, o mutante A .

Para melhor entendermos os resultados de simulação, vamos desenvolver uma aproximação analítica para as probabilidades de fixação dos mutantes. Esta aproximação pode ser então inserida na relação

$$\pi_a^i = \frac{\pi_f^i}{\pi_f^A + \pi_f^B}, \quad (3.7)$$

onde $i \in \{A, B\}$, para obter as probabilidades de absorção. Suponha que todos os mutantes que podem ocorrer surjam por mutação em uma mesma geração, uma hipótese que simplifica a análise. A probabilidade de um conjunto de subpopulações (N_W, N_A, N_B) , na presença de mutação e deriva na mesma geração é

$$p(N_W, N_A, N_B) = \frac{N!}{N_W!N_A!N_B!} (1 - r - r/\rho)^{N_W} \times r^{N_A} (r/\rho)^{N_B} \quad (3.8)$$

Quando o tipo B possui aptidão infinita, o tipo A irá se fixar apenas quando o tipo B não surgir, isto é, $N_B = 0$. Na presença de seleção somente, a probabilidade de fixação de N_A indivíduos em uma população com $N - N_A$ indivíduos selvagens é dado pelo resultado clássico (EWENS, 2004; GILLESPIE, 2004)

$$\Pi_A(N_A) = \frac{1 - e^{-2N_A s_A}}{1 - e^{-2N s_A}} = \frac{1 - \omega_A^{N_A}}{1 - \omega_A^N}, \quad (3.9)$$

onde $w_A = 1 + s_A$ e $\omega_A = e^{-2s_A}$.

Assim, a probabilidade de fixação de A pode ser aproximada como o produto da probabilidade na configuração $(N_W, N_A, 0)$ e a probabilidade de fixação partindo desta configuração apenas na presença de seleção somada sobre todas as configurações sem indivíduos B , isto é,

$$\begin{aligned} \pi_f^A &= \sum_{N_A=1}^N \frac{N!}{N_W!N_A!} (1 - r - r/\rho)^{N_W} r^{N_A} \frac{1 - \omega_A^{N_A}}{1 - \omega_A^N} \\ &= \frac{\left(1 - \frac{(\rho+1)r}{\rho}\right)^N}{\omega_A^N - 1} \times \\ &\quad \left[\left(1 - \frac{\rho r \omega_A}{\rho(r-1) + r}\right)^N - \left(\frac{r - \rho}{\rho(r-1) + r}\right)^N \right]. \end{aligned} \quad (3.10)$$

Para uma mutabilidade muito baixa r , o termo com $N_A = 1$ domina a soma, e a expressão se reduz a

$$\pi_f^A = \theta_A \frac{1 - \omega_A}{1 - \omega_A^N},$$

onde $\theta_A = Nr$ é o suprimento mutacional do tipo A em uma população selvagem, e o segundo fator é a probabilidade de um único indivíduo A se fixar. A restrição para $r \ll 1/N$ indica que a absorção do tipo A é fortemente afetada pelos mutantes que surgem nas primeiras gerações, isso é, pelo suprimento mutacional inicial.

De forma semelhante, a probabilidade de fixação do tipo B é dada pela probabilidade de que pelo menos um mutante B surja ($N_B > 0$) e não depende da presença do tipo A , isto é

$$\begin{aligned} \pi_f^B &= \sum_{N_A=0}^{N-1} \sum_{N_B=1}^{N-N_A} \frac{N!}{N_W! N_A! N_B!} (1-r-r/\rho)^{N_W} r^{N_A} (r/\rho)^{N_B} \\ &= \left(1 - \frac{(\rho+1)r}{\rho}\right)^N \left[\left(\frac{1}{1-r}\right)^N \left(\frac{\rho(r-1)}{\rho(r-1)+r}\right)^N - \left(\frac{r-\rho}{\rho(r-1)+r}\right)^N \right] \end{aligned}$$

Novamente, para uma baixa mutabilidade, esta fórmula simplifica para

$$\pi_f^B = \theta_B,$$

onde $\theta_B = \theta_A/\rho$ é o suprimento mutacional para o tipo B de uma população selvagem. A comparação com o limite de baixa mutabilidade da probabilidade de absorção do tipo A mostra que o suprimento mutacional é um fator determinante. No entanto, para o tipo B , é como se houvesse um termo de probabilidade de fixação, apenas na presença de seleção, igual a um multiplicando o suprimento mutacional. Em outras palavras, a aproximação de r baixo mostra que apesar do tipo A ter uma pequena probabilidade de fixação apenas na presença de seleção, isto pode ser compensado por um alto suprimento mutacional.

As expressões para as probabilidades de absorção estão representadas como linhas sólidas nas figuras 16(a)–16(c). Elas funcionam bem no regime de baixa mutabilidade, mas a mutabilidade mais alta elas superestimam π_a^A . Isto é explicado pelo papel mais importante de mutações recorrentes a mutabilidades intermediárias, facilitando a absorção por B . Naturalmente, quanto maior a aptidão de B , mais forte será o efeito.

A hipótese usada para derivar a aproximação analítica lança luz sobre o mecanismo que rege a absorção. Considerando o suprimento mutacional, isto é, o produto do tamanho da população com a taxa de mutação total, nós vemos que com um suprimento mutacional baixo ($Nr \ll 1$) a dinâmica é bem descrita como “um mutante por vez”. Em contraste, a não-concordância entre a aproximação e a simulação a níveis intermediários de mutabilidade indica o papel crucial das mutações recorrentes, que ocorrem apenas porque existem tipos selvagens em número suficiente para sofrerem mutação, mesmo depois da primeira geração de mutantes. O platô observado na Figura 16(b) sinaliza uma região em que ainda se tem N pequeno de modo que $Nr \ll 1$, mas já se tenha N suficientemente grande de modo que $\omega_A^N \approx 0$. Nessa região, π_a^A deve ficar praticamente independente de r e de N . Em algum ponto, para tamanhos de população suficientemente grandes, as mutações recorrentes são responsáveis pelo decréscimo abrupto de π_a^A .

A rota para a absorção nem sempre é direta. Os mutantes podem surgir e serem extintos antes que se espalhem. A probabilidade de fixação do tipo selvagem é uma medida da chance de mutações desaparecerem antes que possam ser fixadas. Definimos o tempo de fixação do tipo A ou B como o número de gerações entre a ocorrência do primeiro mutante daquele tipo e sua fixação. É também possível definir o tempo de fixação do tipo selvagem como o número de gerações entre a ocorrência do primeiro mutante A ou B e a fixação de W . A Figura 17 mostra a dependência da probabilidade de fixação e os tempos de fixação com a taxa de mutação.

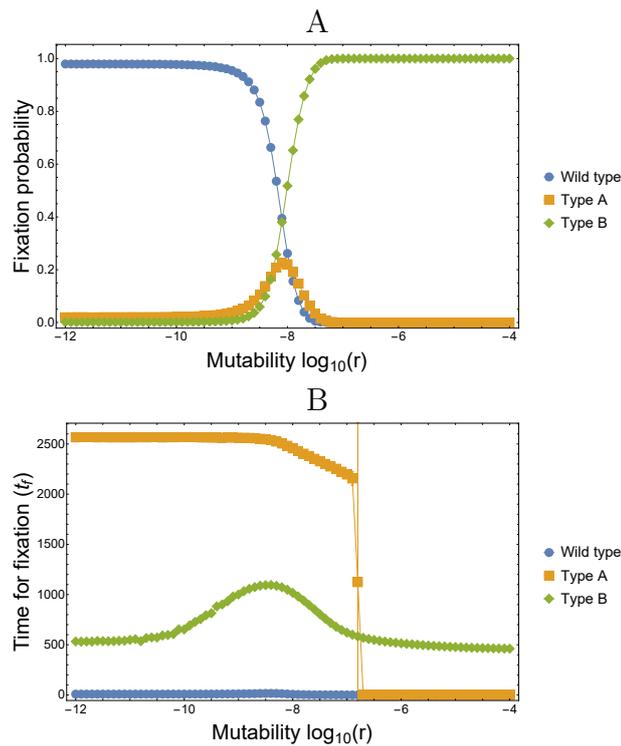


Figura 17 – Probabilidade de fixação de cada tipo (A) e o tempo de fixação (B) como função da mutabilidade. A queda abrupta do tempo de fixação do tipo A e seu desvio padrão é devido ao tamanho finito da amostra. Os parâmetros são $N = 10^7$, $s_A = 0,01$, $s_B = 0,056$ e $\rho = 100$.

A baixas taxas de mutação, o tipo selvagem se fixa mais do que as outras mutabilidades por causa de sua superioridade numérica sobre os mutantes raros. Apesar da probabilidade de fixação dos mutantes parecer ser baixa para mutabilidades baixas, note que, no final das contas, nenhum selvagem permanece. Porque o tempo de fixação de W mede o tempo de retorno para o estado homogêneo depois que ele foi modificado por um único mutante, se espera que o tempo de fixação seja bem pequeno quando comparado ao dos mutantes, como mostrado na Figura 17. De fato, a fixação do tipo selvagem pode ser visto como um processo de renovação (*renewal process*) (KAMPEN, 2011) no qual a população pode voltar ao estado todo- W depois da primeira mutação. A probabilidade que a primeira mutação seja do tipo A ou tipo B é igual a $r_A/(r_A + r_B)$ ou $r_B/(r_A + r_B)$, respectivamente. Assim, no regime de mutações raras, a probabilidade de fixação do tipo

selvagem é dada por

$$\pi_f^W = \frac{\rho}{\rho + 1} \frac{1 - \left(\frac{1}{\omega_A}\right)^{N-1}}{1 - \left(\frac{1}{\omega_A}\right)^N} + \frac{1}{\rho + 1} \frac{1 - \left(\frac{1}{\omega_B}\right)^{N-1}}{1 - \left(\frac{1}{\omega_B}\right)^N}.$$

Note que se N é grande, então $\pi_f^W \approx 1$, isto é, na maior parte do tempo as mutações são perdidas e a população retorna ao estado homogêneo selvagem.

A Figura 17 mostra que no regime de baixa mutabilidade, apesar do tipo A se fixar mais que o tipo B , o tipo A leva mais tempo para se fixar. De maneira interessante, o tempo de fixação do tipo A é reduzido na medida em que a mutabilidade aumenta, apesar da chance de competição com o tipo B aumentar também. O tempo de fixação de B mostra um pico em uma taxa de mutação próxima do valor no qual a probabilidade de fixação do tipo A é maior que a dos outros variantes. Este tempo de fixação aumenta porque existem mais indivíduos do tipo A para competir. E o tempo de fixação de A diminui porque são justamente os processos em que a fixação é rápida que A se fixa sem competição significativa. Na medida em que a mutabilidade é ainda mais aumentada, o número de ambos os mutantes aumenta, e as flutuações não favorecem mais uma maior competitividade do tipo A , de forma que o tempo de fixação de B diminui.

Para se ter uma imagem melhor dos mecanismos de fixação de cada tipo, nós analisamos uma população composta somente do tipo selvagem e N_e mutantes iniciais do mesmo tipo com o coeficiente de seleção efetivo s_e , sem qualquer mutação no processo evolutivo. Então, nós podemos perguntar quais são os valores efetivos N_e e s_e que levam à mesma probabilidade de fixação e tempos de fixação do modelo completo. Essa foi uma sugestão de um revisor do nosso artigo. Para calcularmos este valor, precisamos de duas equações. A primeira é a Equação (3.9); como segunda equação nós podemos usar o tempo de fixação no regime difusivo (CHARLESWORTH, 2020) para as nossas condições (alto $\alpha = 2Ns$), dado por

$$\tau = \int_q^1 \frac{\left(1 - e^{-2(1-q')Ns}\right) \left(1 - e^{-2q'Ns}\right)}{(1 - q')q's(1 - e^{-2Ns})} dq'. \quad (3.11)$$

onde q é a frequência inicial do tipo mutante. Para nosso $\alpha = 10^5$ o integrando entre zero e q é essencialmente nulo. O número inicial de mutantes é então Nq . Os resultados mostrados na Figura 17 são utilizados como dados de entrada para as Equações (3.9) e (3.11). O sistema de equações resultantes é resolvido usando-se o algoritmo QAGS (rotina de quadratura, integrando adaptativo, função geral, singularidades podem ser mais facilmente integradas) e o algoritmo `gls_multiroot_fsolver`, ambos da *GNU Scientific Library*.

A Figura 18 mostra o tamanho efetivo da subpopulação de mutantes e o coeficiente de seleção efetivo para ambos os tipos de mutantes. No limite difusivo, a mudança esperada

do número de mutantes em um passo é proporcional ao coeficiente de seleção e a variância da mudança é inversamente proporcional ao tamanho. Logo, para o tipo B , a variância decresce consistentemente com o aumento da mutabilidade, enquanto seu passo médio tem um mínimo onde a competição é máxima. Assim, a esta mutabilidade, o tipo B varia em passos pequenos enquanto os passos do tipo A variam de uma maneira menos regular. Esta pode ser a razão para a relativa competitividade que B encontra na vizinhança deste valor de mutabilidade. A pequenas mutabilidades, o número efetivo de mutantes inicial é próximo da probabilidade de se ter um determinado mutante no primeiro passo do modelo completo, $\rho/(\rho + 1)$ e $1/(\rho + 1)$ para os tipos A e B , respectivamente. Esse resultado corrobora o regime de “um mutante por vez”. Na medida que o valor da mutabilidade aumenta, o mesmo comportamento qualitativo mostrado na Figura 17 é observado, com um pico de N_e próximo do valor de r associado com o pico na probabilidade de fixação. Em contraste à grande variação no tamanho efetivo, a seleção efetiva é praticamente constante para o tipo A e tem pequenas variações para o tipo B .

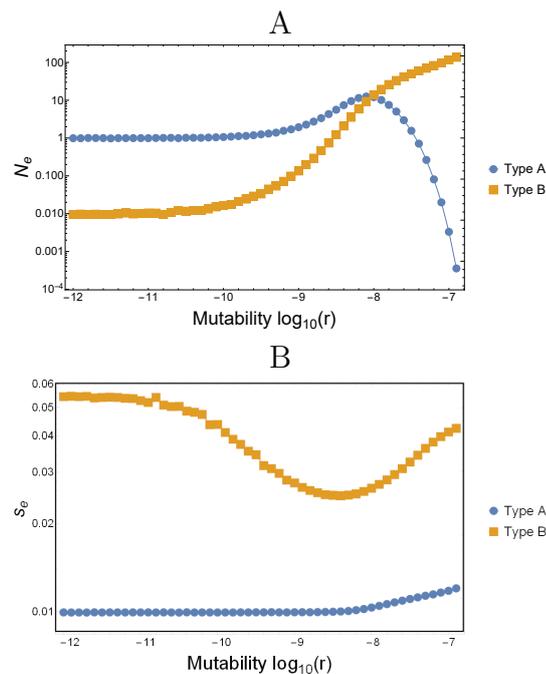


Figura 18 – Tamanhos efetivos de cada população mutante e coeficientes de seleção efetivos. Os dois gráficos, (A) e (B), devem ser lidos em pares: para cada r , os valores de N_e e s_e correspondem à solução de um sistema de equações composto com o tipo selvagem e N_e mutantes de um único tipo mutante que conduz às mesmas probabilidades de fixação e tempos de fixação. O dado original vem de simulação. Os parâmetros são $N = 10^7$, $s_A = 0,01$, $s_B = 0,056$ e $\rho = 100$.

A evolução temporal das abundâncias dos tipos A e B , mostrada na Figura 19, exibe muitas características interessantes. A baixas taxas de mutação, não existe coexistência A - B em qualquer ponto da evolução, mas na medida que deixamos o regime de baixa mutabilidade, a coexistência ocorre. Interessantemente, a evolução em casos exibindo a fixação de A é sempre monotônica, enquanto aqueles em que B se fixa pode ser não

monotônico. A razão para a não-monotonicidade no último caso é simples: uma vez que o tipo B é mais apto, ele pode se recuperar depois de um surgimento tardio. Se a taxa de mutação é alta, então o tipo B domina facilmente.

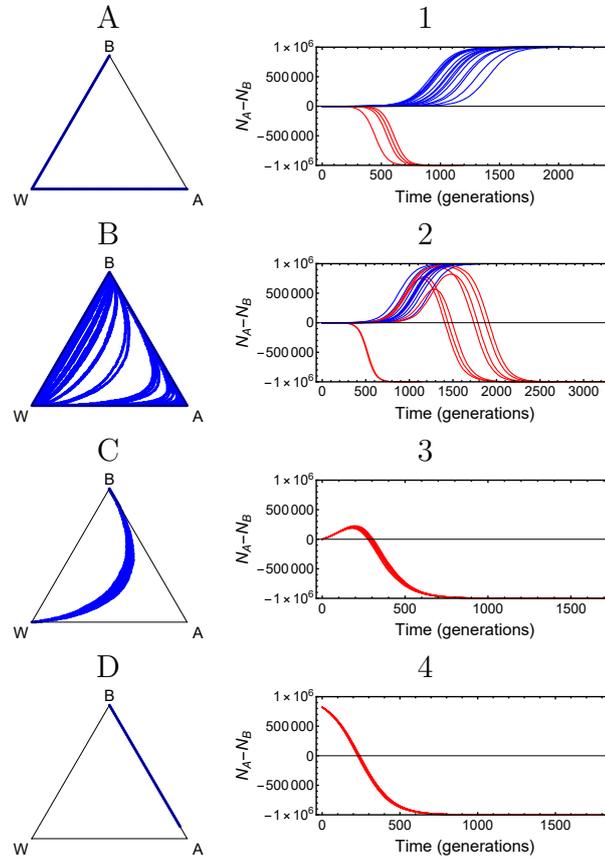


Figura 19 – Histórias simuladas típicas da composição da população, encerradas na absorção de A ou B . Os gráficos da esquerda (A–D) mostram a evolução no simplex $N_W + N_A + N_B = N$, com os vértices correspondendo a estados de único tipo e os lados a estados com dois tipos. Os gráficos da direita (1–4) mostram a diferença $N_A - N_B$ (independentemente do valor de N_W) ao longo do tempo. Histórias que acabam na absorção de A (B) são mostradas em azul (vermelho). As linhas correspondem às mutabilidades (de cima para baixo) $r = 10^{-11}$, $r = 10^{-7}$, $r = 10^{-3}$ e $r = 10/11$, que é o maior valor possível de r com $\rho = 10$ em que $r_a + r_B = (1 + 1/\rho)r \leq 1$, condição do modelo de Wright-Fisher. Os parâmetros são $N = 10^6$, $s_A = 0,01$, $s_B = 0,02$ e $\rho = 10$.

Finalmente, a evolução predita pela aproximação usada anteriormente, onde se supôs que todos os mutantes aparecem de uma só vez, com altas mutabilidades, é mostrado na última fileira da Figura 19. As trajetórias ocorrem em sua maioria no segmento AB (exceto pelo vértice todo- W) começando em torno de $N_A = N\rho/(\rho + 1)$ e $N_B = N - N_A$, dependendo dos parâmetros, algumas vezes tendendo para o vértice todo- A ou todo- B . Para tais parâmetros, o destino da fixação é quase exclusivamente para B .

No regime em que as mutações acontecem mais lentamente que a seleção, a fixação ocorre antes da próxima mutação acontecer. Este é o regime de “seleção forte-mutação

fraca” (SSWM, do inglês *strong selection-weak mutation*, como definido por Gillespie (1983) (GILLESPIE, 1983)). Neste regime, a análise é muito simplificada e provê suporte analítico para as simulações no limite de baixo suprimento mutacional.

Para se determinar as probabilidades de absorção nós notamos que, no regime SSWM, somente um mutante coexiste com o tipo selvagem. Assim, a próxima equação fornece a probabilidade de fixação apenas sob seleção (i.e. sem considerar processos de mutação dando origem aos mutantes iniciais ou outros mutantes) tanto de A como de B , isto é,

$$\Pi_f^i = \frac{1 - \omega_i}{1 - \omega_i^N}. \quad (3.12)$$

A probabilidade de absorção do tipo i é portanto dada por

$$\pi_a^i = \frac{r_i \Pi_f^i}{r_A \Pi_f^A + r_B \Pi_f^B}. \quad (3.13)$$

a qual, após substituição, se torna

$$\pi_a^A = \frac{1}{1 + \frac{1 - \omega_B}{\rho} \frac{1 - \omega_A^N}{1 - \omega_A}}, \quad (3.14)$$

onde $\omega_A = e^{-2s_A}$ e $\omega_B = e^{-2s_B}$.

Para se verificar esta fórmula, o caso $\omega_A = \omega_B$ é útil. Neste caso não existe vantagem adaptativa entre os mutantes e a probabilidade de absorção deve depender apenas das taxas de mutação. De fato, a Equação 3.14 simplifica para

$$\pi_a^A = \frac{\rho}{1 + \rho}. \quad (3.15)$$

Em nosso problema original, o tipo A é menos apto que o tipo B , mas surge mais frequentemente que o tipo B , de forma que a prevalência do tipo A permanece possível. Para quantificar a prevalência de A sobre B , nós determinamos a condição $\pi_a^A > 1/2$. Rearranjando os termos na Equação 3.14, encontramos

$$\rho > \frac{1 - \omega_B}{1 - \omega_B^N} \frac{1 - \omega_A^N}{1 - \omega_A}. \quad (3.16)$$

Uma vez que analisamos mutações adaptativas, temos que $0 < \omega_B < \omega_A < 1$. No regime $N \rightarrow \infty$, a condição simplifica para

$$\rho > \frac{1 - \omega_B}{1 - \omega_A}. \quad (3.17)$$

Como $\omega_B < \omega_A$, segue-se que ρ deve ser sempre maior do que um. Para seleção fraca, $e^{s_A} \approx w_A = 1 + s_A$ e $e^{s_B} \approx w_B = 1 + s_B$, e a Equação 3.17 pode ser mais simplificada para

$$\rho \gtrsim s_B/s_A. \quad (3.18)$$

A condição na Equação (3.16), ou a simplificação na Equação (3.18), mostra que, para o tipo A ter uma vantagem evolutiva maior que o tipo B , então a taxa de mutação do tipo A deve ser maior que a do tipo B , pelo menos o mesmo tanto como determinado pela Equação (3.16). Note que, quanto maior a vantagem adaptativa de B maior deve ser ρ , como mostrado na Figura 20.

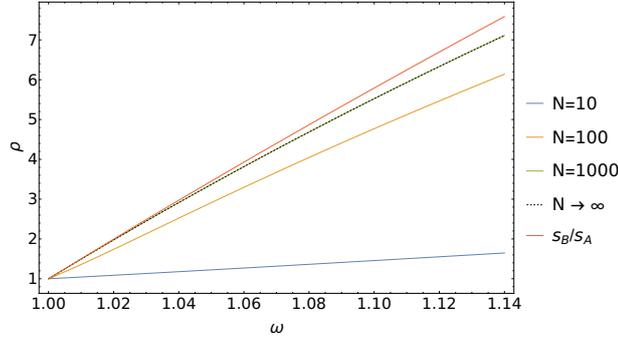


Figura 20 – Limiar para ρ acima do qual o tipo A prevalece no regime de mutações raras da Eq. (3.16) para $N = 10$ (azul), $N = 100$ (amarelo), e $N = 1000$ (verde). O limite $N \rightarrow \infty$, Eq. (3.17) (preto, tracejado) e a aproximação da Eq. (3.18) (vermelho) são também mostrados. O eixo ω corresponde a ω_A/ω_B . Quanto maior ω , maior a vantagem seletiva de B sobre A . O coeficiente $s_A = 0,01$, o que implica $\omega_A \approx 0.98$. Note que a aproximação $N \rightarrow \infty$ tem um bom ajuste para $N = 1000$. Quanto maior a vantagem de B sobre A ou quanto maior o tamanho da população, maior é a razão ρ necessária para A prevalecer.

No regime de mutações frequentes $r_A + r_B \lesssim 1$, todas as mutações acontecem aproximadamente ao mesmo tempo. Assim, a população é composta por tipos A e B apenas, e o equivalente da Eq. (3.9) (trocando w_W , que é igual a um, por w_B) é

$$\pi_f^A = \frac{1 - \omega^{N_A}}{1 - \omega^N},$$

onde $\omega = \omega_A/\omega_B$. A probabilidade de se ter N_A mutantes na primeira geração é dada pela distribuição binomial $B(N, \frac{\rho}{\rho+1})$ com função massa de probabilidade $f(N_A)$. Assim, a probabilidade de absorção de A é

$$\begin{aligned} \pi_a^A &= \sum_{N_A=1}^N f(N_A) \frac{1 - \omega^{N_A}}{1 - \omega^N} \\ &= \sum_{N_A=1}^N \binom{N}{N_A} \left(\frac{\rho}{\rho+1}\right)^{N_A} \left(\frac{1}{\rho+1}\right)^{N-N_A} \frac{1 - \omega^{N_A}}{1 - \omega^N} \\ &= \frac{1 - \left(\frac{\rho\omega+1}{\rho+1}\right)^N}{1 - \omega^N}. \end{aligned} \quad (3.19)$$

Esta equação também simplifica para a Eq. (3.15) se $s_A = s_B$, o que de novo é um resultado consistente.

A condição para $\pi_a^A > 1/2$ é

$$\rho > \frac{T-1}{\omega-T}, \quad (3.20)$$

onde

$$\begin{aligned}
T &= \left[\frac{1}{2}(\omega^N + 1) \right]^{1/N} \\
&= \omega e^{\frac{-\ln 2}{N}} \left[1 + \frac{1}{\omega^N} \right]^{1/N} \\
&= \omega \left[1 - \frac{\ln 2}{N} + \frac{1}{2!} \left(\frac{\ln 2}{N} \right)^2 + \dots \right] \left[1 + \frac{1}{N} \frac{1}{\omega^N} + \frac{1}{2} \frac{1}{N} \left(\frac{1}{N} - 1 \right) \left(\frac{1}{\omega} \right)^2 + \dots \right] \\
&= \omega \left(1 - \frac{\ln 2}{N} + \frac{1}{N\omega^N} + \frac{\ln 2}{N^2} - \frac{\ln 2}{N^2\omega^N} + \dots \right).
\end{aligned}$$

Para $N \gg 1$, porque $\omega > 1$, T pode ser bem aproximado pelos termos iniciais,

$$T \approx \omega \left(1 - \frac{\ln 2}{N} \right),$$

e a Eq. (3.20) se torna,

$$\rho > N \frac{(\omega - 1)}{\omega \log 2}. \quad (3.21)$$

Logo, quanto maior o tamanho ou maior a vantagem seletiva de B sobre A ($\omega = e^{2(s_B - s_A)}$), mais difícil se torna para A vencer. Ainda é possível que A prevaleça, mas a taxa de mutação para este tipo deve ser muito alta.

Os resultados obtidos com nosso modelo com dois mutantes adaptativos são também observados em sistemas que possuem três mutantes adaptativos, como mostrado na Figura 21. Neste último modelo todos os mutantes são gerados do tipo W . Os coeficientes de seleção dos tipos A , B e C são $s_A = 0,01$, $s_B = 0,02$ e $s_C = 0,04$. As taxas de mutação são $r_A = r$, $r_B = r_A/\rho$ e $r_C = r_B/\rho = r_A/\rho^2$, com $\rho = 10$. O tamanho da população é $N = 10^6$. A amostragem de novos indivíduos é realizada de acordo com

$$\begin{aligned}
P_W^{t+1} &= N_W^t (1 - r_A - r_B - r_C), \\
P_A^{t+1} &= w_A N_A^t + N_W^t r_A, \\
P_B^{t+1} &= w_B N_B^t + N_W^t r_B, \\
P_C^{t+1} &= w_C N_C^t + N_W^t r_C,
\end{aligned}$$

O comportamento qualitativo geral persiste: na medida em que aumentamos a mutabilidade, mutantes com taxas de mutação menores e aptidões maiores prevalecem.

3.4 Discussão

A compreensão tradicional desde a síntese moderna do papel das mutações é que nos casos mais típicos, a maioria da variabilidade de uma população, depois de algumas rodadas de mutação, já está disponível e o papel das mutações é meramente mudar

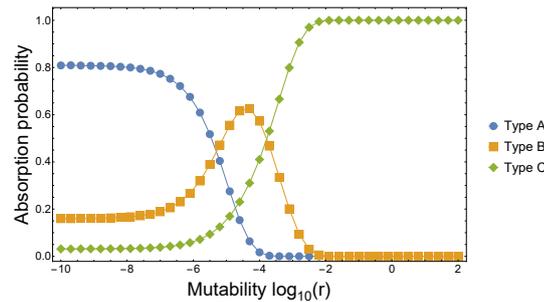


Figura 21 – Sistema com três mutantes adaptativos, com aptidões crescentes e taxas de mutação decrescentes. Os parâmetros são $N = 10^6$, $s_A = 0,01$, $s_B = 0,02$, $s_C = 0,04$, $\rho = 10$ de A para B e de B para C .

a frequência de variantes pré-existentes (HALDANE, 1933; MAYNARD-SMITH et al., 1985; BULL et al., 1997). No entanto, Yampolsky e Stoltzfus (2001) (YAMPOLSKY; STOLTZFUS, 2001) mostraram que, sob competição entre mutação e seleção, o viés na taxa de mutação pode mudar o alelo que prevalece. Nós apresentamos aqui uma análise exaustiva das condições para a qual o viés de mutação muda o tipo que prevalece.

Nossa análise emprega um modelo de Wright-Fisher começando com uma população selvagem que pode sofrer mutações de mão única para dois outros mutantes, A e B , onde B é mais apto que A , enquanto a taxa de mutação para o tipo A é maior que para B . Nós verificamos que um viés de mutação suficientemente grande pode alterar o tipo que prevaleceria se apenas a seleção estivesse em ação. Observamos isto para mutabilidade, deriva genética, vantagem seletiva e fluxo gênico variáveis.

Este resultado implica que as condições sob as quais o destino evolutivo muda são muito mais amplas que o comumente reconhecido. Um foco no efeito da mutabilidade mostra que cinéticas distintas são possíveis, mas que para dois alelos mutantes existem apenas dois regimes básicos: um com mutações raras e outro com mutações frequentes. No regime raro, o tipo menos apto pode prevalecer, enquanto no regime de mutações frequentes isto é praticamente impossível.

3.5 Conclusão

Nós esperamos que este trabalho embase explicações para a dinâmica em paisagens adaptativas mais complexas, revelando porquê um caminho prevalece sobre outros. Isto leva a questões de preditibilidade e repetibilidade dos processos evolutivos (MEYER et al., 2012; BURKE; LITI; LONG, 2014; LACHAPPELLE; REID; COLEGRAVE, 2015; BAUER; GOKHALE, 2015; DIJK et al., 2017; NICHOL et al., 2019).

Nos próximos capítulos nós pretendemos investigar a trajetória evolutiva em paisagens adaptativas simples, completamente suaves com tamanhos de população, aptidão e mutabilidade variadas. Isto deve nos conduzir a compreensões adicionais para se compre-

ender a cinética e o destino evolutivo em paisagens adaptativas realísticas e experimentais (LUNZER et al., 2005; GOULART et al., 2013; MIRA et al., 2015; PALMER et al., 2015; RODRIGUES et al., 2016; TAMER et al., 2019; KNOPP; ANDERSSON, 2018).

4 Ordem das mutações

4.1 Introdução

Após examinarmos como a competição entre seleção e mutação levam a mutabilidade a influenciar o passo adaptativo a ser tomado, examinamos se a mutabilidade pode levar a um processo adaptativo a tomar uma trajetória ou outra. Neste capítulo, o genótipo de destino é igual para todas as trajetórias. O que muda entre as trajetórias é a ordem com que as mutações ocorrem. Ainda não estamos respondendo aqui se a mutabilidade pode levar a destinos evolutivos diferentes, mas os resultados deste capítulo preenchem uma lacuna entre o capítulo anterior e a pergunta se a mutabilidade muda o destino da população, a ser respondida no capítulo 5.

Entender a ordem como as mutações se sucedem em uma linhagem vencedora nos permite compreender mais integralmente a história evolutiva, e nos habilita a interferir em processos adaptativos em andamento, como em melhoramentos genéticos, a evolução de resistência de patógenos a antibióticos, e na produção de vacinas contra variantes de vírus.

Uma distinção se impõe: não estamos examinando aqui a sequência de mutantes que prevalecem com o tempo. Estamos sim analisando a história de sucessões de mutações no indivíduo que domina a população dado tempo suficiente. A simulação é interrompida assim que todos os indivíduos que seguiram diferentes trajetórias evolutivas acumulam todas as mutações adaptativas possíveis. A probabilidade de uma trajetória se fixar será então dada pela proporção de indivíduos que seguiram este mesmo caminho em relação a toda a população, porque a partir daí a evolução é neutra.

4.2 Métodos

Para investigarmos caminhos evolutivos, primeiramente definimos o número G de sítios envolvidos no processo evolutivo, que podem corresponder a locos ou a sítios distintos de mutação em um mesmo loco. Cada sítio é numerado de 1 a G , em ordem crescente de coeficiente de seleção, que inicialmente é dado por ns , onde n é a numeração do sítio. A aptidão total é igual a um mais os coeficientes de seleção de cada sítio (aptidão aditiva). A taxa de mutação do sítio 1 com aptidão típica s é dada por r , para qualquer dos modelos usados.

Definindo $\rho = e$, temos como antes $r_1 = r$, $r_2 = r/\rho$, $r_3 = r/\rho^2$ etc. A taxa de mutação para o sítio n é, portanto, $r_n = r/\rho^{n-1} = re^{-(n-1)}$. Quanto maior n , menor a taxa de mutação. O parâmetro r pode ser identificado como a mutabilidade.

Para testarmos se os resultados são sensíveis a detalhes da distribuição de mutações adaptativas, utilizamos uma classe de distribuições que inclui a distribuição exponencial, chamada de Gumbel. Ela é dada pela seguinte fórmula:

$$f(s) = \frac{1}{\sigma} \frac{e^{-(s/\sigma)^\beta}}{\Gamma(1 + \beta^{-1})}, \quad (4.1)$$

onde σ é a média da distribuição quando ela se reduz à distribuição exponencial, e β é um parâmetro do modelo que torna as mutações mais vantajosas mais raras quando aumentada. Para obtermos a distribuição exponencial, fazemos $\beta = 1$. Em ambas as distribuições, a probabilidade de uma mutação mais vantajosa sempre é menor. Para calcularmos a taxa de uma mutação com coeficiente de seleção ns com $s = \sigma$ e que segue a classe de distribuições de Gumbel, definimos a taxa como

$$r_n = r e^{-(n-1)^\beta}, \quad (4.2)$$

para que a taxa de mutação de referência do primeiro sítio seja r e que as demais taxas sejam proporcionais à densidade de probabilidade da classe de Gumbel.

Exceto o tipo selvagem, indivíduos em uma dada trajetória adaptativa podem ser distinguidos pela sequência de mutações de seus ancestrais. Eles formam o que chamamos de uma classe.

A amostragem é feita por classes (portanto distingue genótipos iguais com trajetórias diferentes) através do processo de Wright-Fisher semelhante ao usado no caminho evolutivo simples, onde a probabilidade de uma classe ser sorteada para a próxima geração é dada por

$$p_c^{t+1} \propto f_s N_s^t r_{cs} + f_c N_c^t (1 - \sum_d r_{dc}), \quad (4.3)$$

onde p_c é a probabilidade de sorteio da classe c , f_c e f_s são as aptidões da classe atual e da qual ela é originada, respectivamente, N_c^t e N_s^t são o número de indivíduos da classe atual e da qual ela é originada, respectivamente, no tempo t , r_{cs} é a taxa de mutação do genótipo da classe original para o genótipo da classe atual, r_{dc} é a taxa de mutação do genótipo da classe atual para um genótipo d dentre os possíveis destinos imediatos, e \sum_d é a soma sobre todos os destinos possíveis. Um destino é possível se sua aptidão é maior que a do genótipo que o dá origem, e está separado do genótipo de origem por apenas uma mutação.

Como no método para o caminho adaptativo simples, esse esquema equivale a uma fase de seleção logo após o surgimento do indivíduo, e outra fase de reprodução, onde as mutações acontecem.

As possíveis trajetórias de cada indivíduo são mostradas na Figura 22.

Duas medidas de entropia são realizadas, usadas tanto aqui como em análises futuras. Em ambas, rodamos cada simulação até que toda a população atinja apenas

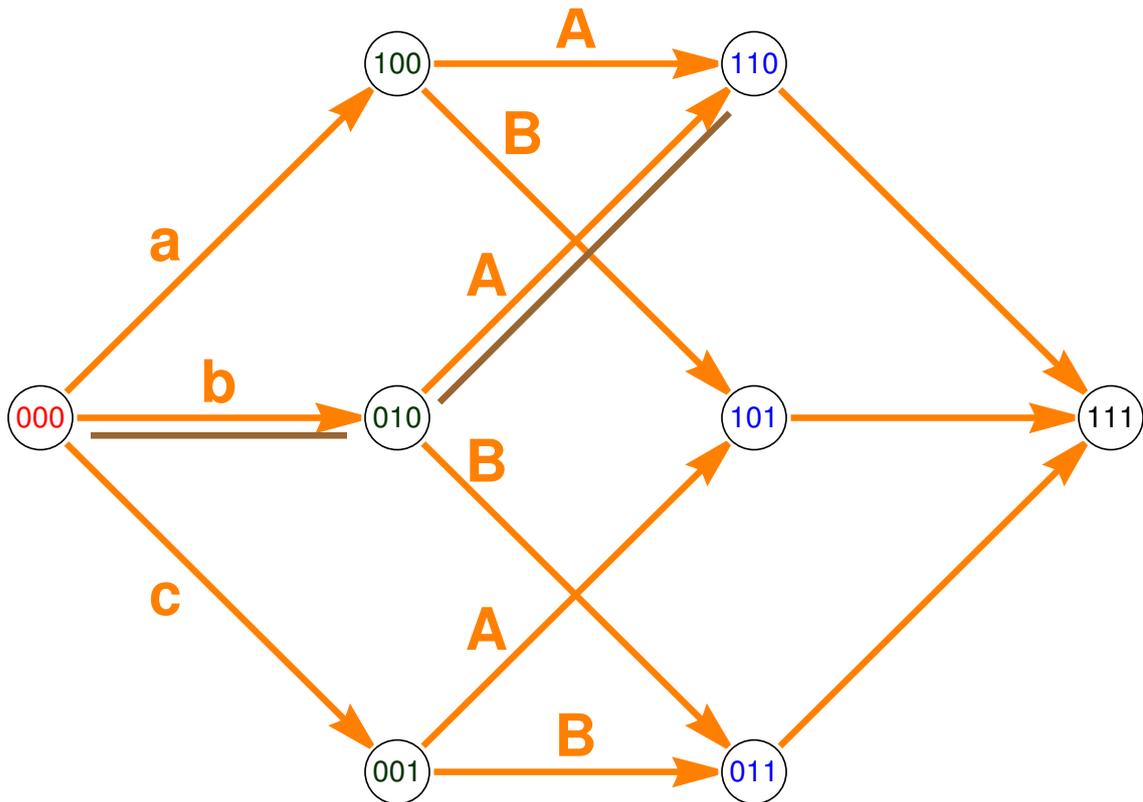


Figura 22 – Há seis trajetórias possíveis entre 000 e 111. O caminho destacado em marrom é bA.

classes com possíveis aptidões finais e iguais entre si, no caso presente, apenas classes com aptidão máxima. Isso pode significar que a população esteja concentrada em uma única classe, ou que esteja distribuída em um conjunto de classes de mesma aptidão. Em seguida, transformamos o vetor com o número de indivíduos por classe, em um vetor de probabilidades ao dividi-lo por N . Na primeira entropia obtemos inicialmente S por simulação da soma de $-p_i \log(p_i)$ de todas as classes finais, onde p_i é a razão do número de indivíduos de uma das classes dividido por N . Na sequência, calculamos a média de S entre as simulações. O desvio padrão é calculado a partir das médias de S de agregados de todos os dados (*binning*).

A segunda entropia S_T é calculada primeiramente calculando-se a média das distribuições e em seguida somando-se $-\langle p_i \rangle \log \langle p_i \rangle$, onde $\langle p_i \rangle$ é a média da razão do número de indivíduos de uma das classes dividido por N sobre as simulações. O desvio padrão é calculado a partir das médias de S_T de agregados de todos os dados.

A distribuição das populações nas diferentes trajetórias até o destino pode ser obtida analiticamente no regime SSWM. Ela corresponde ao produto da probabilidade P_i de cada passo i de uma dada trajetória possível. Essa probabilidade é proporcional ao peso W_i dado pelo produto da taxa de mutação para gerar um único mutante do tipo i

com a probabilidade de fixação na presença apenas de seleção deste único indivíduo.

$$W_i = r_i \frac{1 - e^{2s_i}}{1 - e^{2Ns_i}}, \quad (4.4)$$

onde r_i é a taxa de mutação do alelo selvagem para o mutante tipo i , e s_i é o coeficiente de seleção deste alelo mutante. A probabilidade do passo P_i é o peso W_i normalizado pelos pesos de todas os alelos mutantes disponíveis no mesmo passo.

$$P_i = \frac{W_i}{\sum_j W_j}, \quad (4.5)$$

onde j corresponde aos alelos disponíveis no passo. Por exemplo, para $G = 3$, na trajetória 3-2-1, temos os três alelos mutantes disponíveis no primeiro passo. A probabilidade deste passo seria

$$P_1 = \frac{W_3}{W_1 + W_2 + W_3}. \quad (4.6)$$

A probabilidade do segundo passo seria, neste caso

$$P_2 = \frac{W_2}{W_1 + W_2}. \quad (4.7)$$

Finalmente, a probabilidade do terceiro passo seria

$$P_3 = \frac{W_1}{W_1} = 1. \quad (4.8)$$

A probabilidade Pr_i da trajetória i é o produto das probabilidades de cada um de seus passos. Para a trajetória 3-2-1, sua probabilidade seria

$$\text{Pr}_i = P_1 P_2 P_3 = \frac{W_3}{W_1 + W_2 + W_3} \frac{W_2}{W_1 + W_2}. \quad (4.9)$$

4.3 Resultados

O primeiro resultado que mostramos é o da distribuição média das trajetórias quando toda a população alcança os destinos. As trajetórias desenhadas na Figura 22 podem ser também representadas pelo grafo de uma árvore, com as folhas correspondendo ao genótipo com todas as mutações possíveis. Para cada genótipo corresponde um fenótipo. Como nas folhas da rede o genótipo é igual, a evolução a partir do momento em que toda a população está nelas acontece com toda ela com a mesma aptidão, o que corresponde a uma evolução neutra. A probabilidade de prevalência de uma subpopulação neste regime é proporcional à fração de indivíduos nela. Assim, esta distribuição média pode ser interpretada como a probabilidade de prevalência de cada trajetória.

Note que aqui nada pode ser dito sobre as folhas individuais alcançadas em cada realização, exceto a média das distribuições dos estados sobre todas as realizações. Veremos mais tarde que as trajetórias no platô inicial da Figura 23 são alcançadas segundo o regime

de seleção forte-mutação fraca (SSWM, do inglês *strong selection-weak mutation*), um único estado por vez.

No regime imediatamente a seguir deste platô vemos que as probabilidades de fixação de cada trajetória estão mais heterogêneas: a trajetória 3-2-1 torna-se a mais provável, enquanto a trajetória 1-2-3 é a menos provável, junto com outras trajetórias. Neste regime esperamos que haja a coexistência de diferentes trajetórias, e que as populações que alcançam aptidões maiores prevaleçam mais rapidamente. De fato, a trajetória 3-2-1 é a que cresce mais rapidamente em aptidão, e a 1-2-3, mais devagar.

Um detalhe interessante é que a primeira trajetória a suceder a 1-2-3 em prevalência é, por um breve intervalo de mutabilidade, a 2-1-3. Esse resultado se parece com o de nosso artigo, em que três mutações em um único locus estão disponíveis. Para os valores de aptidão e de taxas de mutação usados, a prevalência das mutações se sucedem em ordem crescente de aptidão, ou decrescente de taxa de mutação. Ambos os resultados se explicam da mesma maneira: ao se aumentar a taxa de mutação um pouco, primeiramente mutantes do tipo intermediário coexistem com mutantes menos aptos, e vencem a competição. No nosso caso, para o segundo passo, o alelo tipo 1 prevalece sobre o tipo 3 porque a taxa para este último ainda é muito baixa.

Quando a trajetória 3-2-1 ascende à maior frequência, a trajetória 2-3-1 também ascende. Ela vence a trajetória 3-1-2 em uma janela significativa de mutabilidade provavelmente porque nela há competição entre mutantes de dois passos, que devem em algum momento surgir em número significativo e vencer a competição com a última trajetória. Com o aumento da mutabilidade além desta janela há uma alternância de prevalência entre essas duas trajetórias.

Finalmente, para taxas muito altas de mutação, o efeito da seleção parece diminuir. As probabilidades parecem convergir para valores semelhantes. Isso pode ser explicado pelo fato de que a mutações muito altas, o tempo para o crescimento de uma subpopulação é menor, e a seleção passa a ser menos relevante. O pico 3-2-1 diminui porque mais trajetórias alcançam o estado de aptidão máxima com mais rapidez.

Na Figura 24 comparamos as probabilidades de cada trajetória com as probabilidades esperadas se elas seguissem o regime SSWM, subtraindo-se uma probabilidade da outra. Vemos de imediato que todas as trajetórias concordam com o regime SSWM a baixas mutabilidades neste regime padrão ($N = 10^6$, $s = 0,01$, e $G = 3$). A validade deste regime se estende até mutabilidades de $r = 10^{-9}$ nas condições usadas. Em todas as condições escolhidas neste capítulo, exceto quando usamos $s = 5,0$, há concordância das curvas com o regime SSWM a baixas mutabilidades.

Deste gráfico, notamos também que podemos distinguir melhor as trajetórias 1-2-3, 1-3-2 e 2-1-3. Elas se distinguem melhor pelo quanto elas se distanciam do regime SSWM

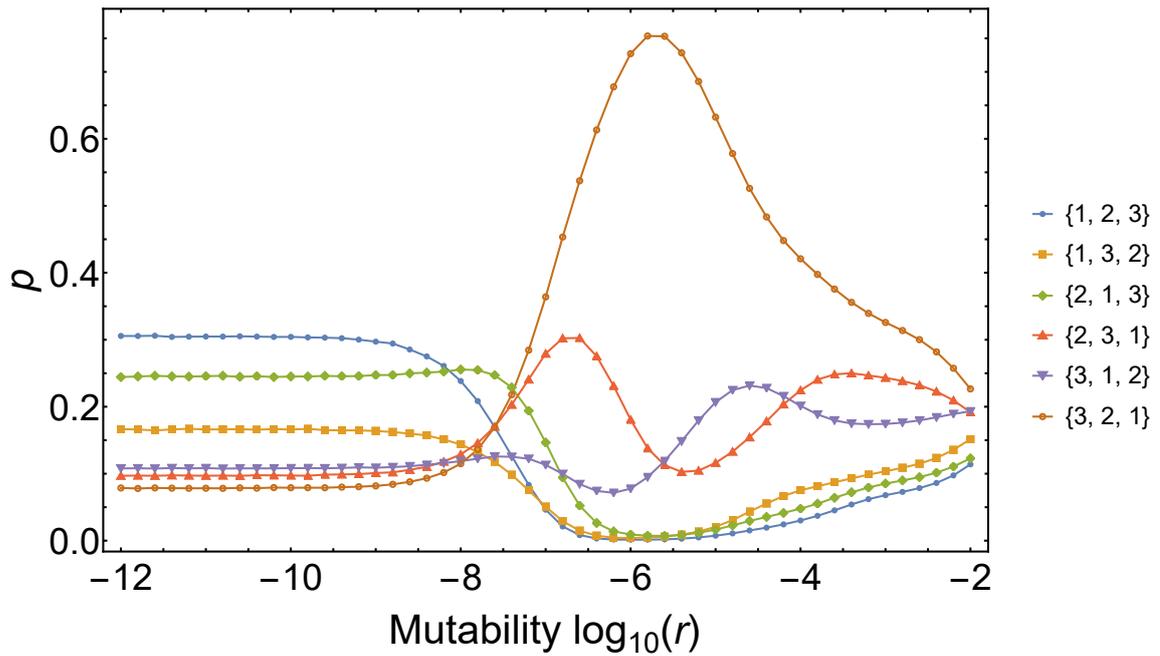


Figura 23 – Probabilidade de cada trajetória em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

com o aumento da mutabilidade. A trajetória 1-2-3 é a mais sensível, porque a probabilidade de fixação é alterada devido ao aumento do número de indivíduos desfavorecer passos iniciais menos adaptativos.

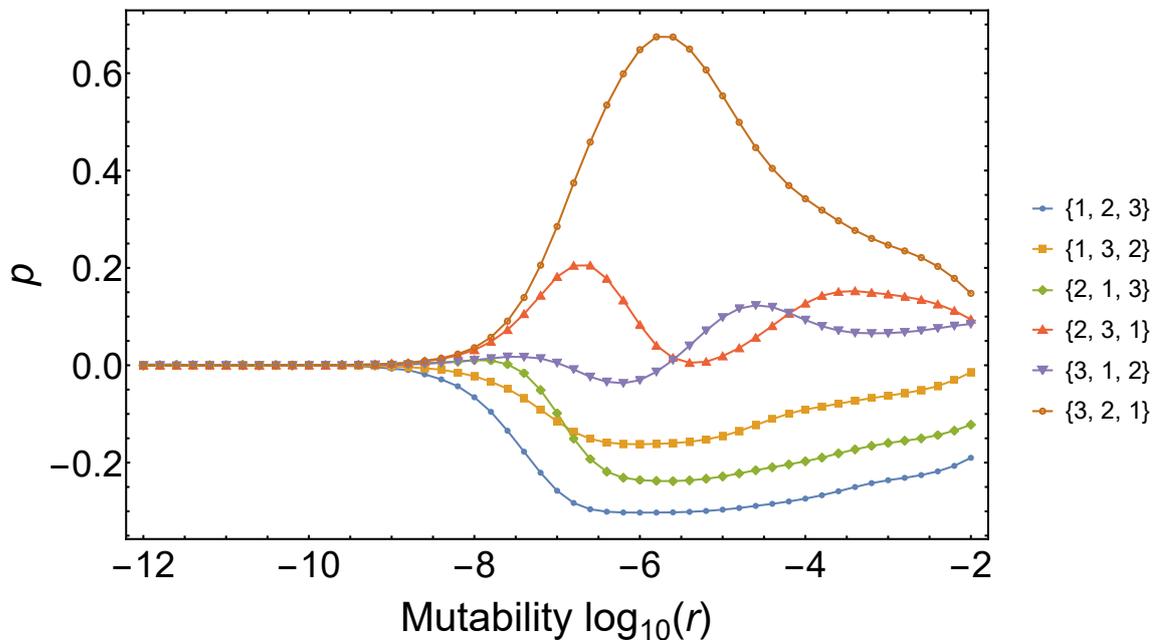


Figura 24 – Diferença da probabilidade de cada trajetória em função da mutabilidade para a probabilidade esperada no regime SSWM. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

Em seguida examinamos a distribuição das trajetórias a cada realização, calculando-

se a entropia de cada uma delas a partir do seu vetor probabilidade. Seu gráfico é mostrado na Figura 25. Vemos que claramente para um regime amplo de baixa mutabilidade a entropia é mínima. Ele corresponde ao regime SSWM na Figura 24. A interpretação disto é simples: de acordo com o regime SSWM, apenas uma trajetória é atravessada por realização, e isso é percebido pela entropia mínima por simulação. O regime SSWM só pode ser definido conjuntamente com as distribuições e a média da entropia por realização.

Com o aumentar da mutabilidade, mais estados são alcançados simultaneamente. Esse fenômeno acontece para todo o regime de mutabilidade estudado, até que a entropia alcance perto seu valor máximo, que equivale ao estado homogêneo. Isso pode ser visto pela convergência das probabilidades das trajetórias na Figura 23. Isso significa que as trajetórias tornam-se igualmente competitivas à mutabilidade máxima escolhida.

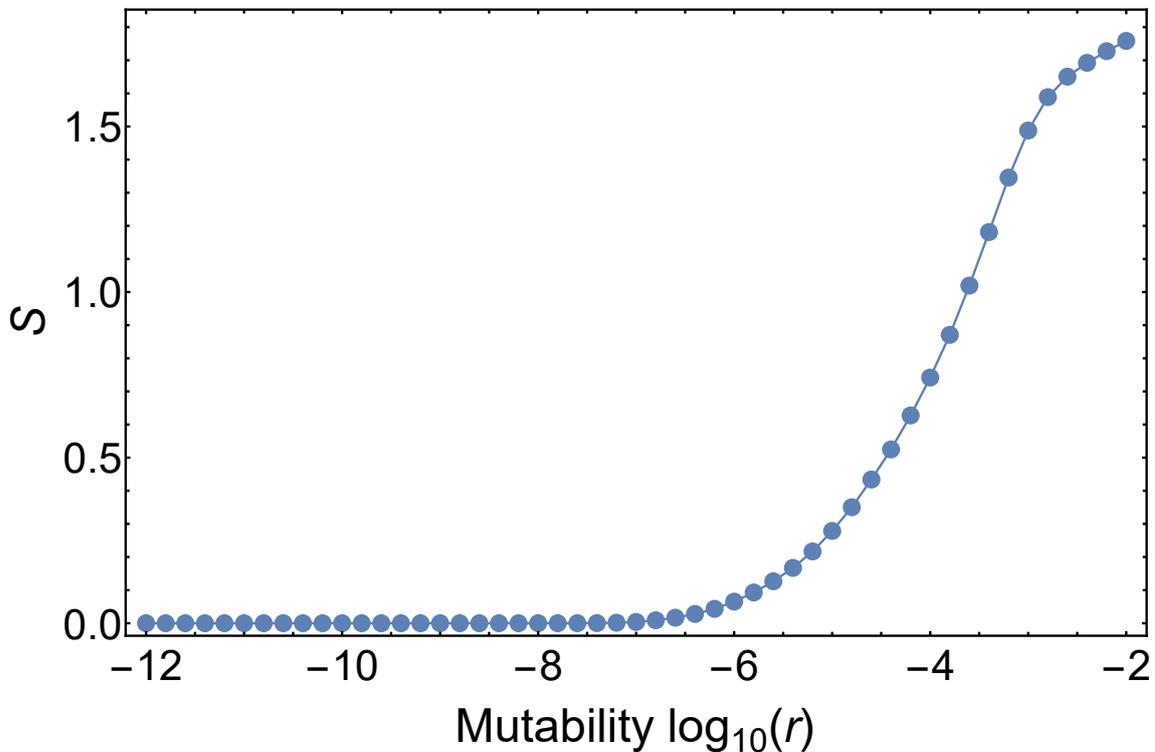


Figura 25 – Média das entropias S por realização, em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

Finalmente, na Figura 26 analisamos a entropia S_T da média das distribuições. Quando menor ela é, mais repetível é o processo adaptativo. Ela se inicia em um alto valor, mas apresenta um mínimo para mutabilidades intermediárias. Ele concorda com o pico 3-2-1 da Figura 23. O aumento da probabilidade de uma das trajetórias devido a seleção em detrimento das demais torna as realizações mais previsíveis, ou mais repetíveis, implicando em uma redução de S_T . Quando a distribuição de probabilidades por realização se torna mais homogênea, isso se reflete também em um aumento de S_T na direção da maior mutabilidade usada.

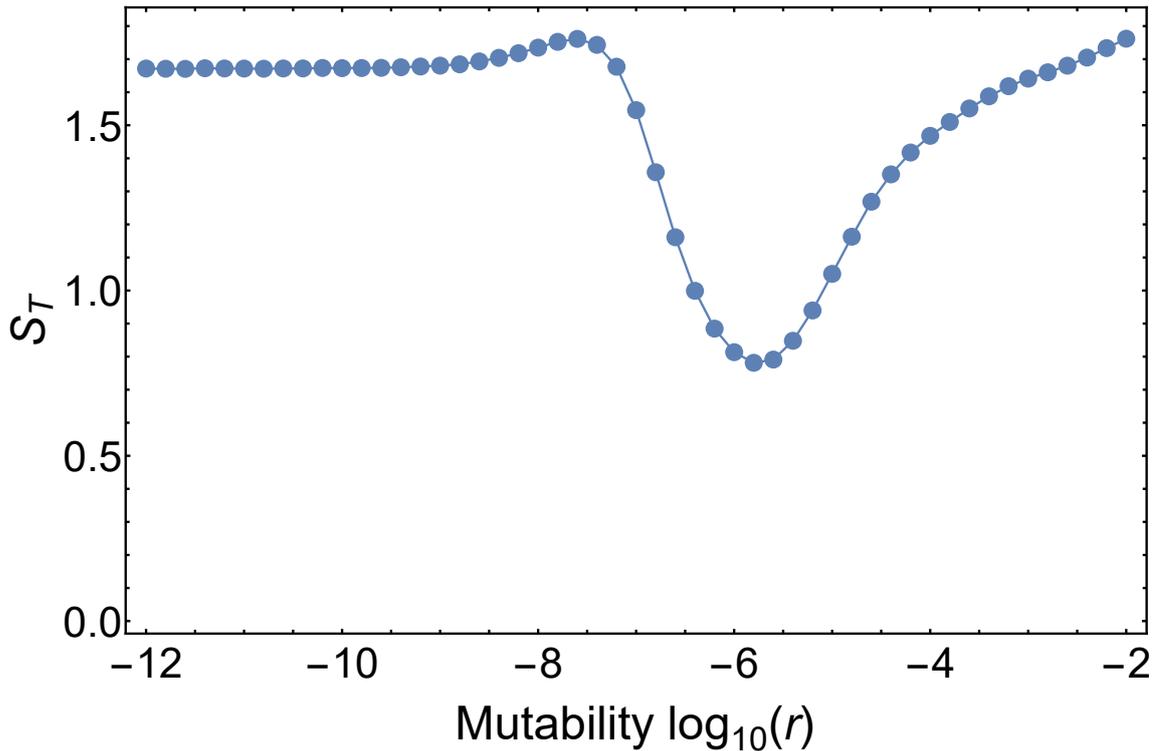


Figura 26 – Entropia S_T da distribuição média, em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

Agora vamos examinar as correlações entre as distribuições das trajetórias quando se aumenta a mutabilidade, mostradas na Figura 27. Notamos de início que todas as correlações entre as trajetórias são negativas, indicando uma competição entre elas pela preferência dos caminhos tomados. A correlação da trajetória 1-2-3 com quase todas as demais apresenta um pico onde praticamente há desconexão entre elas. Ela só apresenta uma correlação negativa no pico em relação a trajetória 3-2-1, a que mais afeta a frequência das trajetórias em geral.

Para a trajetória 1-3-2 também há picos em uma mutabilidade semelhante, e novamente a trajetória 3-2-1 é a que apresenta a correlação mais negativa no pico. Se observarmos todos os gráficos, notaremos que correlação mais negativa das trajetórias é com a trajetória 3-2-1, mais abundante. Quanto mais ela é enriquecida, mais as demais são depletadas.

Estudaremos agora cinco experimentos-controle em relação aos resultados com os parâmetros em seus valores “padrão”. No primeiro controle, as taxas de mutação são conservadas mas as aptidões de todos os genótipos são iguais. A evolução é neutra. O segundo controle relaxa a evolução neutra, atribuindo a cada mutação um mesmo coeficiente de seleção. A vantagem adaptativa situa-se entre as camadas da rede. O terceiro controle tem todas as taxas de mutação iguais e a evolução é neutra. No quarto controle, as taxas de mutação são iguais, mas as aptidões do genótipos são iguais à do experimento

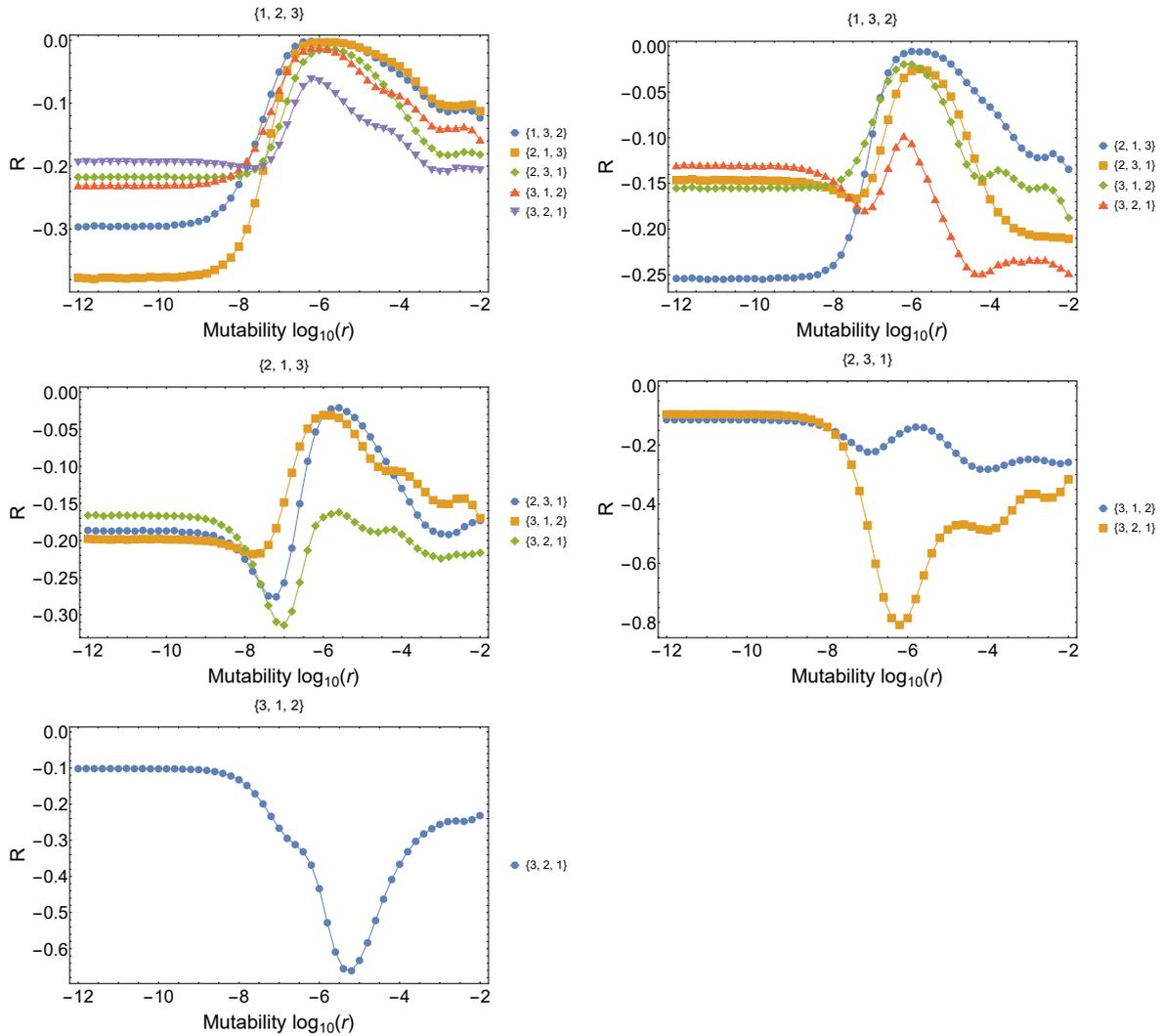


Figura 27 – Correlação entre as distribuições das trajetórias indicadas acima dos gráficos e das legendas, em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

padrão. Finalmente, no quinto e último controle, as taxas de mutação são iguais, e o coeficiente de seleção de cada mutação é igual.

Nos experimentos mostrados na Figura 28 notamos que praticamente não há influência da mutabilidade quando a evolução é neutra ou quando as trajetórias são iguais. O grande ruído notado nos controles com apenas $s = 0$ é devido a lentidão das simulações que possibilitou apenas um tamanho reduzido de amostra. Quando as taxas são diferentes e a evolução é neutra, a distribuição obedece a distribuição prevista apenas pelas taxas de mutação, mesmo quando as taxas são mais altas. A população de uma trajetória não compete com a outra, nem a deriva pode alterar a distribuição média.

Quando as taxas são iguais e a evolução é neutra todos os caminhos são equivalentes. A distribuição de cada caminho é de $1/6$. O mesmo acontece quando cada mutação contribui com um mesmo coeficiente de seleção.

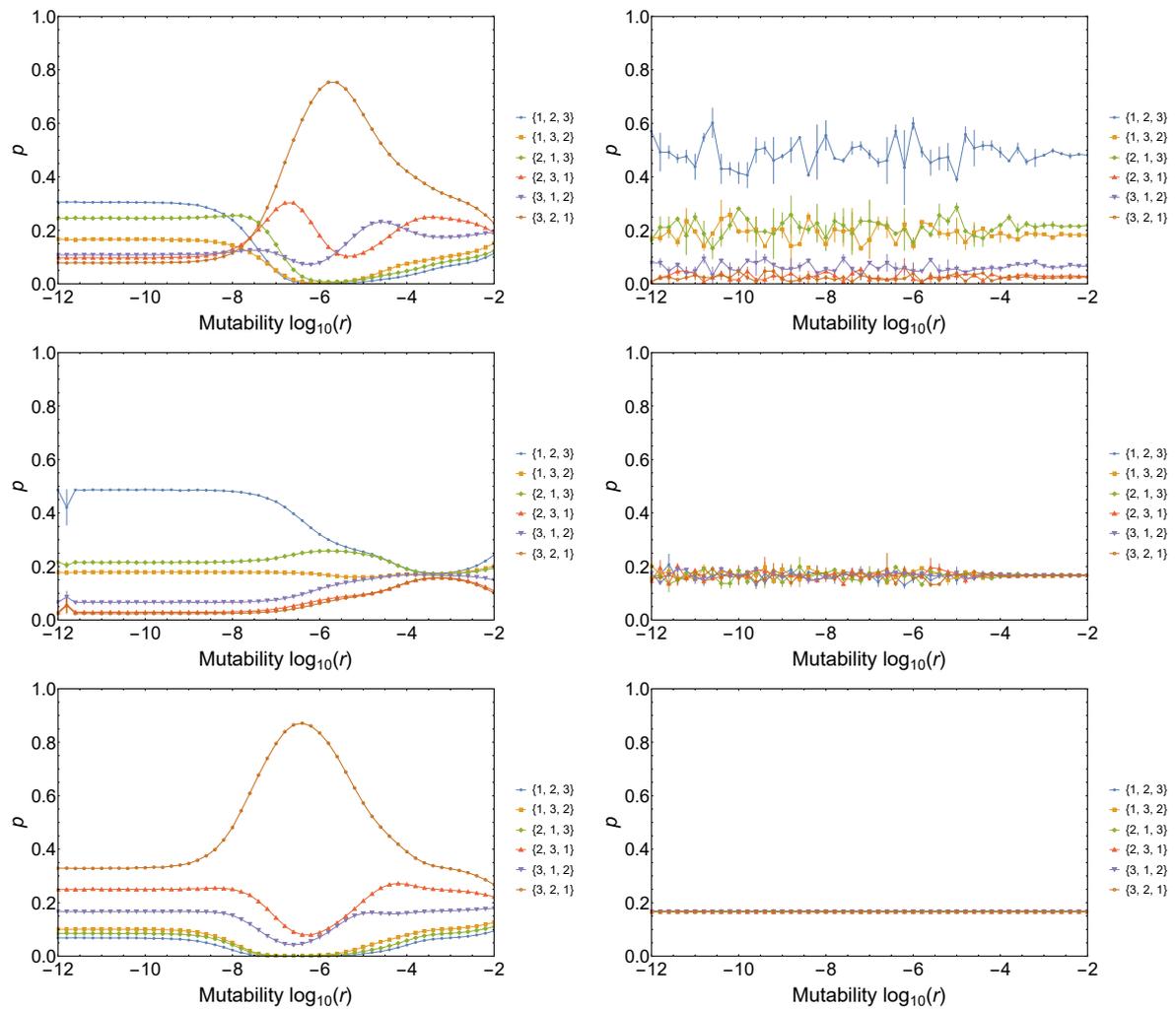


Figura 28 – Distribuições de probabilidade das trajetórias em diferentes experimentos. Da esquerda para a direita, e de cima para baixo: condições padrão; condições padrão exceto por aptidões neutras ($s = 0$); condições padrão com cada mutação contribuindo com o mesmo coeficiente de seleção ($s = 0,01$); as taxas de mutação são iguais e a evolução é neutra ($s = 0$); as taxas de mutação são iguais, mas os coeficientes de seleção dependem do locus, como originalmente; e as taxas de mutação são iguais, e cada mutação contribui com o mesmo coeficiente de seleção ($s = 0,01$). Nas condições padrão, o tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico, exceto para simulações neutras onde foram realizadas 2048 simulações.

Quando as taxas de mutação são diferentes, mas as contribuições de cada mutação são iguais, existe uma alteração de cada trajetória com a mutabilidade. Mas aparentemente, o *ranking* das trajetórias de acordo com sua frequência não muda. Quando as taxas de mutação e as contribuições para a aptidão de cada mutação são iguais, este *ranking* também não muda. Portanto, para haver mudança das frequências das trajetórias com a mutabilidade, é preciso haver mutações adaptativas, e para haver mudança de *ranking*, é preciso haver aptidões diferentes para genótipos com o mesmo número de mutações.

Um trabalho anterior nosso sugere que as trajetórias evolutivas mudam em preferência quando o coeficiente de seleção de uma mutação e sua taxa de mutação estão em oposição, i.e. quando maior o coeficiente de seleção, menor sua taxa de mutação. É o que ocorre no experimento padrão, e isso deve ser a verdadeira explicação para a mudança de *ranking*.

Agora compararemos a entropia por realização entre os diferentes controles a partir da Figura 29. Vemos que nesta Figura todas as entropias aumentam com a mutabilidade. Isto significa que mais trajetórias distintas são alcançadas ao mesmo tempo. Outra característica é que todas as entropias crescem até seu valor máximo, exceto o experimento de evolução neutra com taxas diferentes. A entropia máxima deste último experimento é próxima ao valor esperado dos indivíduos seguirem independentemente suas trajetórias. Os demais processos homogenizam a distribuição das trajetórias.

Outro ponto interessante é que uma homogeneização do processo padrão, seja através de evolução neutra, de mutações com s iguais ou de taxas de mutação iguais, leva a uma entropia maior para um regime significativo de mutabilidade, o que significa estados finais mais homogêneos. O processo mais homogeneizante é o de taxas de mutações iguais e evolução neutra, seguido do processo de taxas iguais e mutações com o mesmo s . Este último homogeneiza menos porque aqui as trajetórias competem umas com as outras via aptidão.

Finalmente, examinamos na Figura 30 a repetibilidade dos processos em questão examinando-se a entropia da média das distribuições por realização. Percebemos imediatamente que o processo com taxas iguais e neutro, e o com taxas iguais e mutações com mesmo s têm entropia máxima, o que é esperado porque todos os seus caminhos são equivalentes. O processo com taxas de mutação padrões e neutro tem uma entropia que aparentemente não depende da mutabilidade, e se comporta de maneira compatível ao processo em que os indivíduos são independentes seguindo as rotas evolutivas exceto pela deriva. Como ela não muda as frequências finais de um processo neutro, o efeito é o mesmo.

O processo com taxas de mutação padrões e mutações com mesmo coeficiente de seleção tem uma entropia em dois platôs, que aumenta com a mutabilidade. Essa mudança coincide com a diminuição da frequência do caminho 1-2-3, que passa ser relativamente

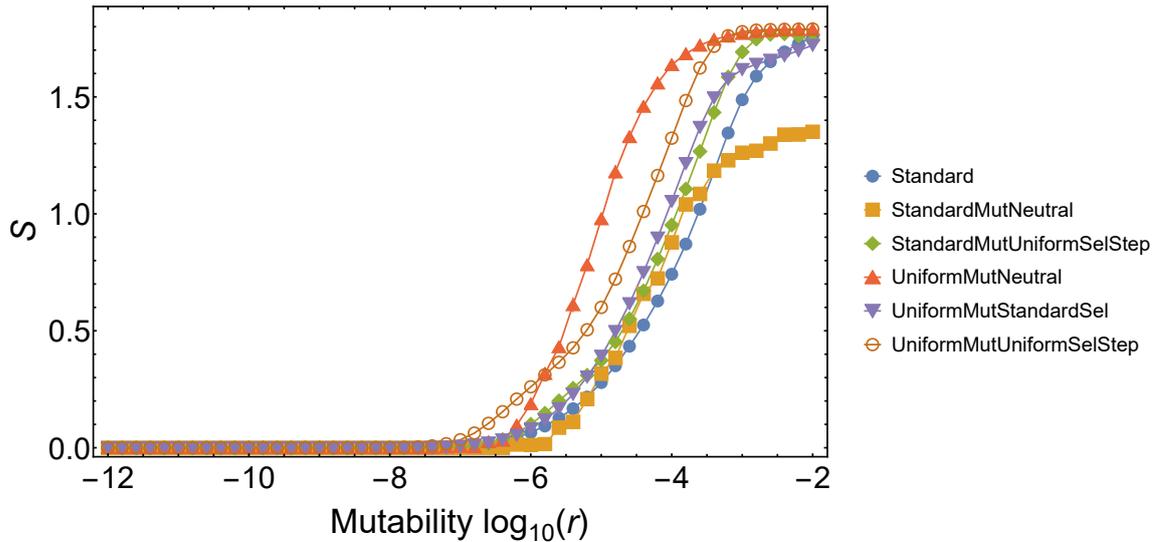


Figura 29 – Média das entropias por realização S no experimento padrão e nos controles. A legenda representa: "Standard" são condições padrão; "StandardMutNeutral" corresponde a condições padrão exceto por aptidões neutras ($s = 0$); "StandardMutUniformSelStep" corresponde a condições padrão com cada mutação contribuindo com o mesmo coeficiente de seleção ($s = 0,01$); "UniformMutNeutral" corresponde a taxas de mutação são iguais e a evolução neutra ($s = 0$); no "UniformMutStandardSel" as taxas de mutação são iguais, mas os coeficientes de seleção dependem do locus, como originalmente; e no "UniformMutUniformSelStep" as taxas de mutação são iguais, e cada mutação contribui com o mesmo coeficiente de seleção ($s = 0,01$). No experimento padrão, o tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico, exceto para simulações neutras onde foram realizadas 2048 simulações.

menos competitivo que os demais (ver Figura 28). Finalmente, as duas últimas curvas (com aptidões padrão) que tem a entropia mudando com a mutabilidade têm um mínimo para mutabilidade intermediária devida ao pico 3-2-1. Esse mínimo acontece com mutabilidade menor quando as taxas são iguais porque a taxa para a mutação 3 é bem maior que no caso padrão, acelerando o favorecimento do pico.

Fizemos também simulações para um modelo mais geral da relação entre mutabilidade e coeficiente de mutação adaptativa, chamado de modelo de Gumbel. Ele inclui o nosso caso exponencial para $\beta = 1$. Não mostramos os resultados aqui porque eles são bem semelhantes ao caso exponencial, exceto para o maior valor usado de $\beta = 2$. Concluimos disso que nossos resultados são robustos em relação à forma como a taxa de mutação decai monotonicamente com o coeficiente de seleção.

Discutiremos agora como a evolução adaptativa muda com o tamanho da população. Estudamos tanto redes com $G = 3$ como com redes $G = 5$, mas para estudarmos como a distribuição das trajetórias muda com a mutabilidade, verificaremos o comportamento apenas para $G = 3$ (que possui 6 trajetórias possíveis), porque o gráfico não seria claro o

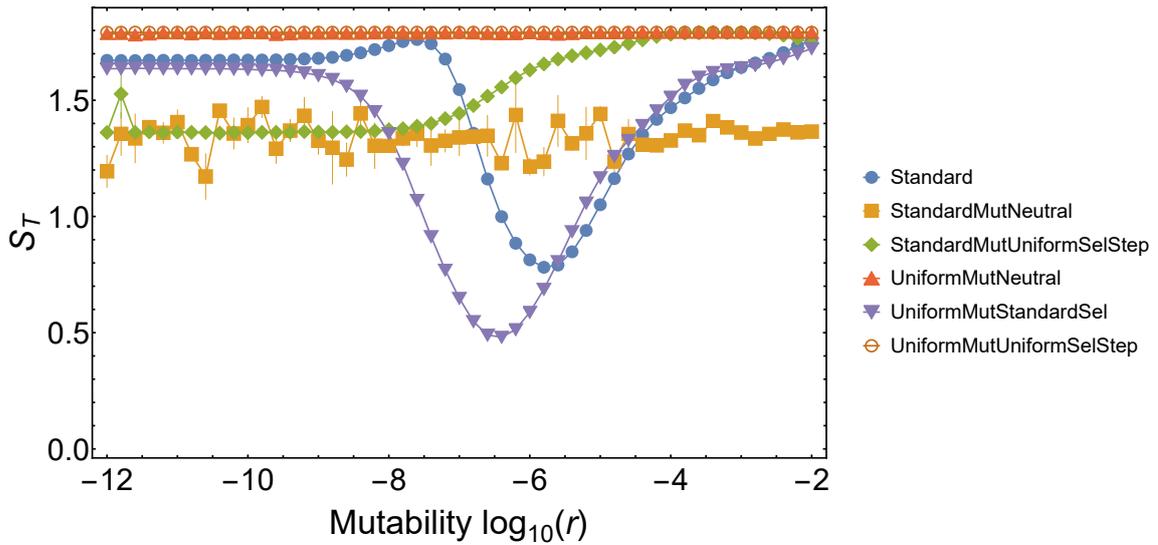


Figura 30 – Entropia da média das distribuições de cada realização em função da mutabilidade. O tamanho da população é $N = 10^6$, a rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

suficiente para representar as 120 trajetórias com $G = 5$.

Como esperado, a Figura 31 mostra que ao aumentarmos o tamanho da população, o regime SSWM encolhe, enquanto o regime em que a seleção torna a distribuição menos homogênea alarga-se para valores menores de r . Para $N = 10^9$ praticamente não há o regime SSWM. Isso é compreensível, porque o aumento do tamanho da população aumenta o suprimento mutacional (o produto das taxas de mutação e o tamanho da população). Duas trajetórias coexistem mais facilmente e competem entre si via aptidão, tornando a distribuição final menos homogênea.

Interessantemente, o *ranking* entre as trajetórias em termos de frequência não parecem mudar com N , mas a repetibilidade do processo evolutivo parece se tornar maior com o aumento de N . O pico 3-2-1 prevalece mais com o aumento de N porque a taxa de mutação menor da mutação 3 é um gargalo cada vez menos eficiente com a maior população. Adicionalmente ao fato de que os caminhos evolutivos experimentais são bastante restringidos por epistasia de sinal (NGHE; KOGENARU; TANS, 2018; CHIOTTI et al., 2014; LALIC; ELENA, 2012; SILVA et al., 2011; KVITEK; SHERLOCK, 2011; DAWID et al., 2010; ORGOGOZO; BROMAN; STERN, 2006; WEINREICH; WATSON; CHAO, 2005), eles parecem ser também muito mais restringidos quando há mais competição entre trajetórias em populações grandes.

Vamos examinar agora como a entropia por realização muda com o tamanho da população. O resultado a ser discutido é mostrado na Figura 32. Vemos aqui que quanto maior o tamanho da população, maior é a entropia, o que é esperado porque quanto mais indivíduos uma população tem, mais acessíveis se tornam simultaneamente as trajetórias. Para $N = 10^3$ a entropia só começa a crescer de seu valor mínimo a partir de uma

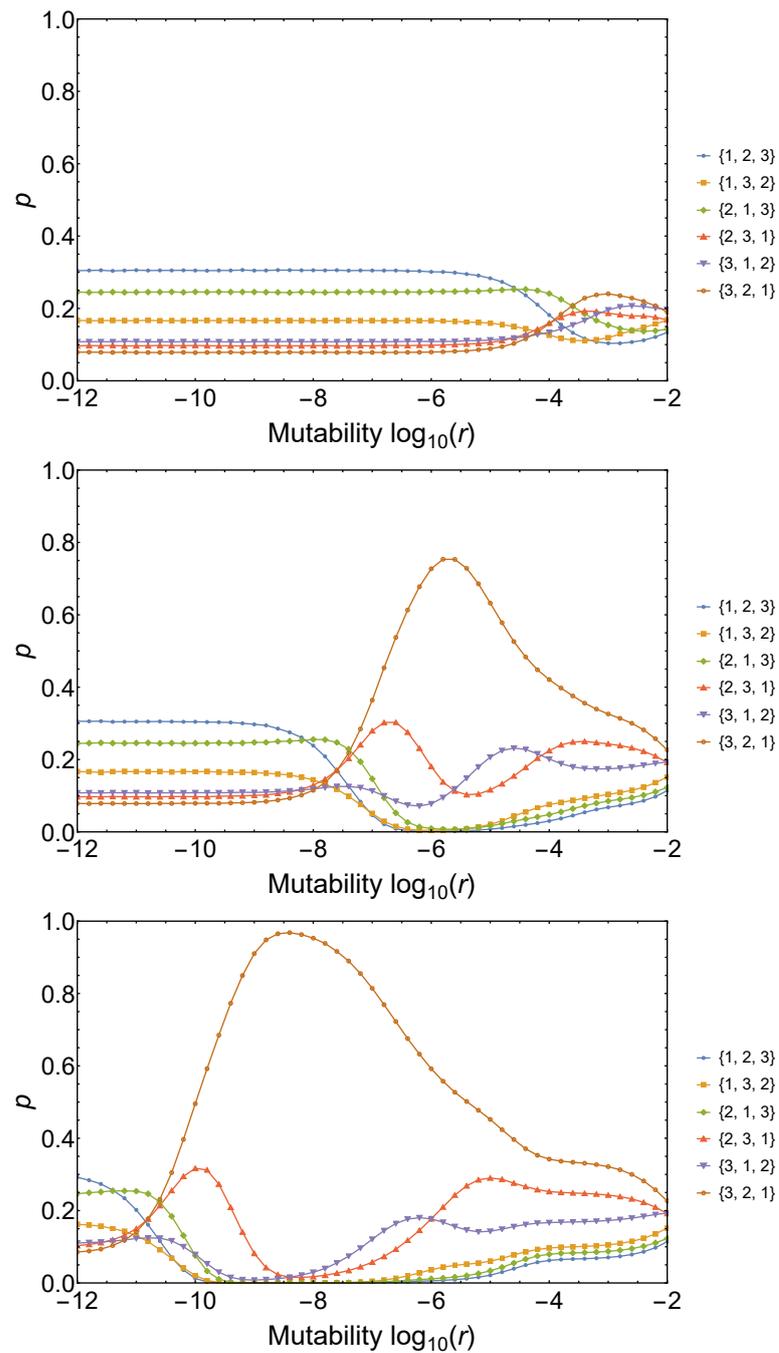


Figura 31 – Distribuições de probabilidade das trajetórias com diferentes tamanhos de população para redes completas com três locos. De cima para baixo: $N = 10^3$, $N = 10^6$ e $N = 10^9$. A rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

mutabilidade alta. Isso se deve a essa população permanecer no regime SSWM num regime muito mais amplo (compare com a Figura 31).

Outro fato interessante é que para $N = 10^9$ a entropia cresce até uma fase de crescimento mais lento, e concorda com a curva para $N = 10^6$ para a mutabilidade mais alta. Este é um sinal de saturação da entropia, que possivelmente ocorreria para uma população infinita. Isso implica uma dependência da entropia com a mutabilidade mesmo para populações que restringem ao mínimo a travessia das possíveis trajetórias. O tempo que um indivíduo leva em média para sofrer uma mutação é importante para a competição entre os genótipos, e conseqüentemente para a heterogeneidade das trajetórias usadas.

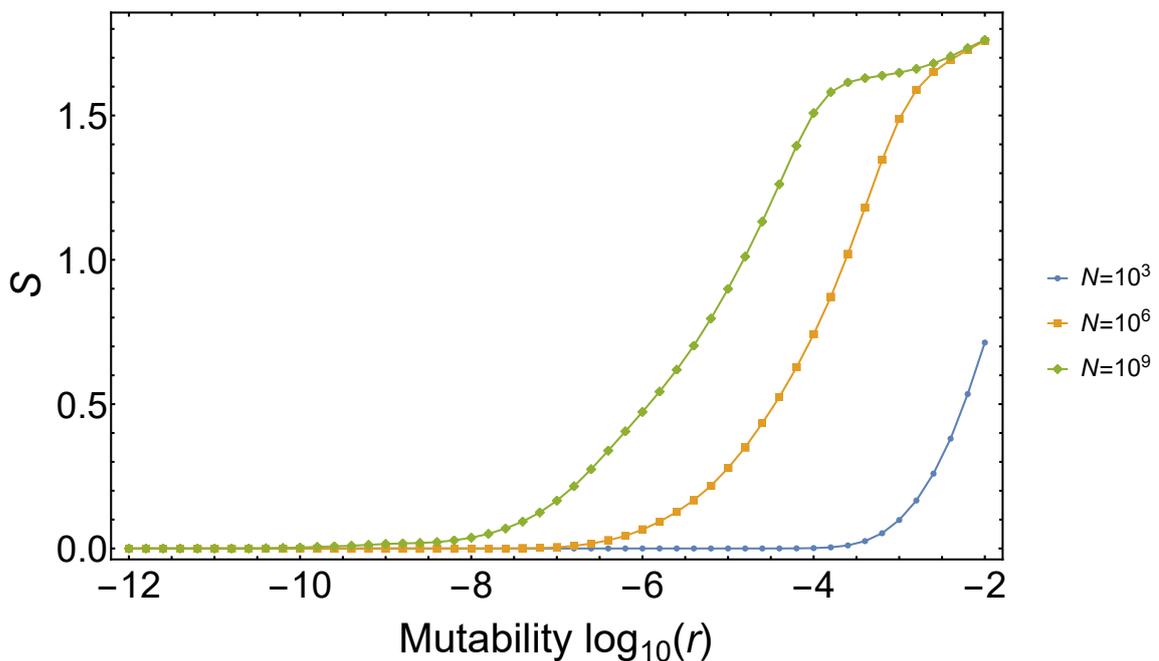


Figura 32 – Média das entropias das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

Agora vamos examinar a repetibilidade do processo para diferentes tamanhos de população, a partir da Figura 33. Para $N = 10^3$ a repetibilidade permanece baixa (entropia alta) para toda a faixa de mutabilidade estudada. Ela aumenta um pouco com a mutabilidade, sem a existência de um mínimo intermediário. Isto coincide com o fato da sequência 3-2-1 não possuir um pico, pela população permanecer essencialmente no regime SSWM por quase toda a mutabilidade estudada.

Para o tamanho de população $N = 10^6$ o mínimo da entropia para mutabilidade intermediária é pronunciado, e ao aumentarmos o tamanho para $N = 10^9$ este mínimo tanto diminui drasticamente como ele acontece em mutabilidade menor. Isso é explicado pelo desvio e acentuação do pico da trajetória 3-2-1 para mutabilidades menores, conforme discutido anteriormente. É possível perceber-se um comportamento assintótico pela concordância das curvas para estes dois N para mutabilidades extremamente altas, reflexo do

comportamento para S na Figura 32.

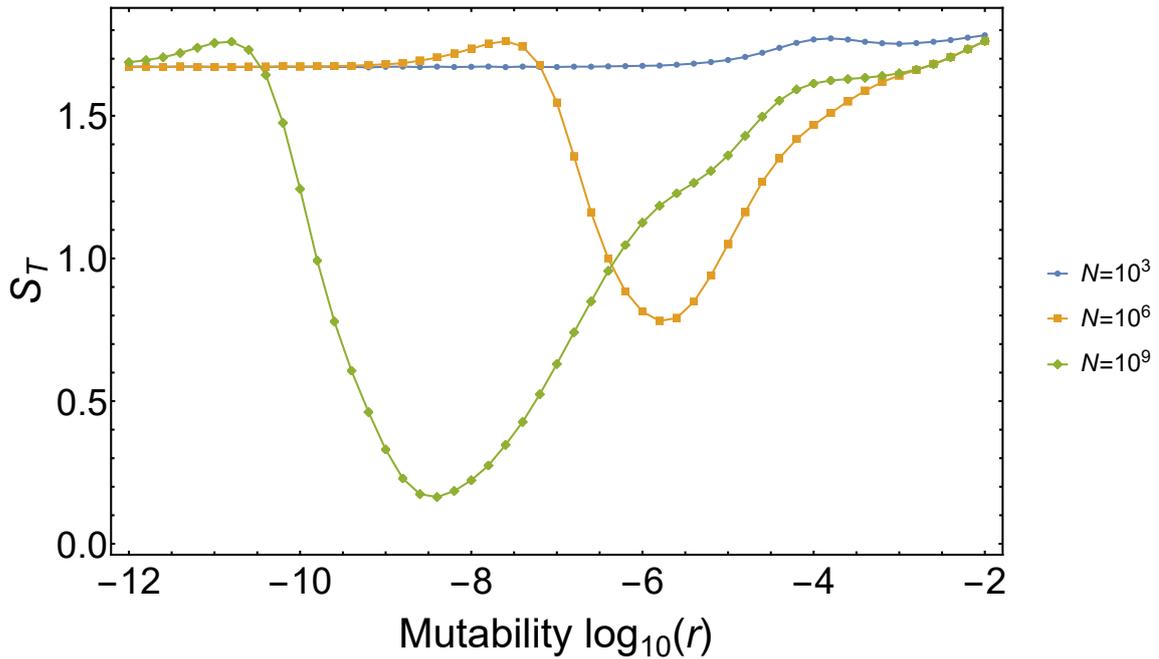


Figura 33 – Entropias da média das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 3$ loci, $\sigma = 0,01$ e $\beta = 1$. Foram realizadas 524.288 simulações por ponto do gráfico.

Agora iremos estudar estas entropias para um tamanho maior de rede, $G = 5$. Primeiramente examinaremos a entropia média S em função da mutabilidade, mostrado na Figura 34. De imediato, percebemos que esses valores são bem mais altos, porque o número de trajetórias possíveis é bem mais alto (120 para $G = 5$, ao invés de 6 para $G = 3$). Também notamos que a entropia S aumenta com a mutabilidade e com o tamanho da população. Ela parece se saturar para $N = 10^9$ e mutabilidade máxima, o que significa que a distribuição por realização é próxima da uniforme neste caso. Diferentemente do caso $G = 3$, a população $N = 10^6$ não satura sua entropia. Parece em geral que para redes maiores é mais difícil saturar a entropia por realização com o aumento de N .

A repetibilidade das trajetórias adaptativas para $G = 5$, medida por S_T e mostrada na Figura 35, segue qualitativamente o mesmo padrão que para $G = 3$. Para $N = 10^3$ não há um mínimo intermediário, enquanto os mínimos intermediários para $N = 10^6$ e $N = 10^9$ se deslocam para mutabilidades mais baixas e são cada vez menores com o aumento de N . A razão é análoga: há um pico (não mostrado) com trajetória onde os coeficientes de seleção de cada mutação decrescem com a ordem da mutação (5-4-3-2-1) que é proeminente para estes tamanhos de rede (alcança cerca de $p = 0,2$ para $N = 10^6$ e cerca de $p = 0,6$ para $N = 10^9$) em mutabilidades cada vez menores com o aumento de N . A explicação para isso deve ser a mesma que para $G = 3$. Os regimes de mutabilidade para os quais a SSWM é válida são bem semelhantes entre os diferentes tamanhos.

Em seguida analisamos o efeito da pressão seletiva mostrado na Figura 36. As

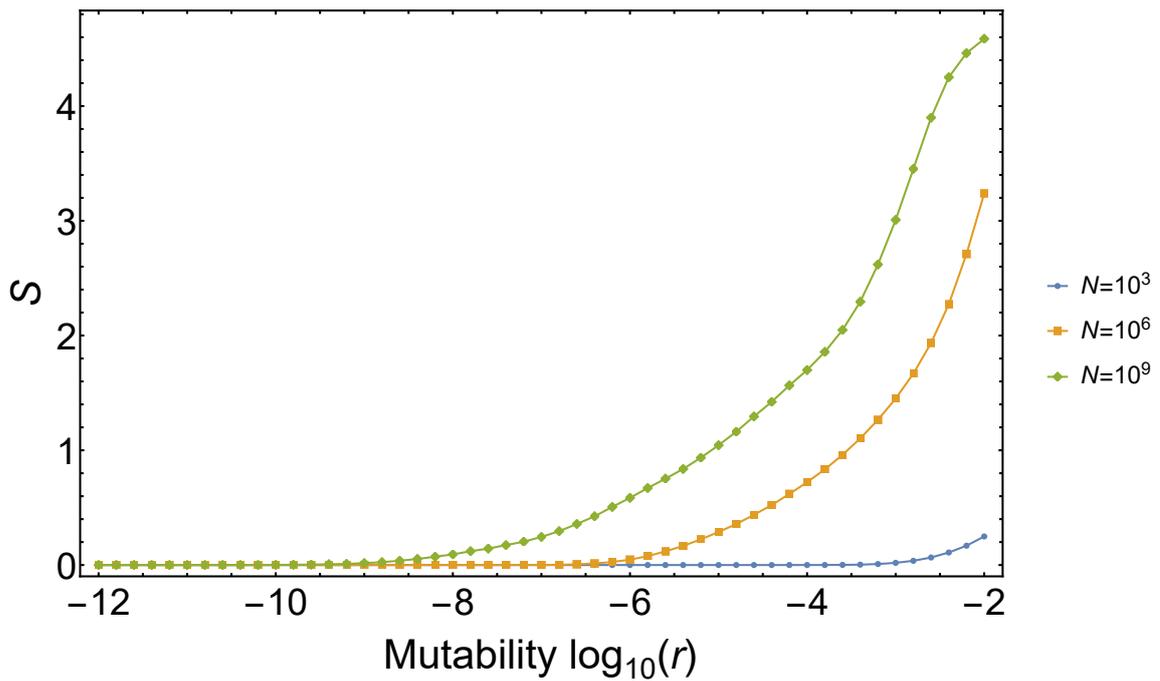


Figura 34 – Média das entropias das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 5$ loci, $\sigma = 0,01$ e $\beta = 1$.

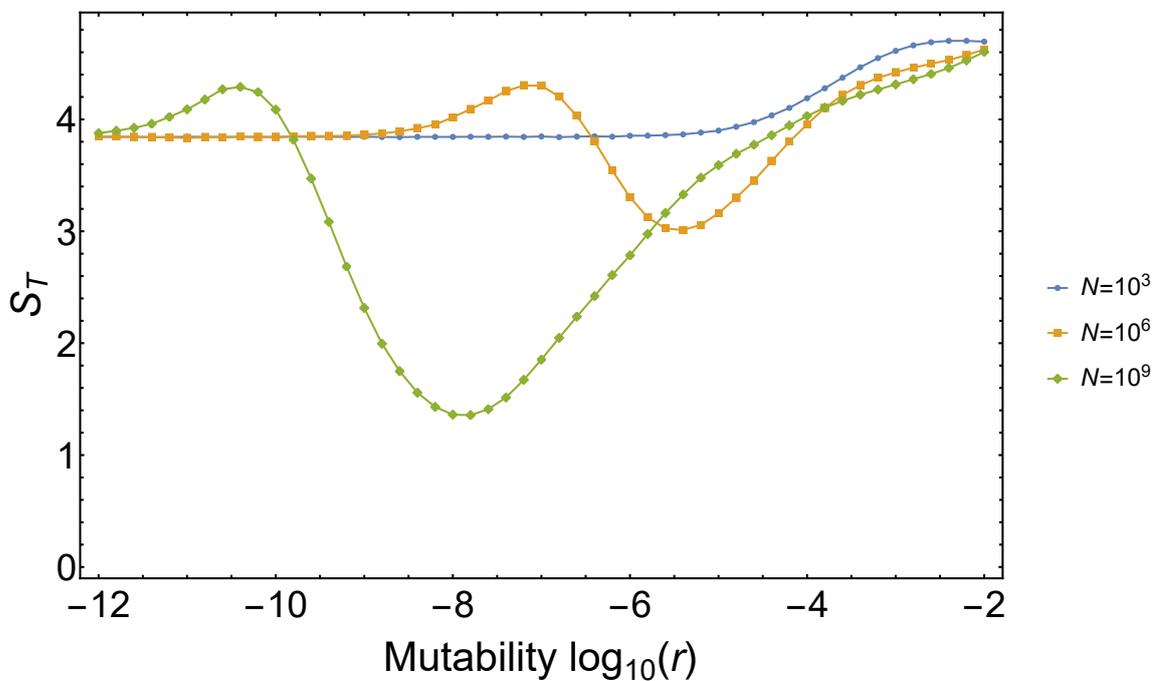


Figura 35 – Entropia da média das distribuições de cada realização S_T em função da mutabilidade para diferentes tamanhos de população. A rede tem $G = 5$ loci, $\sigma = 0,01$ e $\beta = 1$.

distribuições de trajetórias com $s = 0,005$ até $s = 0,04$ são bastante parecidas. Para o valor irrealista de $s = 5,0$ é que elas mudam significativamente. Neste valor de s , basta que haja apenas um indivíduo com uma mutação extra para ela se fixar com probabilidade bem próxima de um. Essas distribuições não serão mostradas aqui. Mostraremos imediatamente o efeito da seleção sobre as entropias S e S_T .

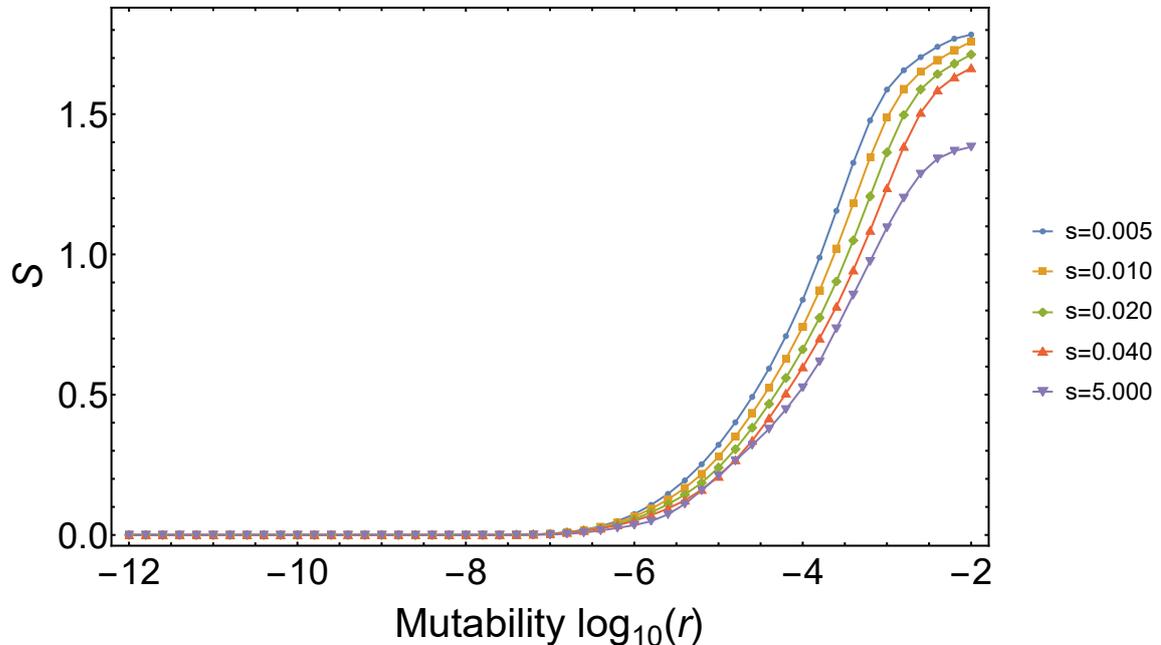


Figura 36 – Média da entropia das distribuições de cada realização S em função da mutabilidade para diferentes coeficientes de seleção médios. A rede tem $G = 5$ loci, $\beta = 1$ e o tamanho da população é $N = 10^6$.

Na Figura 36 vê-se claramente que quanto menor o coeficiente de seleção maior é a entropia S . Isto é semelhante ao que vimos ao compararmos a evolução neutra com o processo adaptativo padrão. Supomos que devido ao valor drástico de $s = 5,0$ sua curva de entropia deve próxima do limite inferior. O regime de mutabilidade válido para SSWM parece não se alterar muito com s .

Quando examinamos S_T , ela parece diminuir verticalmente com o aumento de s em valores típicos. A redução desta entropia total deve ser reflexo do aumento do pico 3-2-1, que existe devido à sua pressão seletiva maior sobre o processo. O aumento de s deve portanto aumentar este pico.

Examinaremos agora o efeito do tamanho de rede sobre as entropias S e S_T . Na Figura 37 vemos que a entropia S aumenta de $G = 3$ para $G = 4$, mas diminui em geral subsequentemente. Parece haver uma tensão entre o número de trajetórias disponível e a seleção que ocorre enquanto cada trajetória é atravessada. Portanto, de $G = 3$ para $G = 4$ o aumento de S é devido à prevalência do aumento do número de trajetórias, enquanto de $G = 4$ para $G = 5$ em diante, o tempo gasto entre o estado selvagem e o totalmente adaptado ou as diferenças maiores de aptidão são os responsáveis pela prevalência de redes

menores em termos de entropia.

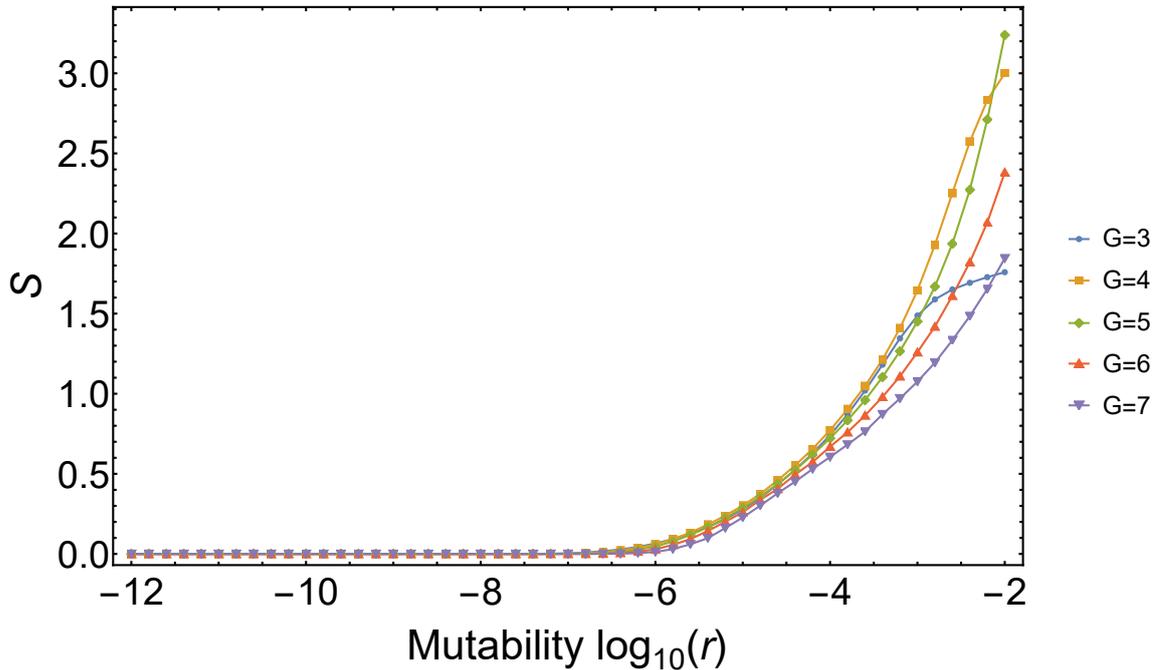


Figura 37 – Média da entropia das distribuições de cada realização S em função da mutabilidade para diferentes tamanhos de rede G . O tamanho da população é $N = 10^6$, $\beta = 1$ e $\sigma = 0,01$.

Quando olhamos para a repetibilidade na Figura 38, é claro que quanto maior o tamanho da rede, menor é a preditibilidade do processo adaptativo. Com o aumento de G , o mínimo global passa a ser um mínimo local para $G = 7$. Talvez o pico de s decrescente (no caso, 7-6-5-4-3-2-1) não seja tão importante como os demais porque as diferenças nas taxas de mutação sejam mais significativas. A taxa para o alelo 7 é extremamente baixa, afetando este pico.

4.4 Conclusão

Verificamos que existem basicamente três regimes de mutabilidade onde o padrão de trajetórias é alterado: um a baixas mutabilidades (SSWM), onde as taxas de mutação têm um papel importante, outro a mutabilidade intermediária, onde a seleção natural prepondera e aumenta a repetibilidade do processo evolutivo, e um terceiro para mutabilidades extremamente altas, em que as altas taxas de mutação não permitem muito tempo para as populações crescerem antes de alcançarem a aptidão máxima.

Progressivamente, ao aumentarmos a mutabilidade, migramos do regime de múltiplas mutações para o de interferência clonal. Isso se vê pelas alternâncias de prevalência das sequências 1-2-3, 2-1-3 e 3-2-1. No regime de baixas mutabilidades, SSWM implica que toda a população ocupa um só estado por vez, seguindo uma única trajetória até o estado final por realização. Mas a cada realização a trajetória pode ser distinta.

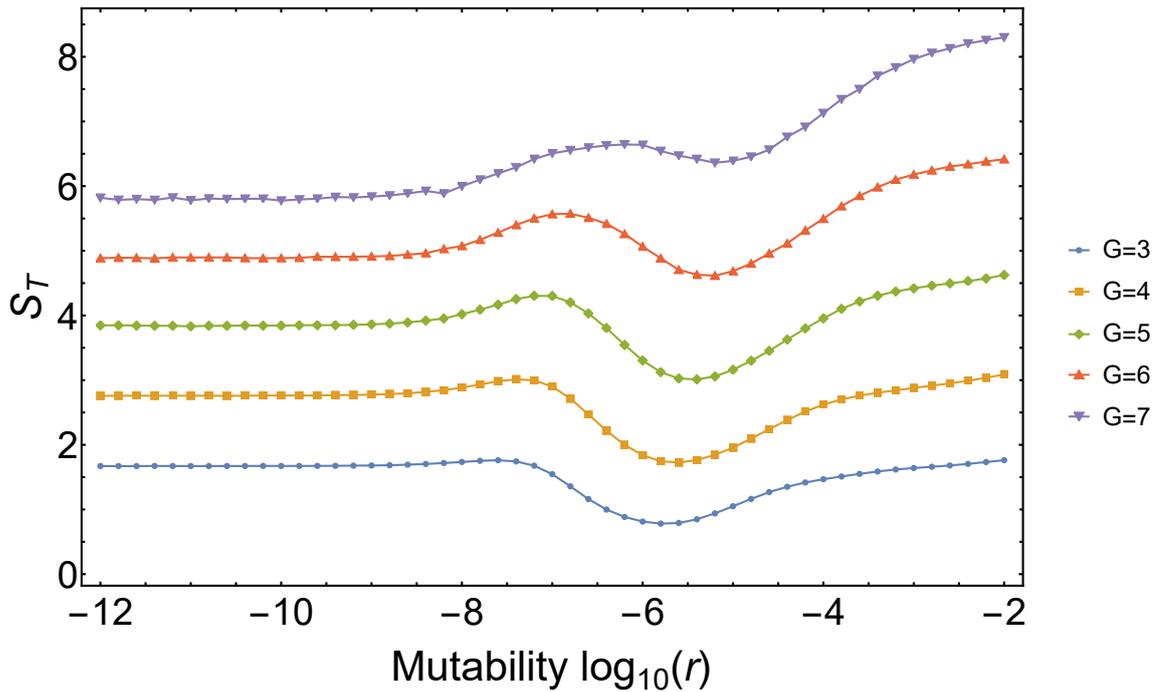


Figura 38 – Entropia da média das distribuições de cada realização S_T em função da mutabilidade para diferentes tamanhos de rede. O tamanho da população é $N = 10^6$, $\beta = 1$ e $\sigma = 0,01$.

A maior repetibilidade do processo evolutivo está associada à maior eliminação de trajetórias alternativas por competição da trajetória que aumenta a aptidão mais rapidamente (3-2-1) com elas. Isso pode ser percebido também com o aumento do tamanho da população e com o aumento de s .

Os controles atestam para a qualidade da simulação principal e revelam alguns fatos interessantes. Eles também apontam para o papel da seleção em reduzir a entropia S_T para mutabilidades intermediárias, e indicam que a mudança de *ranking* entre as trajetórias com a mutabilidade depende de aptidões distintas associadas a genótipos distintos, aparentemente em oposição com as respectivas taxas de mutação, como demonstrado no capítulo 2.

Nossos resultados também não dependem de detalhes específicos da distribuição de efeitos de aptidão, como quando usamos o modelo de Gumbel com diferentes parâmetros. E essencialmente os mesmos fenômenos vistos para $G = 3$ foram vistos para $G = 5$. No entanto, a taxa de mutação para a mutação mais apta parece ter um papel em permitir que haja um mínimo de entropia S_T no regime de maior seleção.

5 Destino evolutivo

5.1 Introdução

No capítulo anterior, vimos que a mutabilidade altera a ordem com que mutações adaptativas aditivas ocorrem até alcançar-se um genótipo final. A história evolutiva portanto depende das taxas de mutação de uma população.

Verificaremos neste capítulo um outro sentido em que a história evolutiva depende das taxas de mutação: quando as mutações acumuladas, ou o destino evolutivo ou genótipo final pode ser modificado. Vimos este efeito em miniatura no capítulo 3, quando apenas dois passos evolutivos estão disponíveis. Agora observaremos se existe algum efeito semelhante quando existe uma rede maior de mutações adaptativas e genótipos.

Para que haja mais de um genótipo-alvo disponível em uma rede de mutações adaptativas, os fenótipos associados aos genótipos precisam apresentar epistasia, isto é, uma interação tal entre as mutações envolvidas que a aptidão não seja aditiva (i.e. o coeficiente de seleção resultante não é uma soma dos coeficientes de seleção de cada mutação isolada).

Trabalharemos com a epistasia aleatória (NEHER; SHRAIMAN, 2009), onde o coeficiente de seleção final é igual à adição dos coeficientes de seleção das mutações individuais mais um ruído gaussiano de variância σ^2 . Primeiramente verificaremos exemplos e propriedades das redes estudadas, e depois monitoraremos como o destino evolutivo muda com a mutabilidade acompanhando a frequência de genótipos-alvo com n mutações, o número médio de mutações nos genótipos-alvo $\langle n \rangle$, e as entropias total e média, à semelhança do que vimos no capítulo 4.

5.2 Métodos

Os métodos são bem semelhantes ao estudo do efeito da mutabilidade sobre a ordem das mutações. Os sítios de mutação são numerados de 1 até $G = 5$, e os estados dos sítios são representados por 0 (estado selvagem) e 1 (com mutação), em ordem crescente de coeficiente de seleção da direita para a esquerda.

Portanto, o genótipo que tem representação 00000 e é o estado selvagem; o genótipo representado por 00001 corresponde ao mutante de menor aptidão, sem contar a epistasia; o genótipo representado por 10000 se refere ao mutante de uma mutação de maior aptidão, sem contar a epistasia; e o genótipo com representação 11111 corresponde ao genoma com todas as mutações disponíveis realizadas.

A aptidão do genótipo é escolhida determinando-se qual seria o coeficiente de seleção final com aptidão aditiva, onde o coeficiente de seleção isolado de cada mutação é is , onde $i = 1, 2, \dots, 5$ é a posição do sítio da direita para a esquerda, e $s = 0,01$ é o coeficiente de seleção de referência. Após este procedimento, soma-se ao coeficiente de seleção final um ruído gaussiano com média zero e variância $\sigma^2 = 0,005$ apenas para os genótipos mutantes. Na simulação apenas caminhos de um genótipo de aptidão menor para outro de aptidão maior são permitidos. Com isso, esse ruído torna possível que certos genótipos sem todas as mutações presentes, vindo de uma certa sequência de mutações, não tenham um genótipo disponível que o possa suceder. Eles se tornam assim genótipos-alvo alternativos, ou folhas da rede.

Como estamos interessados em que para o sítio $i = 1$ a taxa seja r , e que as taxas de mutações de sítios com diferentes coeficientes de seleção sigam uma progressão geométrica como nos capítulos passados, calculamos a taxa de um sítio da seguinte forma:

$$r_i = r e^{-(i-1)}, \quad (5.1)$$

Como nos capítulos anteriores, o parâmetro r pode ser identificado como a mutabilidade. Com as informações acima temos as propriedades de topologia das redes definidas. A Figura 39 mostra uma rede completa com $G = 5$ que irá servir de referência para a investigação estrutural das redes epistáticas aleatórias.

Esta Figura representa a rede sem epistasia aleatória. Portanto todas as mutações possíveis a partir de cada genótipo aumentam a aptidão. Cada nó dela está associado a um genótipo, mas todos os genótipos com pelo menos duas mutações estão associados a vários nós. Em particular, todos os nós da camada mais exterior representam o mesmo genótipo, aquele em que todas as mutações ocorreram. Cada caminho sobre o grafo, mesmo que não alcance um último genótipo, representa uma classe. No caso com epistasia aleatória, nem todas as classes com máxima aptidão são possíveis, pois em vários casos, ao se chegar a um determinado nó nenhuma das mutações disponíveis irá aumentar a aptidão. Uma folha é um nó final de uma classe.

A amostragem é feita por classes (portanto distingue genótipos iguais com trajetórias diferentes) através do processo de Wright-Fisher semelhante ao usado no caminho evolutivo simples, onde a probabilidade de uma classe ser sorteada para a próxima geração é dada por

$$p_c^{t+1} \propto w_s N_s^t r_{cs} + w_c N_c^t (1 - \sum_d r_{dc}), \quad (5.2)$$

onde p_c é a probabilidade de sorteio da classe c , w_c e w_s são as aptidões da classe atual e da qual ela é originada, respectivamente, N_c^t e N_s^t são o número de indivíduos da classe atual e da qual ela é originada, respectivamente, no tempo t , r_{cs} é a taxa de mutação do genótipo da classe original para o genótipo da classe atual, r_{dc} é a taxa de mutação do

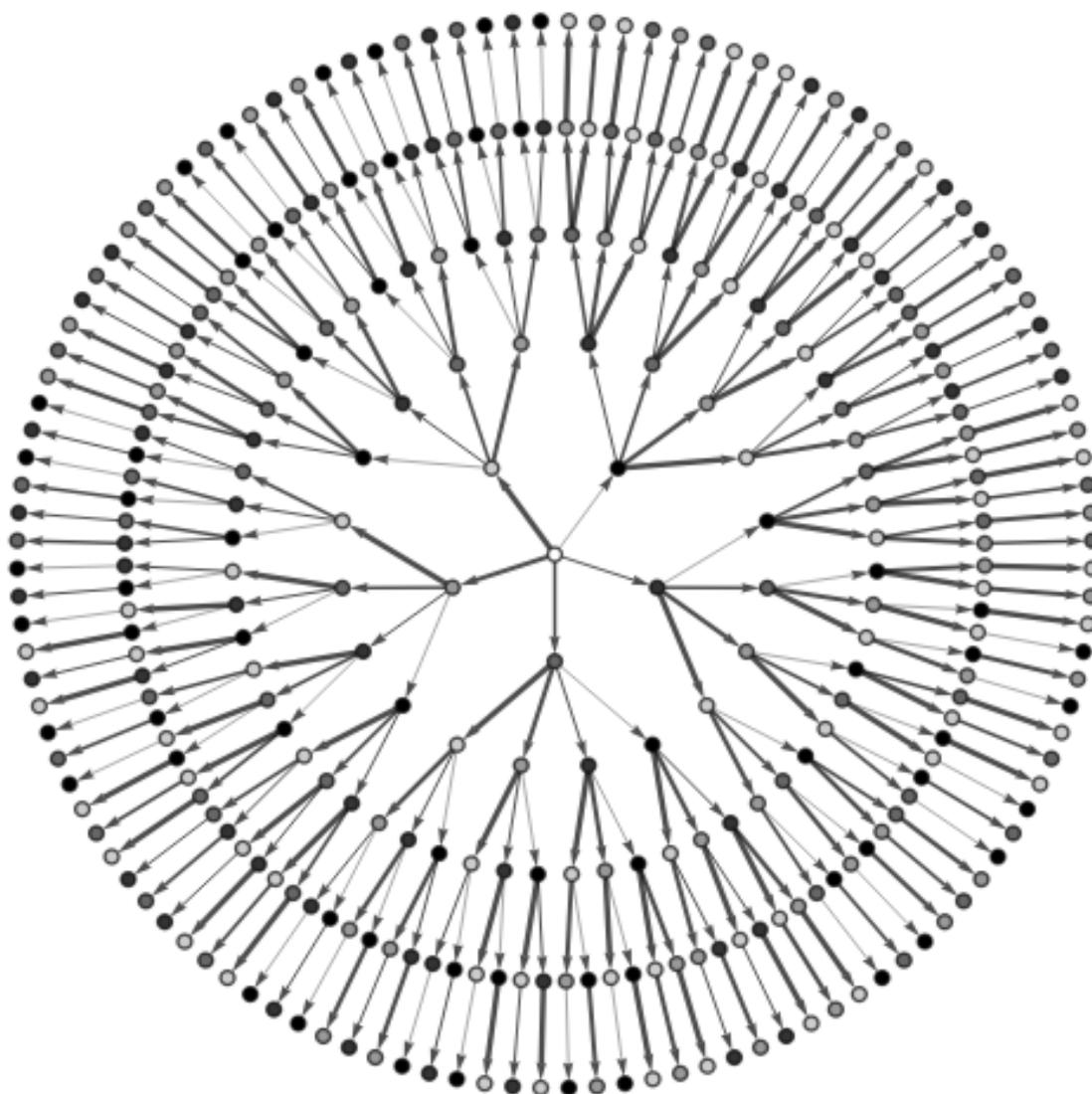


Figura 39 – Rede completa a partir da qual serão construídas as redes com epistasia aleatória. As setas correspondem a transições entre os nós. Quanto mais escuras e mais largas as setas, maior é a taxa de transição entre os respectivos nós. Quanto mais escuros os nós, maior é a aptidão. Note uma relação inversa entre a tonalidade da seta que chega ao nó e a do próprio nó.

genótipo da classe atual para um genótipo d dentre os possíveis destinos imediatos, e \sum_d é a soma sobre todos os destinos possíveis. Um destino é possível se sua aptidão é maior que a do genótipo que lhe dá origem, e está separado do genótipo de origem por apenas uma mutação. Uma classe (nó) só pode ser ancestral de outra se sua aptidão for menor do que uma classe seguinte. Dada a epistasia aleatória, um nó que seria ancestral de outros nós numa rede completa pode se tornar uma folha, isto é, um possível destino evolutivo.

Como no método para o caminho adaptativo simples e o da rede completa, esse esquema equivale a uma fase de seleção logo após o surgimento do indivíduo, e outra fase de reprodução, onde as mutações acontecem.

De cada simulação registramos em formato binário o grafo (em pares nó de origem, nó de destino), os nós que pertencem ao grafo em uma lista, as aptidões dos genótipos de cada simulação, a lista de folhas de cada grafo, os genótipos-folha dos grafos, e o número de pares do grafo, de nós, de folhas, e de folhas do genótipo de destino.

Como nos experimentos de ordem de mutação, as mesmas duas medidas de entropia S e S_T são realizadas. Para ambas, rodamos cada simulação até que toda a população atinja apenas um genótipo final, que corresponde a uma única aptidão final, mas que pode corresponder a mais de uma folha na rede. Em seguida, transformamos o vetor com o número de indivíduos por folha do genótipo escolhido em um vetor de probabilidades ao dividi-lo por N . Na primeira entropia S obtemos inicialmente por simulação a soma de $-p_i \log(p_i)$ de todas as folhas pertinentes, onde p_i é a razão do número de indivíduos de uma das classes dividido por N . Na sequência, calculamos a média de S entre as simulações. O desvio padrão é calculado por *binning*, como indicado a seguir.

A segunda entropia S_T é calculada primeiramente calculando-se a média das distribuições e em seguida somando-se $-\langle p_i \rangle \log \langle p_i \rangle$, onde $\langle p_i \rangle$ é a média da razão do número de indivíduos de uma das classes dividido por N sobre as simulações. O desvio padrão é calculado também por *binning*, descrito a seguir.

Os erros dos histogramas foram realizados através de *bootstrap*, sorteando-se 128 amostras a partir dos dados originais e com mesmo tamanho de amostra destes dados, e partir daí estimamos as quantidades de cada amostra e o desvio entre elas. Para dados da estrutura da rede epistática aleatória, juntamos os dados correspondentes de todas as 51 mutabilidades uniformemente espaçadas de $\log_{10} r = -12$ até $\log_{10} r = -2$. Para cada mutabilidade processamos 16.384 (2^{14}) realizações. O total de realizações para uma dada condição e todas as mutabilidades é de 835.584. A cada uma dessas realizações corresponde a uma única rede com certa associação entre genótipos e aptidões. O erro de cada barra foi tão pequeno que para não prejudicar a visualização não apresentamos as barras.

Para o cálculo dos erros do número médio de mutantes ou das entropias usamos o *binning*. Agrupamos os dados em 2, 4, 8, ou 16 grupos, e calculamos o desvio padrão

das entropias entre os grupos. Quando medimos S , calculamos a média da entropia de cada grupo para calcular o desvio. Quando medimos S_T , agregamos a distribuição de probabilidade dentro dos grupos, calculamos a entropia desta distribuição e então calculamos o desvio. O menor desvio foi escolhido como o erro correspondente a certa mutabilidade. Na prática, os erros foram tão pequenos que em geral não incluímos as suas barras nos gráficos correspondentes.

5.3 Resultados

Primeiramente, vamos explorar as redes epistáticas aleatórias com $s = 0,01$ e $\sigma^2 = 0,005$ apresentando uma amostra delas na Figura 40. Note que para esta configuração a grande maioria dos possíveis nós estão excluídos das redes efetivas. Apenas poucos caminhos estão disponíveis, que em muitos casos mas não todos, alcança a quinta camada correspondente ao genótipo 11111. O fato de não se ter em alguns casos este genótipo de maneira alguma significa que tais redes são iguais a redes $G = 4$ ou menos. Pode ser que as cinco mutações sejam usadas sem que apareçam juntas, o que torna cada *ensemble* de redes G peculiar.

Em uma população infinita, podemos admitir que apenas o genótipo com maior aptidão, distribuído por suas classes presentes na rede epistática, irá sobreviver em tempos muito longos. Mas para populações finitas, uma subpopulação em uma folha mais apta pode ganhar de indivíduos em um nó menos apto, mas que conduziria a população a uma folha de maior aptidão que a alternativa. Além disso, no regime SSWM a população pode facilmente seguir um caminho que não conduza à maior aptidão disponível. Com isso, já intuimos que a mutabilidade deve ter efeito nos destinos evolutivos que são tomados.

Antes de realizar sorteios da epistasia aleatória, temos um grafo orientado (ou rede completa) com nós que correspondem a genótipos, cada genótipo, sendo, em geral, representado por mais de um nó. A rede completa está ilustrada na Figura 39. Cada caminho sobre a rede completa partindo do nó central define uma classe, ou seja, uma história que indica quais mutações ocorreram e em qual ordem, não necessariamente terminando no genótipo com todas as mutações presentes.

Quando realizamos um sorteio da epistasia pode acontecer que um nó tenha aptidão menor que o de seu ancestral imediato na rede completa. Todas as classes da rede completa em que, em alguma aresta, aconteça uma diminuição da aptidão por conta da epistasia são eliminadas. Assim, para cada realização do sorteio da epistasia teremos uma rede, em geral incompleta. Alguns exemplos destas redes são mostrados na Figura 40.

A parte superior da Figura 41 representa, após várias realizações do sorteio da epistasia, uma para cada genótipo, a probabilidade de cada nó da rede completa estar presente nas redes simuladas. Em cada realização, definimos um nó da rede incompleta

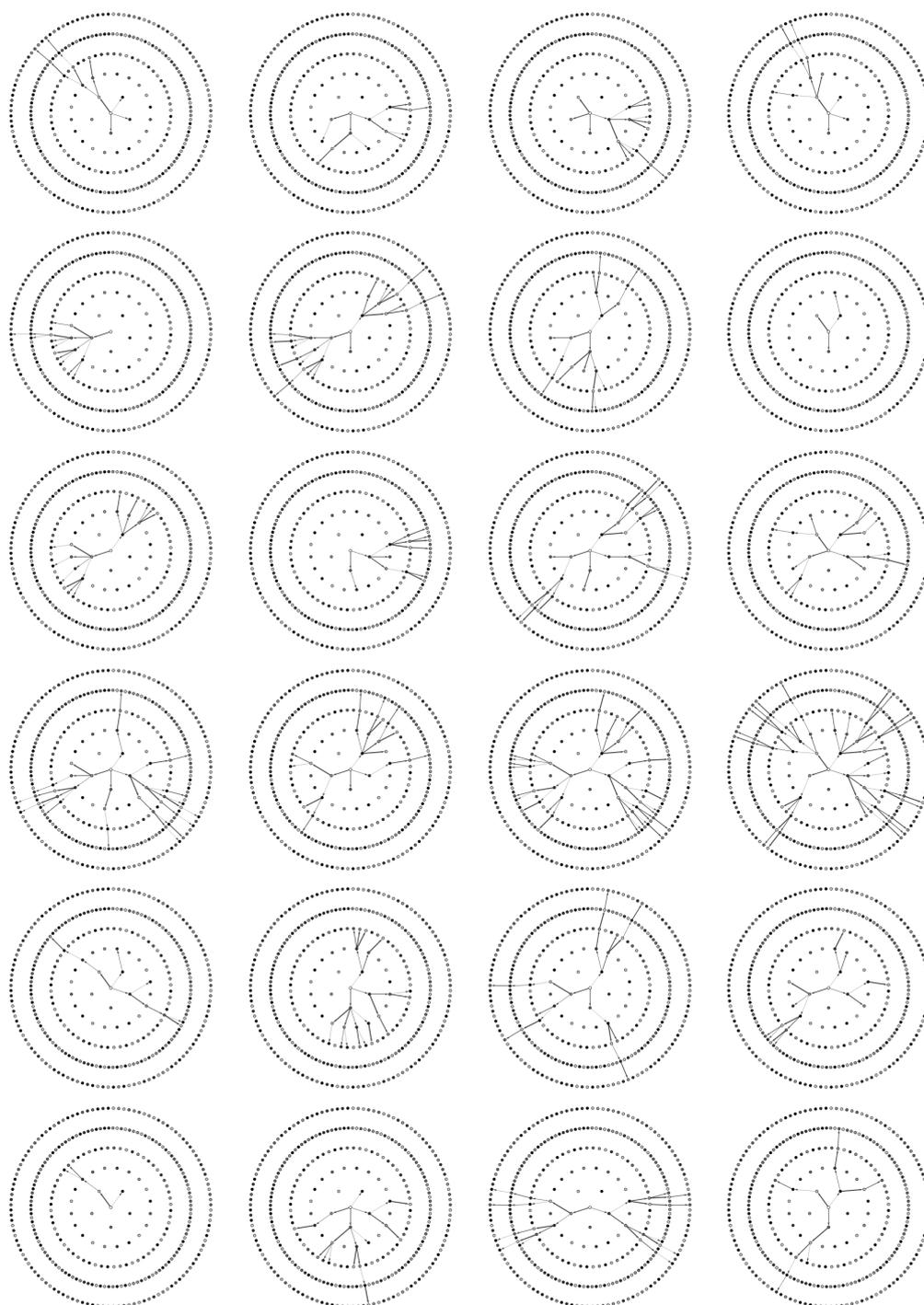


Figura 40 – Exemplos de redes com epistasia aleatória. Todos os nós da rede completa são mostrados como referência. Os nós que pertencem à rede estão ligados por setas a partir do nó raiz. Quanto mais escuras as setas, maior é a taxa de transição entre os respectivos nós. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$.

como sendo uma folha se ele é o último nó da maior classe à qual pertence. Uma folha pode estar associada ao genótipo com todas as mutações, ou não. O segundo caso acontece quando todas as mutações possíveis a partir do genótipo da folha diminuem a aptidão.

A parte inferior da Figura 41 representa a probabilidade de que cada nó da rede completa seja folha nas realizações da epistasia.

Em ambos os casos (nó e folha), a chance de um nó específico estar presente no grafo final como nó ou folha diminui com o número de mutações. Mas o número de nós aumenta com o número de mutações, de certa forma compensando o primeiro efeito.

Para a chance de um nó estar presente como nó na rede final, vemos que as chances de classes com o mesmo número de mutação são semelhantes entre si. As chances de uma mutação com maior aptidão é maior porque a chance do ruído gaussiano tornar sua aptidão menor que a do ancestral, e portanto de estar ausente na rede final, é menor. O nó central naturalmente tem a maior chance de ocorrer porque a raiz da rede é obrigatória. Nenhum outro nó é obrigatório.

Quando olhamos para as chances das folhas acontecerem em certas classes, primeiramente notamos que existe a probabilidade não-nula de que a raiz seja uma folha. Obviamente, se isso ocorrer, toda a rede consiste na raiz, e em um único destino evolutivo. A chance é pequena, mas não desprezível. Em geral, a chance de uma classe ser truncada mais cedo está relacionada a seus ancestrais terem menor aptidão que em outro caminho. Assim, esperamos classes mais longas, com destino provavelmente com maior aptidão, a partir de nós mais aptos.

A Figura 42 mostra a distribuição dos genótipos em todos os grafos. Note que não há uma relação tão clara entre a aptidão aditiva esperada e a frequência, exceto para as maiores aptidões. Estas últimas frequências podem refletir o fato de que existem mais nós com cada um desses genótipos.

A distribuição dos genótipos que são potenciais destinos evolutivos ou folhas é ainda mais irregular que aqueles que simplesmente são nós, como mostrado na Figura 43. A tendência do aumento da frequência de genótipos finais com a aptidão aditiva esperada pode ser explicada como anteriormente: pela maior frequência de nós com cada um desses genótipos.

A Figura 44 mostra a distribuição do número de mutações nas folhas. Esse gráfico traz em particular maior entendimento de gráficos seguintes que relacionam o número de mutações no destino final efetivo com a mutabilidade. É interessante que para $n = 5$, onde esperamos o maior número de nós em um grafo completo (juntamente com $n = 4$) não encontramos a maior frequência de folhas. Isso reflete que a chance de um ramo ser “podado”, não permitindo folhas com maior número de mutações, aumenta com n a ponto deste fenômeno acontecer.

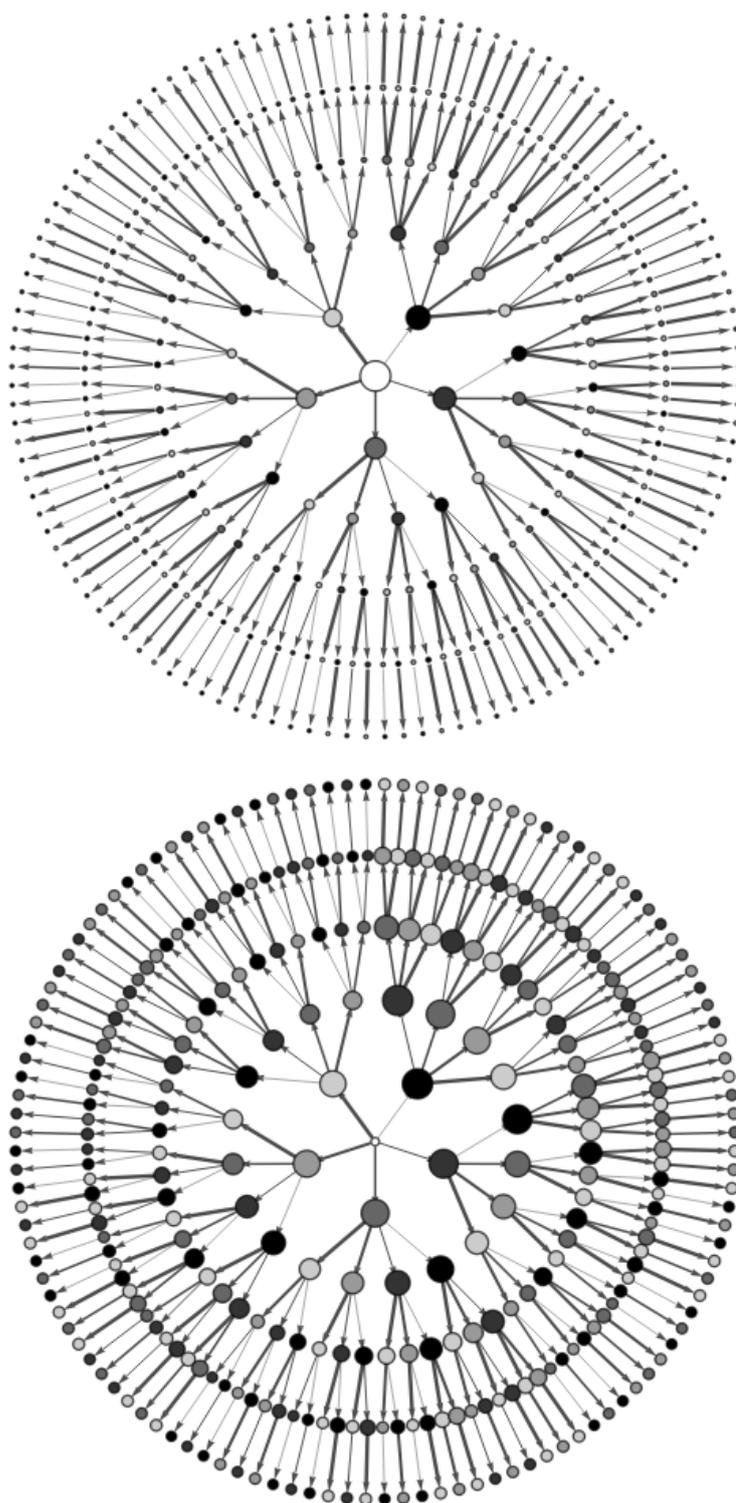


Figura 41 – Representação da frequência de nós (acima) e folhas (abaixo) da rede com epistasia aleatória. Quanto maior o tamanho do nó, maior é a sua frequência relativa. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$.

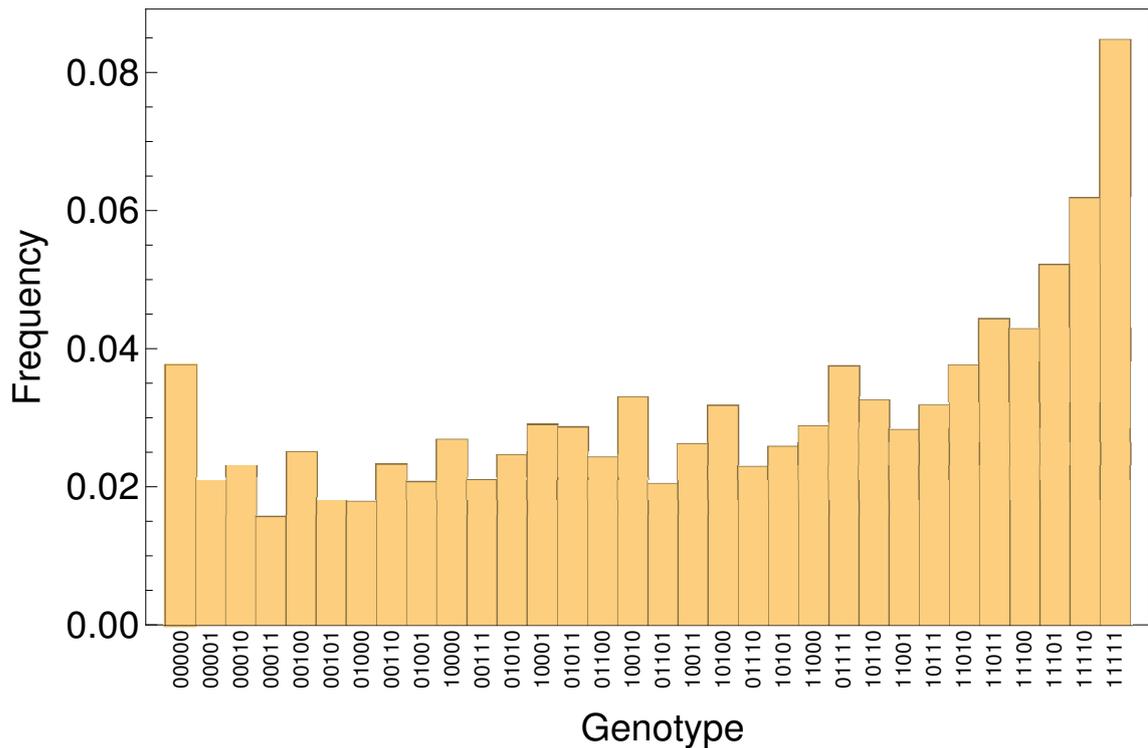


Figura 42 – Histograma da probabilidade de um nó ser de um dado genótipo. Os genótipos são representados por números binários que codificam a representação das mutações em cada sítio (veja Métodos), em ordem crescente de aptidão aditiva. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$.

A Figura 45 é o primeiro resultado a mostrar que a mutabilidade altera o destino evolutivo, mostrando que os padrões com que os genótipos finais são alcançados mudam com a taxa de mutação basal. Note que para $n = 0$ a frequência é basicamente constante, uma vez que se a raiz é folha, ela será sempre o único destino, e sua frequência é semelhante nas amostras de redes de cada mutabilidade escolhida. Para valores altos de r , quanto maior a mutabilidade, mais fácil é para indivíduos ganharem mutações, e acessarem destinos mais distantes, que devem se correlacionar com os de maior aptidão. Isso se reflete em ambos os gráficos.

É interessante que para mutabilidades mais baixas, destinos com duas mutações ($n = 2$) prevaleçam. Isso pode ser um reflexo do número maior destes destinos nas redes que $n = 1$, que por essa prevalência aumentam sua chance de alcançarem tais números de mutações. Ao mesmo tempo, a mutabilidade menor não permite que genótipos com $n = 1$ e aptidões altas sejam ocupados porque a taxa de mutação é baixa para eles, o que dificulta ainda mais que folhas com $n = 1$ prevaleçam. Ainda, o regime de baixa mutabilidade também dificulta que folhas com $n > 2$ sejam alcançadas tão facilmente, porque populações nesses caminhos devem ocupar nós com $n \leq 2$ enquanto competem com folhas com o mesmo número de mutações. Como é mais fácil truncar um caminho quando a aptidão do nó é maior (tornando-o folha), espera-se que folhas com n mutações

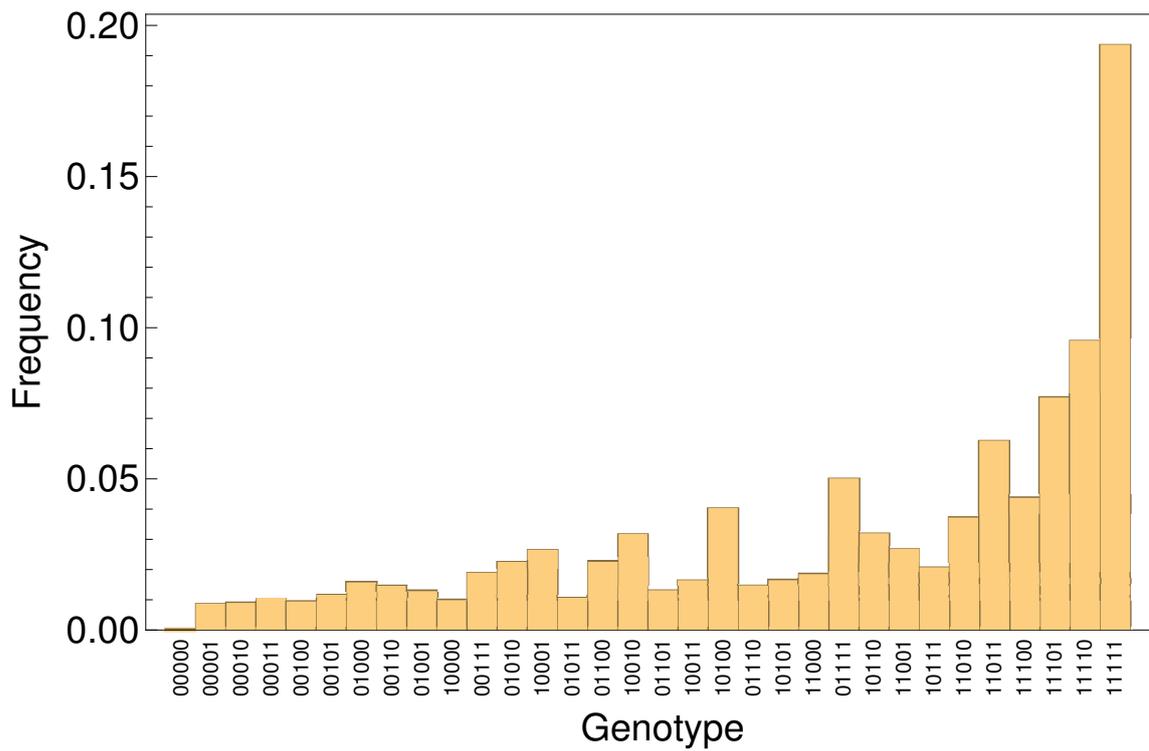


Figura 43 – Histogramas da probabilidade de uma folha ser de um dado genótipo. Os genótipos são representados por números binários que codificam a representação das mutações em cada sítio (veja Métodos), em ordem crescente de aptidão aditiva. . A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$.

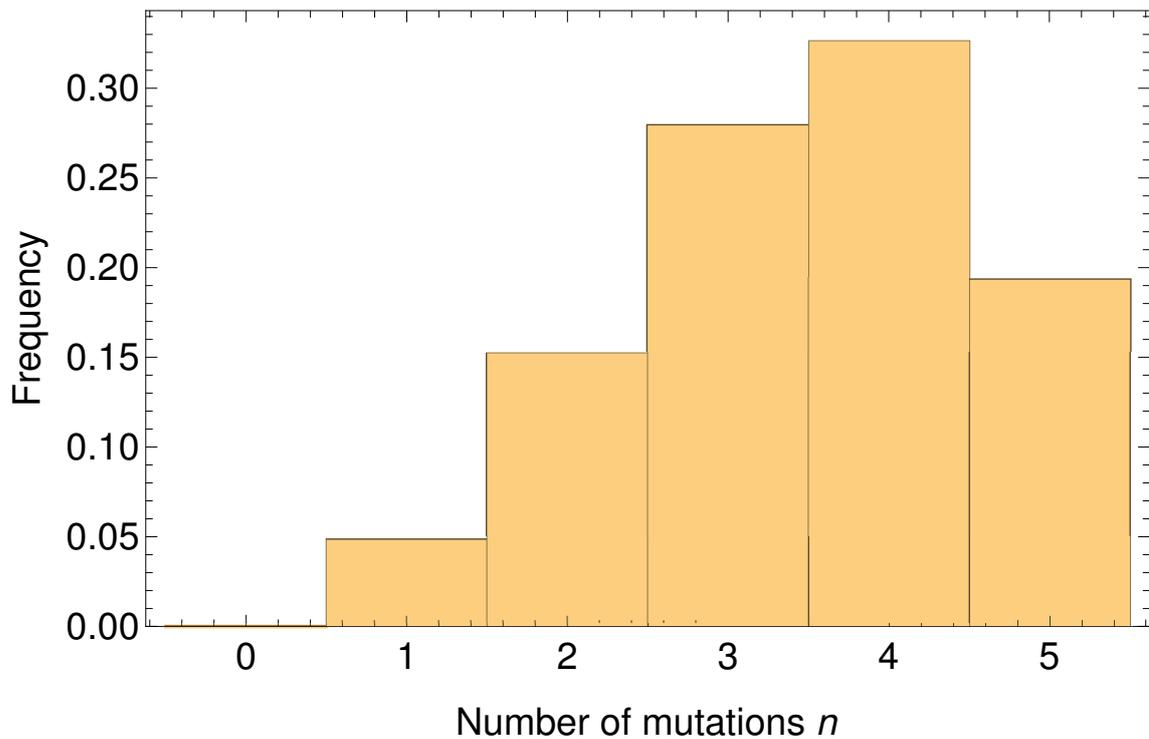


Figura 44 – Histograma da probabilidade de uma folha ter n mutações. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$.

tenham maior aptidão que nós com o mesmo número de mutações. Como resultado, folhas com menor n podem prevalecer.

Mas estranhamente, os genótipos com $n = 1$ mutações prevalecem a uma mutabilidade maior, de cerca de $r = 10^{-6}$. Talvez isso corresponda ao maior acesso ao genótipo $n = 1$ de maior aptidão aditiva, que tem a menor taxa e maior probabilidade de ser folha que outros nós com $n = 1$. Ao aumentá-la, ele suplanta mutantes $n = 2$ com aptidões menores. A prevalência dos tipos $n = 1$ e $n = 2$ justificam o vale do gráfico de $\langle n \rangle$ em uma janela de mutabilidades.

Outro ponto importante é que a maior disponibilidade de mais folhas com $n = 5$ (como se pode ver na Figura 43) não afetou a distribuição de n para uma faixa ampla de mutabilidade. Os indivíduos da população aparentemente não chegavam a estas folhas mais altas devido à competição entre folhas mais baixas e nós intermediários até folhas em n maior, que não eram pouco numerosos (Figura 42).

Outra maneira de observarmos que a mutabilidade está alterando o padrão dos destinos evolutivos é medir a entropia do genótipo final através de todos os seus nós presentes nas redes (Figura 46). A entropia S mede de certa forma o número de destinos alcançados por realização. Quando ela é zero, apenas um destino é alcançado. Para uma ampla faixa de mutabilidade (até aproximadamente 10^{-6}) temos apenas uma folha alcançada por vez. Isso não é reflexo da frequência de chegada a destinos com uma só mutação, onde só corresponde um nó a cada genótipo, porque sua frequência não é muito maior que as demais. Com o aumento da mutabilidade, mais destinos evolutivos podem ser alcançados, e por isso a entropia aumenta monotonicamente.

A entropia total S_T corresponde a uma medida da repetibilidade do processo dado que a rede apresenta epistasia aleatória. É interessante que, apesar de distinta do gráfico com o número médio de mutantes, ambos se parecem bastante. Isso indica que o aumento da repetibilidade está associada ao menor número de nós associados aos mutantes $n = 1$ e $n = 2$, que prevalecem neste regime de mutabilidade.

5.4 Conclusão

Verificamos neste capítulo que os destinos evolutivos, seja a sequência de mutantes até o genótipo-alvo ou a identidade do próprio genótipo final, varia com a mutabilidade. A taxa de mutação, além de acelerar a formação de variantes, também afeta a direção que a evolução irá tomar. Este é um resultado fundamental em Genética de Populações, que tem repercussões nas diferentes áreas que ela afeta.

Pelo número de nós ser numeroso para o tamanho de rede que usamos $G = 5$, agregamos as frequências da população neles em termos de genótipos e número de mutação,

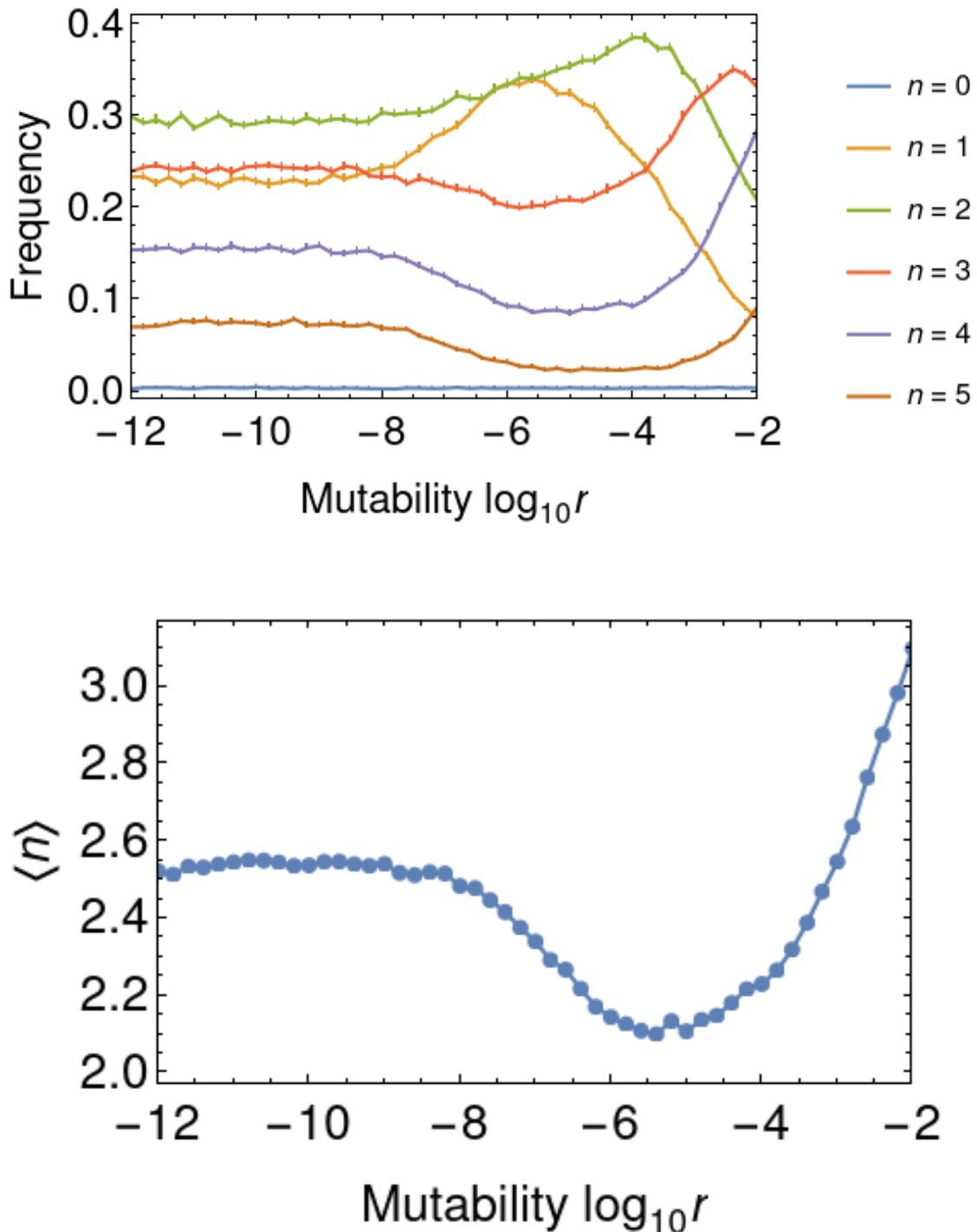


Figura 45 – Frequências dos genótipos de acordo com o número de mutações que carrega (acima), e número médio de mutações por genoma, como função da mutabilidade. O tamanho da população é $N = 10^6$ e o coeficiente de seleção unitário é $s = 0,01$. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$. Para cada mutabilidade foram construídas 16384 novas redes, com uma simulação por rede. A barra de erro é mostrada em ambos os gráficos, mas não é grande o suficiente no segundo gráfico.

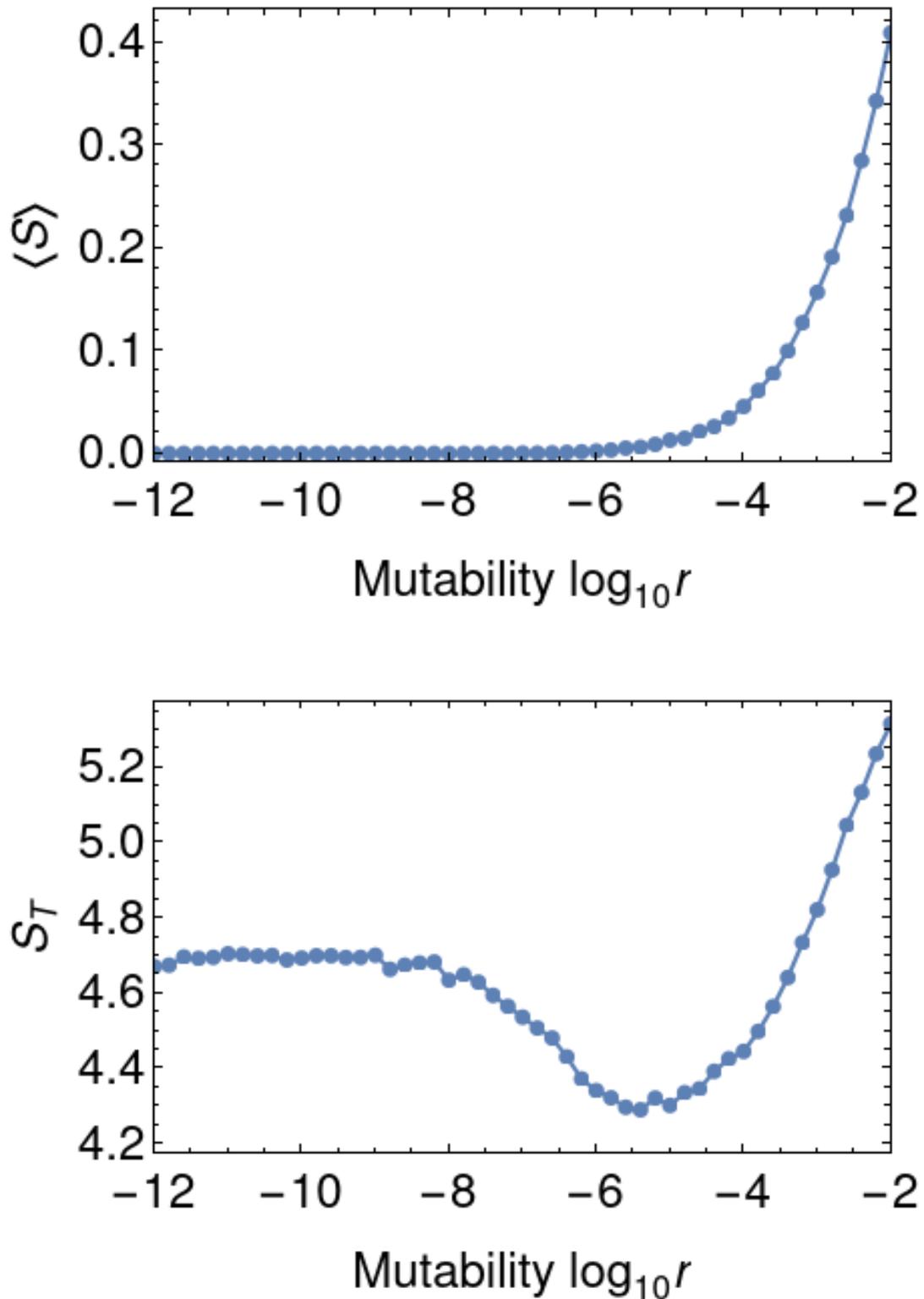


Figura 46 – Média das entropias de cada realização, e entropia da média das distribuições de cada realização, como função da mutabilidade. O tamanho da população é $N = 10^6$ e o coeficiente de seleção unitário é $s = 0,01$. A variância do ruído gaussiano sobre as aptidões é $\sigma^2 = 0,005$.

o que deixa evidente tal efeito da mutabilidade.

As medidas de entropia revelaram que apenas uma folha é alcançada por realização a mutabilidades suficientemente baixas, mas que a repetibilidade do processo adaptativo em redes epistáticas, que devem representar qualitativamente redes experimentais e incompletas, só é máxima (a entropia é mínima) para regimes de mutabilidade intermediária.

Muitos outros experimentos foram realizados, mas não puderam ser analisados a tempo, como o efeito de σ^2 , de s (ambos afetam a topologia da rede, e o último afeta a pressão seletiva), do tamanho da população N , o efeito do modelo generalizado de Gumbel para a distribuição de efeitos de aptidão, do tamanho das redes e de vários controles. Também deixamos de apresentar resultados explícitos da distribuição da população sobre os genótipos a diferentes mutabilidades.

Esperamos que em um artigo dedicado ao assunto possamos cobrir todos estes resultados e trazer um panorama ainda mais completo de como a mutabilidade afeta o padrão de determinação dos destinos evolutivos.

6 Conclusão

Após encaminharmos a solução analítica para o modelo adaptativo com aptidão aditiva, estudamos como a mutabilidade afeta a história adaptativa. Uma das inovações que fizemos foi a de identificar um nó numa rede mutacional com um genótipo que percorreu uma sequência específica de mutações. Também medimos a repetibilidade por realização nesta rede, e também do processo como um todo, reunindo os resultados de cada realização. Realizamos também vários controles para o estudo da ordem das mutações.

Nossa abordagem inicial começou com a investigação do processo adaptativo para populações infinitas. Existem diferenças esperadas para populações finitas, que são mais realistas, mas aquelas nos trouxeram intuição para o processo. Além disso, a fórmula para o tempo de adaptação nos sugeriu que a ordem dos tempos de aptidão para diferentes alelos poderia se alterar com a mutabilidade. Infelizmente nossa dedução da fórmula da frequência dos diferentes números de mutação não foi totalmente transparente, porque um pequeno passo ainda não foi possível de ser provado à mão, em contraste com a resposta afirmativa do CAS.

Há basicamente duas maneiras de se alterar uma história adaptativa: mudando a ordem com que as mutações ocorrem, e mudando aonde a sequência evolutiva é interrompida. Neste ponto de trajetórias diferentes, os genótipos podem ser diferentes ou não, o que torna-se um objeto de estudo adicional.

Em um sistema de um passo e duas alternativas, objeto de estudo do Capítulo 3, o número de mutações até o estado final é a mesma, mas o genótipo muda, juntamente com a taxa de mutação e a aptidão, mantidas em oposição. No sistema estudado, a mutabilidade pode definir qual destino prevalece. Determinamos os regimes nos parâmetros de interesse onde o destino menos apto prevalece.

Em redes completas, estudadas no Capítulo 4, o número de passos é o mesmo até os possíveis destinos, o genótipo final é o mesmo, mas a ordem com que as mutações ocorrem é distinta. Em diferentes regimes de mutabilidade, diferentes ordens são privilegiadas. A repetibilidade aumenta para valores intermediários de mutabilidade. Isso se deve porque o regime em vigor é o da competição e seleção entre as trajetórias. Dependendo da taxa de mutação para mutação mais apta, a trajetória que prevalece é a que a aptidão cresce mais rapidamente. Esse processo elimina mais sub-populações, e diminui a entropia. O processo se torna mais previsível.

Para redes incompletas, com possíveis destinos com diferentes genótipos e números de mutação, e a mutabilidades menores, a dinâmica privilegia tipos mais rapidamente acessíveis. O fato de $\langle n \rangle$ ficar aproximadamente constante para mutabilidades intermediárias

rias indica a ocorrência de algo mais complexo, ainda a ser entendido. Com mutabilidades maiores se acessam genótipos com mais mutações, que são mais prováveis de terem maior aptidão e prevalecerem quando em competição com outros destinos. O fato mais importante aqui é destacar que a mutabilidade, ao invés de apenas apressar as mutações, ela é determinante para onde a evolução seguirá.

Schenk *et al.* (SCHENK *et al.*, 2022) estudaram o efeito de tamanhos N médios e grandes de bactérias *E. coli* sobre a adaptação, examinando com cuidado a história não só de substituições de nucleotídeos, mas variantes estruturais (deleções e inserções) e outros tipos de mutação, realizando o sequenciamento de vários clones ao longo do tempo. Eles notaram diferenças adaptativas entre estas populações, em particular a vantagem de populações maiores devido à interferência clonal. Do nosso ponto de vista, as explicações concordam, porque quanto maior a população, maior o suprimento mutacional $\theta = Nr$, e mais fácil é ela estar no regime de seleção, e de adaptação mais rápida. Esperamos, e precisamos confirmar com dados não-analisados ainda, que populações grandes também alcançam aptidões finais maiores.

O trabalho de Weireich *et al.*, denominado “Evolução darwiniana pode seguir apenas poucos caminhos mutacionais” (WEINREICH *et al.*, 2006) (tradução minha) na verdade não considerou o efeito da competição entre os diferentes caminhos, que restringe ainda mais os caminhos que são de fato trilhados e aumenta significativamente a repetibilidade da evolução. Esse efeito é dependente da mutabilidade, que em última instância depende dos alelos mutadores que a população carrega.

Esperamos com isso que diferentes linhagens mutadoras sigam diferentes caminhos evolutivos, e que somados à taxa de mutação adaptativa, taxa deletéria, deriva e seleção direta sobre o alelo mutador determinem a linhagem que prevalecerá em uma população com linhagens em competição. Esperamos que a mutabilidade, via a determinação da ordem e do destino evolutivo, determine a evolução dela mesma. Este é um aspecto inédito, até onde sabemos.

Um outro ponto importante é que nossos resultados sugerem que diferentes sistemas de dominância podem seguir histórias adaptativas bem distintas, dados que os regimes em que as mutabilidades associadas a sistemas dominantes e recessivos podem ser bem distintos, como o SSWM e o de seleção. Além disso, a prevalência de um sistema ou outro quando eles estão em competição devem favorecer, do ponto de vista da história de adaptações, o sistema dominante, cuja mutabilidade é muito maior.

Outros aspectos a serem explorados são a investigação do efeito da mutabilidade sobre trajetórias em paisagens adaptativas experimentais, do efeito do espectro mutacional, isto é, das frequências relativas dos diferentes tipos de mutação possíveis no material genético, de sistemas diplóides e da presença de recombinação. Isso expandirá nossa compreensão para situações mais realistas e amplas.

Referências

- AITA, T.; HUSIMI, Y. Fitness spectrum among random mutants on mt. fuji fitness landscape. *J. Theor. Biol.*, v. 182, n. 4, p. 469–485, May 1996. Citado na página 37.
- BAER, C. F.; MIYAMOTO, M. M.; DENVER, D. R. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, v. 8, n. 8, p. 619–631, AUG 2007. ISSN 1471-0056. Citado 2 vezes nas páginas 28 e 29.
- BARTON, N. H.; COE, J. B. On the application of statistical physics to evolutionary biology. *J. Theor. Biol.*, v. 259, n. 2, p. 317–324, JUL 21 2009. ISSN 0022-5193. Citado na página 27.
- BAUER, B.; GOKHALE, C. S. Repeatability of evolution on epistatic landscapes. *Sci. Rep.*, v. 5, 5 2015. ISSN 2045-2322. Citado na página 73.
- BJEDOV, I. et al. Stress-induced mutagenesis in bacteria. *Science*, v. 300, n. 5624, p. 1404–1409, MAY 30 2003. ISSN 0036-8075. Citado na página 30.
- BJORKHOLM, B. et al. Mutation frequency and biological cost of antibiotic resistance in helicobacter pylori. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, v. 98, n. 25, p. 14607–14612, DEC 4 2001. ISSN 0027-8424. Citado na página 29.
- BOLTON, K. L. et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *NATURE GENETICS*, v. 52, n. 11, p. 1219+, NOV 2020. ISSN 1061-4036. Citado na página 28.
- BOURGUET, D. The evolution of dominance. *HEREDITY*, v. 83, n. 1, p. 1–4, JUL 1999. ISSN 0018-067X. Citado na página 35.
- BULL, J. et al. Exceptional convergent evolution in a virus. *Genetics*, v. 147, n. 4, p. 1497–1507, DEC 1997. ISSN 0016-6731. Citado na página 73.
- BURKE, M. K.; LITI, G.; LONG, A. D. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of saccharomyces cerevisiae. *Mol. Biol. Evol.*, v. 31, n. 12, p. 3228–3239, DEC 2014. ISSN 0737-4038. Citado na página 73.
- CAMPO, R. del et al. Population structure, antimicrobial resistance, and mutation frequencies of streptococcus pneumoniae isolates from cystic fibrosis patients. *JOURNAL OF CLINICAL MICROBIOLOGY*, v. 43, n. 5, p. 2207–2214, MAY 2005. ISSN 0095-1137. Citado na página 29.
- CHAO, L.; COX, E. Competition between high and low mutating strains of escherichia-coli. *Evolution*, v. 37, n. 1, p. 125–134, 1983. ISSN 0014-3820. Citado na página 29.
- CHAO, L. et al. Transposable elements as mutator genes in evolution. *Nature*, v. 303, n. 5918, p. 633–635, 1983. ISSN 0028-0836. Citado na página 29.
- CHARLESWORTH, B. How long does it take to fix a favorable mutation, and why should we care? *Am. Nat.*, v. 195, n. 5, p. 753–771, MAY 2020. Citado na página 67.

- CHEN, J. Z.; FOWLER, D. M.; TOKURIKI, N. Environmental selection and epistasis in an empirical phenotype-environment-fitness landscape. *NATURE ECOLOGY & EVOLUTION*, v. 6, n. 4, p. 427+, APR 2022. ISSN 2397-334X. Citado na página 28.
- CHIOTTI, K. E. et al. The valley-of-death: Reciprocal sign epistasis constrains adaptive trajectories in a constant, nutrient limiting environment. *Genomics*, v. 104, n. 6, A, p. 431–437, DEC 2014. ISSN 0888-7543. Citado na página 87.
- COSSIO-DIAZ, J. Fernandez-de; UGUZZONI, G.; PAGNANI, A. Unsupervised inference of protein fitness landscape from deep mutational scan. *MOLECULAR BIOLOGY AND EVOLUTION*, v. 38, n. 1, p. 318–328, JAN 2021. ISSN 0737-4038. Citado na página 28.
- COX, E.; GIBSON, T. Selection for high mutation rates in chemostats. *Genetics*, v. 77, n. 2, p. 169–184, 1974. ISSN 0016-6731. Citado na página 29.
- DAWID, A. et al. Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape. *Chaos*, v. 20, n. 2, JUN 2010. ISSN 1054-1500. Citado na página 87.
- DENAMUR, E. et al. High frequency of mutator strains among human uropathogenic escherichia coli isolates. *JOURNAL OF BACTERIOLOGY*, v. 184, n. 2, p. 605–609, JAN 2002. ISSN 0021-9193. Citado na página 29.
- DENAMUR, E.; MATIC, I. Evolution of mutation rates in bacteria. *Molecular Microbiology*, v. 60, n. 4, p. 820–827, MAY 2006. ISSN 0950-382X. Citado 2 vezes nas páginas 28 e 29.
- DIJK, T. van et al. Mutation supply and the repeatability of selection for antibiotic resistance. *Phys. Biol.*, v. 14, 03 2017. Citado na página 73.
- DONNELLY, P.; KURTZ, T. Genealogical processes for fleming-viot models with selection and recombination. *Annals of Applied Probability*, v. 9, n. 4, p. 1091–1148, 1999. Citado na página 36.
- DROSSEL, B. Biological evolution and statistical physics. 2001. Citado 2 vezes nas páginas 26 e 27.
- EVANS, S. N.; STEINSALTZ, D.; WACHTER, K. W. A mutation-selection model with recombination for general genotypes. *Memoirs of the American Mathematical Society*, v. 222, n. 1044, p. 1+, 2013. Citado na página 36.
- EWENS, W. *Mathematical Population Genetics 1: Theoretical Introduction*. [S.l.]: Springer New York, 2004. (Interdisciplinary Applied Mathematics). ISBN 9780387201917. Citado 4 vezes nas páginas 33, 36, 64 e 123.
- FOGLE, C. A.; NAGLE, J. L.; DESAI, M. M. Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics*, v. 180, n. 4, p. 2163–2173, DEC 2008. ISSN 0016-6731. Citado na página 32.
- GABZI, T.; PILPEL, Y.; FRIEDLANDER, T. Fitness landscape analysis of a trna gene reveals that the wild type allele is sub-optimal, yet mutationally robust. *MOLECULAR BIOLOGY AND EVOLUTION*, v. 39, n. 9, SEP 1 2022. ISSN 0737-4038. Citado na página 28.

- GENTILE, C. F. et al. Competition between high- and higher-mutating strains of escherichia coli. *Biology Letters*, v. 7, n. 3, p. 422–424, JUN 23 2011. ISSN 1744-9561. Citado na página 33.
- GERRISH, P.; LENSKI, R. The fate of competing beneficial mutations in an asexual population. *Genetica*, v. 102-3, n. SI, p. 127–144, 1998. ISSN 0016-6707. Citado na página 29.
- GILLESPIE, J. Genetic drift in an infinite population: The pseudohitchhiking model. *Genetics*, v. 155, n. 2, p. 909–919, 2000. Citado na página 36.
- GILLESPIE, J. *Population Genetics: A Concise Guide*. [S.l.]: Johns Hopkins University Press, 2004. (Population Genetics). ISBN 9781421401706. Citado 2 vezes nas páginas 64 e 121.
- GILLESPIE, J. H. Some properties of finite populations experiencing strong selection and weak mutation. *Am. Nat.*, v. 121, n. 5, p. 691, 1983. Citado na página 70.
- GIRAUD, A. et al. The rise and fall of mutator bacteria. *Current Opinion in Microbiology*, v. 4, n. 5, p. 582–585, OCT 2001. ISSN 1369-5274. Citado na página 29.
- GOULART, C. P. et al. Designing antibiotic cycling strategies by determining and understanding local adaptive landscapes. *PLOS ONE*, v. 8, n. 2, FEB 13 2013. ISSN 1932-6203. Citado na página 74.
- GRECH, L. et al. Fitness landscape of the fission yeast genome. *MOLECULAR BIOLOGY AND EVOLUTION*, v. 36, n. 8, p. 1612–1623, AUG 2019. ISSN 0737-4038. Citado na página 28.
- GROSS, M.; SIEGEL, E. Incidence of mutator strains in escherichia-coli and coliforms in nature. *Mutation Research*, v. 91, n. 2, p. 107–110, 1981. ISSN 0921-8262. Citado na página 29.
- HALDANE, J. The part played by recurrent mutation in evolution. *Am. Nat.*, v. 67, n. 708, p. 5–19, Jan 1933. Citado na página 73.
- HEGRENESS, M. et al. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science*, American Association for the Advancement of Science, Washington, DC, v. 311, n. 5767, p. 1615–1617, 2006. Citado na página 45.
- HERR, A. J.; WILLIAMS, L. N.; PRESTON, B. D. Antimutator variants of dna polymerases. *Critical Reviews in Biochemistry and Molecular Biology*, v. 46, n. 6, p. 548–570, DEC 2011. ISSN 1040-9238. Citado na página 33.
- HIGGS, P. Frequency distributions in population genetics parallel those in statistical physics. *Phys. Rev. E*, v. 51, n. 1, p. 95–101, JAN 1995. ISSN 2470-0045. Citado na página 27.
- JOHN, S.; JAIN, K. Effect of drift, selection and recombination on the equilibrium frequency of deleterious mutations. *J. Theor. Biol.*, v. 365, p. 238–246, 2015. Citado 2 vezes nas páginas 37 e 58.

- JOHNSON, T. The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, v. 266, n. 1436, p. 2389–2397, DEC 7 1999. ISSN 0962-8452. Citado na página 29.
- JOHNSON, T. Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *GENETICS*, v. 151, n. 4, p. 1621–1631, APR 1999. ISSN 0016-6731. Citado na página 29.
- KAMPEN, N. V. *Stochastic Processes in Physics and Chemistry*. [S.l.]: Elsevier Science, 2011. (North-Holland Personal Library). ISBN 9780080475363. Citado na página 66.
- KIMURA, M. On the probability of fixation of mutant genes in a population. *Genetics*, Oxford University Press, v. 47, n. 6, p. 713, 1962. Citado na página 126.
- KNOPP, M.; ANDERSSON, D. Predictable phenotypes of antibiotic resistance mutations. *mBio*, v. 9, p. e00770–18, 05 2018. Citado na página 74.
- KVITEK, D. J.; SHERLOCK, G. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLOS Genetics*, v. 7, n. 4, APR 2011. ISSN 1553-7404. Citado na página 87.
- LABAT, F. et al. Mutator phenotype confers advantage in escherichia coli chronic urinary tract infection pathogenesis. *FEMS IMMUNOLOGY AND MEDICAL MICROBIOLOGY*, v. 44, n. 3, p. 317–321, JUN 1 2005. ISSN 0928-8244. Citado na página 29.
- LACHAPELLE, J.; REID, J.; COLEGRAVE, N. Repeatability of adaptation in experimental populations of different sizes. *Proc. R. Soc. B*, v. 282, 04 2015. Citado na página 73.
- LALANNE, J.-B.; PARKER, D. J.; LI, G.-W. Spurious regulatory connections dictate the expression-fitness landscape of translation factors. *MOLECULAR SYSTEMS BIOLOGY*, v. 17, n. 4, APR 2021. ISSN 1744-4292. Citado na página 28.
- LALIC, J.; ELENA, S. F. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant rna virus. *Heredity*, v. 109, n. 2, p. 71–77, AUG 2012. ISSN 0018-067X. Citado na página 87.
- LANG, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, v. 500, n. 7464, p. 571+, AUG 29 2013. ISSN 0028-0836. Citado na página 32.
- LECLERC, J. et al. High mutation frequencies among escherichia coli and salmonella pathogens. *Science*, v. 274, n. 5290, p. 1208–1211, NOV 15 1996. ISSN 0036-8075. Citado 2 vezes nas páginas 29 e 30.
- LECLERC, J. et al. Detection of mutator subpopulations in salmonella typhimurium lt2 by reversion of his alleles. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, v. 400, n. 1-2, SI, p. 89–97, MAY 25 1998. ISSN 0027-5107. Citado na página 29.
- LENORMAND, T.; ROZE, D.; ROUSSET, F. Stochasticity in evolution. *Trends Ecol. Evol.*, v. 24, n. 3, p. 157–165, MAR 2009. ISSN 0169-5347. Citado na página 35.

- LUNZER, M. et al. The biochemical architecture of an ancient adaptive landscape. *Science*, v. 310, n. 5747, p. 499–501, OCT 21 2005. ISSN 0036-8075. Citado na página 74.
- MACLEAN, R. C.; BUCKLING, A. The distribution of fitness effects of beneficial mutations in *Pseudomonas aeruginosa*. *PLOS Genetics*, v. 5, n. 3, MAR 2009. ISSN 1553-7390. Citado na página 34.
- MAO, E. et al. Proliferation of mutators in a cell population. *Journal of Bacteriology*, v. 179, n. 2, p. 417–422, JAN 1997. ISSN 0021-9193. Citado na página 29.
- MARTIN, C. H.; GOULD, K. J. Surprising spatiotemporal stability of a multi-peak fitness landscape revealed by independent field experiments measuring hybrid fitness. *EVOLUTION LETTERS*, v. 4, n. 6, p. 530–544, DEC 2020. Citado na página 28.
- MATIC, I. et al. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *SCIENCE*, v. 277, n. 5333, p. 1833–1834, SEP 19 1997. ISSN 0036-8075. Citado na página 29.
- MATSUDA, H.; ISHII, K. A synthetic theory of molecular evolution. *Genes & Genetic Systems*, v. 76, n. 3, p. 149–158, 2001. Citado na página 36.
- MAYNARD-SMITH, J. et al. Developmental constraints and evolution: A perspective from the mountain lake conference on development and evolution. *Q. Rev. Biol.*, v. 60, n. 3, p. 265–287, Sep 1985. Citado na página 73.
- MCCANDLISH, D. M.; STOLTZFUS, A. Modeling evolution using the probability of fixation: History and implications. *Quarterly Review of Biology*, v. 89, n. 3, p. 225–252, SEP 2014. ISSN 0033-5770. Citado na página 32.
- METZGAR, D.; WILLS, C. Evidence for the adaptive evolution of mutation rates. *Cell*, v. 101, n. 6, p. 581–584, JUN 9 2000. ISSN 0092-8674. Citado na página 30.
- MEYER, J. et al. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science (New York, N.Y.)*, v. 335, p. 428–32, 01 2012. Citado na página 73.
- MIRA, P. M. et al. Adaptive landscapes of resistance genes change as antibiotic concentrations change. *Mol. Biol. Evol.*, v. 32, n. 10, p. 2707–2715, OCT 2015. ISSN 0737-4038. Citado na página 74.
- NEHER, R. A.; SHRAIMAN, B. I. Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proc. Natl. Acad. Sci. U.S.A.*, National Acad Sciences, v. 106, n. 16, p. 6866–6871, 2009. Citado na página 95.
- NGHE, P.; KOGENARU, M.; TANS, S. J. Sign epistasis caused by hierarchy within signalling cascades. *Nature Communications*, v. 9, APR 13 2018. ISSN 2041-1723. Citado na página 87.
- NICHOL, D. et al. Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nat. Commun.*, v. 10, 12 2019. Citado na página 73.
- NOTLEY-MCROBB, L.; SEETO, S.; FERENCI, T. Enrichment and elimination of muty mutators in *Escherichia coli* populations. *Genetics*, v. 162, n. 3, p. 1055–1062, NOV 2002. ISSN 0016-6731. Citado na página 29.

- OLIVER, A. et al. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *SCIENCE*, v. 288, n. 5469, p. 1251–1253, MAY 19 2000. ISSN 0036-8075. Citado 2 vezes nas páginas 29 e 30.
- ORGOGOZO, V.; BROMAN, K.; STERN, D. High-resolution quantitative trait locus mapping reveals sign epistasis controlling ovariole number between two *Drosophila* species. *Genetics*, v. 173, n. 1, p. 197–205, MAY 2006. ISSN 0016-6731. Citado na página 87.
- ORR, H. The population genetics of adaptation: The adaptation of DNA sequences. *Evolution*, v. 56, n. 7, p. 1317–1330, JUL 2002. ISSN 0014-3820. Citado na página 34.
- OTTO, S. P.; GERSTEIN, A. C. The evolution of haploidy and diploidy. *Current Biology*, v. 18, n. 24, p. R1121–R1124, DEC 23 2008. ISSN 0960-9822. Citado na página 35.
- PAL, C. et al. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, v. 450, n. 7172, p. 1079–1081, DEC 13 2007. ISSN 0028-0836. Citado na página 29.
- PALMER, A. C. et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.*, v. 6, JUN 2015. ISSN 2041-1723. Citado na página 74.
- PARK, S.-C.; SIMON, D.; KRUG, J. The speed of evolution in large asexual populations. *Journal of Statistical Physics*, v. 138, n. 1-3, p. 381–410, 2010. Citado na página 36.
- PATRICK, C. et al. Biophysical fitness landscape of the SARS-CoV-2 delta variant receptor binding domain. *JOURNAL OF MOLECULAR BIOLOGY*, v. 434, n. 13, JUL 15 2022. ISSN 0022-2836. Citado na página 28.
- PERFEITO, L. et al. Adaptive mutations in bacteria: high rate and small effects. *Science*, v. 317, n. 5839, p. 813–815, AUG 10 2007. ISSN 0036-8075. Citado na página 35.
- PISTONO, P. E. et al. Fitness landscape-guided engineering of locally supercharged virus-like particles with enhanced cell uptake properties. *ACS CHEMICAL BIOLOGY*, 2022. ISSN 1554-8929. Citado na página 28.
- POKUSAIEVA, V. O. et al. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLOS GENETICS*, v. 15, n. 4, APR 2019. ISSN 1553-7404. Citado na página 28.
- PRUNIER, A. et al. High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. *JOURNAL OF INFECTIOUS DISEASES*, v. 187, n. 11, p. 1709–1716, JUN 1 2003. ISSN 0022-1899. 41st Interscience Conference on Antimicrobial Agents and Chemotherapy, CHICAGO, ILLINOIS, DEC 16-20, 2001. Citado na página 29.
- RICHARDSON, A. et al. Mutator clones of *Neisseria meningitidis* in epidemic serogroup a disease. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, v. 99, n. 9, p. 6103–6107, APR 30 2002. ISSN 0027-8424. Citado na página 29.
- RIO, M. B. D.; STEPHENS, C. R.; ROSENBLUETH, D. A. Fitness landscape epistasis and recombination. *Advances in Complex Systems*, v. 18, n. 7-8, 2015. Citado na página 36.

- RODRIGUES, J. V. et al. Biophysical principles predict fitness landscapes of drug resistance. *Proc. Natl. Acad. Sci. U.S.A.*, v. 113, n. 11, p. E1470–E1478, MAR 15 2016. ISSN 0027-8424. Citado na página 74.
- ROSENBERG, S.; THULIN, C.; HARRIS, R. Transient and heritable mutators in adaptive evolution in the lab and in nature. *Genetics*, v. 148, n. 4, p. 1559–1566, APR 1998. ISSN 0016-6731. Citado na página 30.
- SAAKIAN, D. B. Exact solution of mutator model with linear fitness and finite genome length. *Journal of the Physical Society of Japan*, v. 86, n. 8, 2017. Citado na página 36.
- SAAKIAN, D. B. Quasispecies model of evolution with migration. *Physical Review E*, v. 98, n. 1, 2018. Citado na página 36.
- SAAKIAN, D. B.; GHAZARYAN, M. H.; HU, C.-K. Punctuated equilibrium and shock waves in molecular models of biological evolution. *Physical Review E*, v. 90, n. 2, 2014. Citado na página 36.
- SAAKIAN, D. B.; KIRAKOSYAN, Z.; HU, C.-K. Biological evolution in a multidimensional fitness landscape. *Physical Review E*, v. 86, n. 3, 1, 2012. Citado na página 36.
- SCHENK, M. F. et al. Population size mediates the contribution of high-rate and large-benefit mutations to parallel evolution. *NATURE ECOLOGY & EVOLUTION*, v. 6, n. 4, p. 439+, APR 2022. ISSN 2397-334X. Citado na página 110.
- SHAVER, A. et al. Fitness evolution and the rise of mutator alleles in experimental escherichia coli populations. *Genetics*, v. 162, n. 2, p. 557–566, OCT 2002. ISSN 0016-6731. Citado 2 vezes nas páginas 29 e 30.
- SILVA, R. F. et al. Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. *PLOS Genetics*, v. 7, n. 7, JUL 2011. ISSN 1553-7404. Citado na página 87.
- SINAI, S. et al. Discovery of highly epistatic aav capsid variants through machine-guided navigation of the fitness landscape. *MOLECULAR THERAPY*, v. 30, n. 4, 1, p. 429, APR 2022. ISSN 1525-0016. Citado na página 28.
- SNIEGOWSKI, P. et al. The evolution of mutation rates: separating causes from consequences. *BIOESSAYS*, v. 22, n. 12, p. 1057–1066, DEC 2000. ISSN 0265-9247. Citado na página 29.
- SNIEGOWSKI, P.; GERRISH, P.; LENSKI, R. Evolution of high mutation rates in experimental populations of e-coli. *Nature*, v. 387, n. 6634, p. 703–705, JUN 12 1997. ISSN 0028-0836. Citado na página 29.
- SONG, S.; ZHANG, J. Unbiased inference of the fitness landscape ruggedness from imprecise fitness estimates. *EVOLUTION*, v. 75, n. 11, p. 2658–2671, NOV 2021. ISSN 0014-3820. Citado na página 28.
- SOO, V. W. C. et al. Fitness landscape of a dynamic rna structure. *PLOS GENETICS*, v. 17, n. 2, FEB 2021. ISSN 1553-7404. Citado na página 28.

- SVENSSON, E. I.; CALSBEEK, R. *The Adaptive Landscape in Evolutionary Biology*. [S.l.]: Oxford Press, 2012. Citado na página 28.
- SZOLLOSI, G. J.; DERENYI, I. The effect of recombination on the neutral evolution of genetic robustness. *Mathematical Biosciences*, v. 214, n. 1-2, p. 58–62, 2008. Citado na página 36.
- TADDEI, F. et al. Role of mutator alleles in adaptive evolution. *NATURE*, v. 387, n. 6634, p. 700–702, JUN 12 1997. ISSN 0028-0836. Citado 2 vezes nas páginas 17 e 45.
- TAMER, Y. T. et al. High-order epistasis in catalytic power of dihydrofolate reductase gives rise to a rugged fitness landscape in the presence of trimethoprim selection. *MOLECULAR BIOLOGY AND EVOLUTION*, v. 36, n. 7, p. 1533–1550, JUL 2019. ISSN 0737-4038. Citado 2 vezes nas páginas 28 e 74.
- TENAILLON, O. et al. Mutators and sex in bacteria: Conflict between adaptive strategies. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, v. 97, n. 19, p. 10465–10470, SEP 12 2000. ISSN 0027-8424. Citado na página 29.
- THOMPSON, D. A.; DESAI, M. M.; MURRAY, A. W. Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. *Current Biology*, v. 16, n. 16, p. 1581–1590, AUG 22 2006. ISSN 0960-9822. Citado na página 29.
- TIAN, P.; BEST, R. B. Exploring the sequence fitness landscape of a bridge between protein folds. *PLOS COMPUTATIONAL BIOLOGY*, v. 16, n. 10, OCT 2020. ISSN 1553-734X. Citado na página 28.
- TROBNER, W.; PIECHOCKI, R. Competition growth between escherichia-coli mutl and mut+ in continuously growing cultures. *Zeitschrift fur Allgemeine Mikrobiologie*, v. 21, n. 4, p. 347–349, 1981. ISSN 0044-2208. Citado na página 29.
- VISSER, J. de et al. Diminishing returns from mutation supply rate in asexual populations. *Science*, v. 283, n. 5400, p. 404–406, JAN 15 1999. ISSN 0036-8075. Citado 2 vezes nas páginas 29 e 30.
- VLADAR, H. P. de; BARTON, N. H. The contribution of statistical physics to evolutionary biology. *Trends Ecol. Evol.*, v. 26, n. 8, p. 424–432, AUG 2011. ISSN 0169-5347. Citado 2 vezes nas páginas 26 e 27.
- WATSON, M.; BURNS, J.; SMITH, A. Hypermutable haemophilus influenzae with mutations in muts are found in cystic fibrosis sputum. *MICROBIOLOGY-SGM*, v. 150, n. 9, p. 2947–2958, SEP 2004. ISSN 1350-0872. Citado na página 29.
- WEINREICH, D. et al. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, v. 312, n. 5770, p. 111–114, APR 7 2006. ISSN 0036-8075. Citado 2 vezes nas páginas 35 e 110.
- WEINREICH, D.; WATSON, R.; CHAO, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, v. 59, n. 6, p. 1165–1174, JUN 2005. ISSN 0014-3820. Citado na página 87.
- WOODCOCK, G.; HIGGS, P. Population evolution on a multiplicative single-peak fitness landscape. *J. Theor. Biol.*, v. 179, n. 1, p. 61–73, 1996. Citado na página 58.

- WU, Z. et al. Expression level is a major modifier of the fitness landscape of a protein coding gene. *NATURE ECOLOGY & EVOLUTION*, v. 6, n. 1, p. 103+, JAN 2022. ISSN 2397-334X. Citado na página 28.
- YAKUSHKINA, T.; SAAKIAN, D. B. Modeling evolution on nearly neutral network fitness landscapes. *Journal of the Physical Society of Japan*, v. 86, n. 8, 2017. Citado na página 36.
- YAKUSHKINA, T.; SAAKIAN, D. B. The exact solution of the mutator model with directed mutations, linear fitness function, and finite genome length. *Chinese Journal of Physics*, v. 56, n. 5, p. 2235–2240, 2018. Citado na página 36.
- YAMPOLSKY, L.; STOLTZFUS, A. Bias in the introduction of variation as an orienting factor in evolution. *Evol. Dev.*, v. 3, n. 2, p. 73–83, MAR-APR 2001. ISSN 1520-541X. Citado na página 73.
- YANG, G. et al. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *NATURE CHEMICAL BIOLOGY*, v. 15, n. 11, p. 1120+, NOV 2019. ISSN 1552-4450. Citado na página 28.
- YANG, G. et al. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme (vol 51, pg 831, 2020). *NATURE CHEMICAL BIOLOGY*, v. 16, n. 8, p. 930, AUG 1 2020. ISSN 1552-4450. Citado na página 28.

Apêndices

APÊNDICE A – Ploidia e propriedades de fixação

Neste Apêndice iremos primariamente demonstrar que propriedades do processo de fixação apenas sob o efeito da seleção, a saber, sua probabilidade e seu tempo, não variam entre uma população de indivíduos haplóides e outra de diplóides. Este é o objetivo pragmático, pois precisamos usar fórmulas do caso diplóide nas nossas situações haplóides.

No caso mais geral, podemos admitir que no processo de Wright-Fisher se forma um pool infinito de gametas, formando um número infinito de zigotos com ploidia qualquer, dos quais apenas um número fixo total de cromossomos é selecionado. As propriedades de fixação não mudam com as ploidias.

Apesar da artificialidade dessa condição geral, ela também implica que se uma população tem ploidia fixa, a condição de “número fixo total de cromossomos” pode ser trocada por “número fixo total de indivíduos”, que é uma condição muito mais natural.

Nossa estratégia é a seguinte: primeiramente demonstramos que as fórmulas para a probabilidade e o tempo de fixação dependem apenas de duas funções, $m(p)$ e $v(p)$, onde p é a frequência do alelo na população, $m(p)$ é o salto médio em frequência a partir de p , e $v(p)$ pode ser entendida como a variância do salto a partir de p . Em seguida mostramos que $m(p)$ e $v(p)$ não dependem da ploidia z no regime difusivo ($N \gg 1$ e $s \ll 1$), e que como consequência as propriedades de fixação estudadas também não mudam. Finalmente, explicamos estes resultados, e sugerimos outros casos em que estas propriedades também não dependam da ploidia e a extrapolação da conservação de outras propriedades de fixação com a mudança de ploidia.

A.1 Probabilidade de fixação

Nesta seção apresentaremos a dedução da fórmula da probabilidade de fixação acompanhando o livro de Gillespie (GILLESPIE, 2004), e concluiremos que ela depende apenas de $m(p)$ e $v(p)$. Esta última variável será relacionada com a variância do tamanho do salto em frequência de uma geração para outra.

A probabilidade de fixação $\pi(p)$ é igual a soma da probabilidade dos possíveis saltos em frequência Δp vezes a probabilidade da população se fixar com frequência $p + \Delta p$. Isso pode ser expresso da seguinte forma:

$$\pi(p) = \sum_{\Delta p} \text{Prob}(\Delta p) \pi(p + \Delta p). \quad (\text{A.1})$$

Isto pode ser interpretado como a esperança da probabilidade de fixação $\pi(p + \Delta p)$ em termos da probabilidade de salto $\text{Prob}(\Delta p)$. Em termos estatísticos,

$$\pi(p) = E_{\Delta p} \pi(p + \Delta p). \quad (\text{A.2})$$

A probabilidade de fixação após o salto $\pi(p + \Delta p)$ pode ser expandida em série de Taylor em torno de p . Truncado a série em três termos, temos que

$$\pi(p + \Delta p) \approx \pi(p) + \pi'(p)\Delta p + \frac{1}{2}\pi''(p)(\Delta p)^2. \quad (\text{A.3})$$

Tomando-se a esperança de ambos os lados da igualdade, temos aproximadamente que

$$\pi(p) \approx \pi(p) + \pi'(p)E\{\Delta p\} + \frac{1}{2}\pi''(p)E\{(\Delta p)^2\}. \quad (\text{A.4})$$

Isto nos leva à seguinte equação:

$$\frac{1}{2}E\{(\Delta p)^2\}\pi''(p) + E\{\Delta p\}\pi'(p) = 0. \quad (\text{A.5})$$

Definimos $m(p) = E\{\Delta p\}$, isto é, a esperança do salto a partir de p , e $v(p) = E\{(\Delta p)^2\}$. Mas temos que

$$v(p) = E\{(\Delta p)^2\} = \text{Var}(p) + E\{\Delta p\}^2, \quad (\text{A.6})$$

onde $\text{Var}(p)$ é a variância da frequência p . Na Equação A.3 da série truncada, admitimos implicitamente que Δp é pequeno. Logo, a esperança ao quadrado $E\{\Delta p\}^2$ é esperada ser ainda menor, e $v(p)$ pode ser na prática interpretado como a variância em p .

A equação diferencial que modela o sistema se torna:

$$\frac{1}{2}v(p)\pi''(p) + m(p)\pi'(p) = 0, \quad (\text{A.7})$$

com condições de contorno $\pi(0) = 0$ e $\pi(1) = 1$. Para resolver a equação acima, façamos primeiramente $f(p) = \pi'(p)$. Após uma pequena manipulação algébrica, obtemos

$$f'(p) + 2\frac{m(p)}{v(p)}f(p) = 0. \quad (\text{A.8})$$

Esta é uma equação diferencial linear de primeira ordem, que possui como fator integrante bem estabelecido

$$h(p) = e^{B(p)}, \quad (\text{A.9})$$

com

$$B(p) = 2 \int_0^p \frac{m(x)}{v(x)} dx. \quad (\text{A.10})$$

Com o fator integrante, a fórmula torna-se em

$$\frac{d}{dp} [f(p)h(p)] = 0. \quad (\text{A.11})$$

Logo,

$$f(p) = \pi'(p) = C_1 e^{-B(p)}, \quad (\text{A.12})$$

o que nos leva a

$$\pi(p) = C_1 \int_0^p e^{-B(y)} dy + C_2. \quad (\text{A.13})$$

Usando-se a condição de contorno $\pi(0) = 0$, temos que $C_2 = 0$. Para a condição $\pi(1) = 1$ obtemos

$$C_1^{-1} = \int_0^1 e^{-B(y)} dy, \quad (\text{A.14})$$

o que nos dá

$$\pi(p) = \frac{\int_0^p e^{-B(y)} dy}{\int_0^1 e^{-B(y)} dy}. \quad (\text{A.15})$$

É bem conhecido para diplóides (EWENS, 2004) que $m(p) = spq/2$ e que $v(p) = pq/2N$, com s aqui sendo o coeficiente de seleção do homozigoto mutante na ausência de dominância (mais adiante redefiniremos este coeficiente como sendo o de cada alelo mutante, com aptidão aditiva), N é o número de indivíduos diplóides (N também será redefinido como o número de alelos), e $q = 1 - p$. Como pretendemos provar que este resultado não depende da ploidia, esta dedução pode ser inicialmente feita com diplóides. Assim,

$$B(y) = 2 \int_0^y \frac{spq/2}{pq/2N} dp = 2Nsy. \quad (\text{A.16})$$

A fórmula para a probabilidade de fixação torna-se

$$\pi(p) = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}. \quad (\text{A.17})$$

Apesar das redefinições de s e N posteriormente, esta fórmula não será alterada.

A.2 Tempo de fixação

A dedução para o tempo de fixação será calculada, sem perda de generalidade, a partir de indivíduos diplóides, acompanhando a dedução de Ewens (EWENS, 2004). Da mesma forma que demonstramos que a probabilidade de fixação depende apenas de $m(p)$ e $v(p)$, provaremos aqui que o mesmo se sucede ao tempo de fixação.

Seja $f(x; p, t)$ a probabilidade de uma população mudar da frequência p para a x antes do, ou no tempo t . Esta probabilidade irá obedecer à seguinte relação:

$$f(x; p, t + \delta t) = \int_0^1 g(\delta p; p) f(x; p + \delta p, t) d(\delta p), \quad (\text{A.18})$$

onde δp é a mudança no valor da variável aleatória no intervalo de tempo $(0, \delta t)$, e $g(\delta p; p)$ é a função densidade probabilidade para a mudança δp a partir de p . Se expandirmos ambos os lados em séries de Taylor e retivermos os primeiros termos, chegaremos à equação

$$\frac{\partial f(x; p, t)}{\partial t} = m(p) \frac{\partial f(x; p, t)}{\partial p} + \frac{1}{2} v(p) \frac{\partial^2 f(x; p, t)}{\partial p^2}. \quad (\text{A.19})$$

Se definimos

$$F(x; p, t) = \int_0^x f(y; p, t) dy, \quad (\text{A.20})$$

a partir da Equação A.19 obtemos

$$\frac{\partial F(x; p, t)}{\partial t} = m(p) \frac{\partial F(x; p, t)}{\partial p} + \frac{1}{2} v(p) \frac{\partial^2 F(x; p, t)}{\partial p^2}. \quad (\text{A.21})$$

Suponha que $x = 0$ e $x = 1$ são estados absorventes do processo de difusão. Da Equação A.21 chegamos à seguinte equação:

$$\frac{\partial P_0(p, t)}{\partial t} = m(p) \frac{\partial P_0(p, t)}{\partial p} + \frac{1}{2} v(p) \frac{\partial^2 P_0(p, t)}{\partial p^2}. \quad (\text{A.22})$$

onde $P_0(p, t)$ é a probabilidade de que a absorção em zero ocorreu antes do ou no tempo t . A probabilidade $P_1(p, t)$ de que a absorção em um ocorreu obedece a mesma equação. O que muda entre P_0 e P_1 são as condições de contorno.

Nós consideramos agora que o tempo médio até que um ou outro estado absorvente seja alcançado no processo de difusão de interesse. Seja $\phi(t; p)$ a função densidade do tempo t até que a absorção ocorra. Assim, a Equação A.22 e a equação análoga para $P_1(p, t)$ nos mostram que ϕ obedece à seguinte equação:

$$\frac{\partial \phi(t; p)}{\partial t} = m(p) \frac{\partial \phi(t; p)}{\partial p} + \frac{1}{2} v(p) \frac{\partial^2 \phi(t; p)}{\partial p^2}. \quad (\text{A.23})$$

Pela definição de $\phi(t; p)$,

$$\begin{aligned} -1 &= - \int_0^\infty \phi(t; p) dt \\ &= - [t\phi(t; p)]_0^\infty + \int_0^\infty t \frac{\partial \phi}{\partial t} \\ &= 0 + \int_0^\infty t \left\{ m(p) \frac{\partial \phi}{\partial p} + \frac{1}{2} v(p) \frac{\partial^2 \phi}{\partial p^2} \right\} dt, \end{aligned} \quad (\text{A.24})$$

de maneira que

$$-1 = m(p) \frac{d\bar{t}(p)}{dp} + \frac{1}{2} v(p) \frac{\partial^2 \bar{t}(p)}{\partial p^2}. \quad (\text{A.25})$$

Note que, aqui,

$$\bar{t}(p) = \int_0^\infty t \phi(t; p) dt. \quad (\text{A.26})$$

é o tempo médio até que um ou outro estado absorvente seja alcançado, dada a frequência inicial p . A solução da Equação A.25, com condições de contorno $\bar{t}(0) = 0$ e $\bar{t}(1) = 0$, é melhor expressa na forma

$$\bar{t}(p) = \int_0^1 t(x; p) dx, \quad (\text{A.27})$$

onde

$$t(x; p) = 2P_0(p) [v(x)h(x)]^{-1} \int_0^x h(y) dy, \quad 0 \leq x \leq p, \quad (\text{A.28})$$

$$t(x; p) = 2P_1(p) [v(x)h(x)]^{-1} \int_x^1 h(y) dy, \quad p \leq x \leq 1. \quad (\text{A.29})$$

A função $h(x)$ é o fator integrante encontrado na Seção A.1. Para o processo condicional à fixação $x = 1$, a função tempo de permanência (*soujorn time*) acima deve ser modificada para

$$t^*(x; p) = t(x; p)P_1(x)/P_1(p). \quad (\text{A.30})$$

Para o tempo de permanência até a absorção em $x = 0$, a fórmula é análoga, trocando-se P_1 por P_0 . Assim, demonstramos que tanto o tempo para absorção em qualquer uma das barreiras $x = 0$ ou $x = 1$, ou para cada uma delas separadamente, depende apenas de $m(p)$ e de $v(p)$. Juntamente com a dedução para a probabilidade de fixação, demonstramos que propriedades-chave da fixação dependem apenas de $m(p)$ e de $v(p)$. Agora é hora de demonstrarmos que essas funções não dependem da ploidia.

A.3 Independência da ploidia

Agora vamos demonstrar que a média $m(p)$ e a variância $v(p)$ não dependem da ploidia z no limite difusivo ($N \gg 1$ e $|s| \ll 1$) quando, para o início da próxima geração, forma-se uma sopa de gametas haplóides. Isso ocorre comumente com organismos diplóides ($z = 2$), e trivialmente com haplóides ($z = 1$).

Imaginemos que existam apenas dois tipos de alelos: o alelo selvagem, que contribui aditivamente com um coeficiente de seleção nulo e o alelo mutante, que contribui aditivamente com um coeficiente de seleção s . Não imporemos nenhuma condição sobre s além daquela do limite difusivo, o que quer dizer que ele pode assumir valores adaptativos típicos, quase neutros, ou deletérios não muito fortes. A aptidão de um indivíduo z -plóide carregando $0 \leq k \leq z$ alelos mutantes é dada por

$$w(k) = 1 + sk. \quad (\text{A.31})$$

Admitindo-se uma amostragem aleatória dos gametas a cada geração, um genótipo antes da seleção terá k alelos com uma distribuição binomial $B(l, x)$, onde l é o número de alelos mutantes em um dado genótipo, e $x = i/N$ é a frequência do alelo mutante na sopa. A variável N é o número total de alelos e i o número de alelos mutantes nesta sopa. Assim, a aptidão média \bar{w} será dada por

$$\begin{aligned} \bar{w} &= \sum_{l=0}^z w(l) \binom{z}{l} \left(\frac{i}{N}\right)^l \left(\frac{N-i}{N}\right)^{z-l} \\ &= 1 + szx. \end{aligned} \quad (\text{A.32})$$

A quantidade w_m a seguir é a média da frequência dos alelos após a seleção, que é multiplicada pela aptidão de seu genótipo. Ela precisará ser normalizada pela aptidão média \bar{w} para resultar na frequência do alelo mutante após a seleção.

$$\begin{aligned} w_m &= \sum_{l=0}^z \frac{l}{z} w(l) \binom{z}{l} \left(\frac{i}{N}\right)^l \left(\frac{N-i}{N}\right)^{z-l} \\ &= x\{1 + s[1 + (z-1)x]\}. \end{aligned} \quad (\text{A.33})$$

Temos que $p'(x) = w_m/\bar{w}$ é a frequência esperada ao fim da seleção. Assim, o passo médio é dado por $m(x) = p' - x$. Expandindo-se esse resultado em primeira ordem em s e $1/N$ a expressão resultante, obtemos que

$$m(x) = sx(1 - x), \quad (\text{A.34})$$

que não depende da ploidia z . Vamos agora calcular a variância. Ela é dada por

$$\begin{aligned} v(x) &= \sum_{j=0}^N \left(\frac{j-i}{N}\right)^2 \binom{j}{N} p'(x)^j (1-p'(x))^{N-j} \\ &= \left(\frac{1}{N} - 2x\right) p'(x) + \left(1 - \frac{1}{N}\right) p'(x)^2 + x^2. \end{aligned} \quad (\text{A.35})$$

Novamente, expandindo-se esta solução em termos de s e $1/N$ até primeira ordem, obtemos

$$v(x) = \frac{1}{N}x(1 - x), \quad (\text{A.36})$$

que também não depende da ploidia. Assim provamos que propriedades importantes da fixação, a saber, os tempos e a probabilidade de fixação, não dependem da ploidia quando a amostragem de alelos se dá de uma sopa de gametas. Isto se aplica em especial para haplóides e diplóides. A base para a independência da ploidia é que no regime difusivo os alelos se comportam independentemente. Isto nos sugere que propriedades de fixação de sistemas com mais do que dois alelos e na presença de mutação não dependam também da ploidia.

Kimura (KIMURA, 1962) já havia argumentado que sua dedução da fórmula de fixação não fazia menção à ploidia, e que portanto era geral para qualquer uma delas. Aqui apenas explicitamos o porquê das propriedades de fixação não dependerem dela.

Anexos

Hidden role of mutations in the evolutionary processAlexandre de Aquino Soares^{✉*} and Lucas Wardil*Departamento de Física, Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais (UFMG), 31270-901 Belo Horizonte, Minas Gerais, Brazil*

Louis Bernard Klaczko

Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas (Unicamp), C. P. 6109, 13083-970 Campinas, São Paulo, Brazil

Ronald Dickman

Departamento de Física and National Institute of Science and Technology for Complex Systems, Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais (UFMG), C. P. 702, 30123-970 Belo Horizonte, Minas Gerais, Brazil (Received 1 February 2021; revised 5 October 2021; accepted 5 October 2021; published 25 October 2021)

Mutations not only alter allele frequencies in a genetic pool but may also determine the fate of an evolutionary process. Here we study which allele fixes in a one-step, one-way model including the wild type and two adaptive mutations. We study the effect of the four basic evolutionary mechanisms—genetic drift, natural selection, mutation, and gene flow—on mutant fixation and its kinetics. Determining which allele is more likely to fix is not simply a question of comparing fitnesses and mutation rates. For instance, if the allele of interest is less fit than the other, then not only must it have a greater mutation rate, but also its mutation rate must exceed a specific threshold for it to prevail. We find exact expressions for such conditions. Our conclusions are based on the mathematical description of two extreme but important regimes, as well as on simulations.

DOI: [10.1103/PhysRevE.104.044413](https://doi.org/10.1103/PhysRevE.104.044413)**I. INTRODUCTION**

The evolution of biological populations—the changing frequencies of genetic types over time—is driven by four principal forces: genetic drift, natural selection, mutation, and gene flow [1]. Genetic drift corresponds to fluctuations in frequencies with no preferred direction, as in unbiased diffusion. Natural selection drives the composition toward the fittest genotype. Mutation creates new types, guaranteeing genetic diversity. Last, gene flow consists of the exchange of genetic types between populations. In the standard view, mutations introduce novelty but do not determine the direction of evolution [2–4]. Once mutation has created new genetic types, the latter compete in a race determined by fitness, that is, the tendency to increase in frequency in subsequent generations, reflecting fertility, viability, and other factors given the present environment [5].

Nevertheless, mutation can have a more prominent role in evolution. Yampolsky and Stoltzfus [6] proposed a model in which one mutant is fitter while the other arises with a greater likelihood and showed that the prevailing type can shift when mutability (an overall factor multiplying all mutation rates) is varied.

Interplay between mutation rates and fitnesses has been studied via experiment and simulation, as well as analytically [7–11]. As an example of mutation having a determining role, we may cite the class of point mutations, called transitions (mutations from purine to purine or pyrimidine to pyrimidine)

which occur at a higher frequency than transversions (mutations from a purine to a pyrimidine, or vice versa).

Stoltzfus and Norris [11] analyzed the hypothesis that transitions are overrepresented because they are less effective in altering proteins. They show, however, that this effectiveness is weak and thus the prevalence of transitions is not likely to be explained as selection on proteins, favoring the hypothesis that ratios between mutation rates (that is, mutation biases), may direct evolution.

In the context of adaptive evolution (the process by which fitness increases), Stoltzfus *et al.* [7] found that transitions are overrepresented among adaptive mutations. More specifically, in Ref. [9] they studied the molecular basis of convergent increases in the affinity of hemoglobin for oxygen in high-altitude birds and identified causative substitutions for the increased affinity. The dinucleotide consisting of a cytosine followed by a guanine in the same strand (CpG), depending on their frequent modification by methylation, is more prone to mutation than other dinucleotides. They find that a disproportionate number of causative amino acid replacements are attributable to CpG, and because this dinucleotide is associated with a higher mutation rate, their results suggest that mutation bias influences the outcome of molecular adaptation.

Mutational dynamics is commonly analyzed using a stochastic birth-and-death process with a finite population. Because the population is finite, it will eventually become homogeneous; homogeneity can be violated if one or more mutations are possible. The probability of attaining a homogeneous state is called the *fixation probability*. Thus, if mutation rates have an effect on the direction of evolution, then they should affect fixation probabilities.

*asoares@fisica.ufmg.br

There are many stochastic models that can be applied to population genetics [12–27]. Here we analyze the Wright-Fisher model in a population starting with the wild type and evolving to one of two mutants. We study which mutation prevails as the mutation rates are varied and show that mutation rates, fitness, population size, and gene flux are all relevant to this process. We investigate the kinetics of the process as mutability is varied, focusing on two regimes, rare mutations and frequent mutations, and find analytical expressions for specific regimes, establishing thresholds for changes in prevalence.

Furthermore, if we suitably redefine the concept of prevalence in the presence of continuous migration, then the competition between mutants (up to a certain value of the rate of exchange of the population with an external wildtype population) is weaker the greater the gene flow. We study these phenomena using computer simulation and analytical approximations that become exact in specified limits. Additionally, we analyze the time evolution of the composition of the population before fixation and how it depends on the evolutionary parameters. Interestingly, coexistence of the two mutants (or clonal interference [28,29]) favors fixation of the fittest type, whereas if both mutants are quasineutral, then the higher mutation rate may dictate which type wins. The larger the population, the stronger the effect of clonal interference.

This work is organized as follows: In Sec. II we define the model. Then, in Sec. III, we demonstrate that all four mechanisms may lead to distinct evolutionary steps. Following this, we concentrate on the mechanisms of drift, selection, and mutation, for which we have robust analytical and numerical results. In the analytical study, we examine limits of the probabilities of absorption of the mutants, and the conditions for one mutant to prevail over the other in different regimes. Numerical results complement the analysis and are used to probe nonextreme regimes. Finally, in Sec. IV, we summarize our main results, compare our work with the literature, and offer perspectives.

II. THE MODEL

We study a generalization of the Wright-Fisher model [30]. Let N_W^t , N_A^t , and N_B^t be the numbers of individuals of type W (the wild type), A , and B (the two mutants), respectively, at generation t . We set $N_W^t + N_A^t + N_B^t = N$, the fixed population size. The corresponding numbers at generation $t + 1$ are drawn from a multinomial distribution with probabilities proportional to:

$$\begin{aligned} p_W^{t+1} &\propto N_W^t(1 - r_A - r_B) + (w_A N_A^t + w_B N_B^t)f, \\ p_A^{t+1} &\propto (1 - f)w_A N_A^t + N_W^t r_A, \\ p_B^{t+1} &\propto (1 - f)w_B N_B^t + N_W^t r_B, \end{aligned} \quad (1)$$

which are normalized by their sum. The variables p_W^t , p_A^t , and p_B^t are the probabilities of the alleles W , A , and B at time t to be sampled at the next generation. $r_A = r$ and $r_B = r/\rho$ are the mutation rates for mutants A and B , r is the mutation rate or *mutability*, and ρ is the mutation bias. w_A and w_B are the fitnesses of types A and B , and f is the rate of exchange of the population with an external wildtype population. Let us fix $w_B \geq w_A \geq w_W$; the fitness of the wild type, w_W , is set

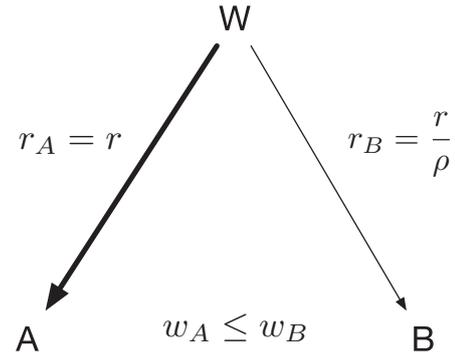


FIG. 1. Model for mutations of wildtype haploid individuals at a single locus. The wild type W can mutate to types A or B . The mutation rate to A (thicker line) is greater than to B , but the fitness of B is higher, $w_W \leq w_A \leq w_B$, in which w_W , the fitness of the wild type, is set to unity. The mutation bias, $\rho > 1$, is set so that $r_A/r_B = \rho$.

to unity. Note that in this article we deal only with average to high values of the fitnesses of the mutants. It is out of our scope to study quasineutral mutations [31–33], which would demand significant changes to our model. It is usual to recast the fitness of the mutants as $w_i = 1 + s_i$, where s_i is called coefficient of selection, with $s_i > 0$, corresponding to adaptive mutations. The actual probability of mutation is equal to $r\Delta t$. But, because the process is discrete with time interval equal to unity, we sometimes refer to r_A and r_B as *mutation rates*.

The basic model is sketched in Fig. 1. The initial, homogeneous population has only wild types. In the presence of one-way mutations and absence of gene flow ($f = 0$), which is the scenario we focus on, the population will certainly reach a state in which all individuals are of type A or all are of type B . Once an all- A or all- B state is attained, there are no further mutations, and the population remains in this state, i.e., the all- A and all- B states are *absorbing*. After the first mutation takes place and before reaching an absorbing state, the population can return, one or more times, to the initial all- W state.

The *fixation probability* of an allele in a heterogeneous population is defined as the probability that the frequency of this allele reaches unity. Because the all- W state may be revisited repeatedly before one of the absorbing states is reached, we make the following distinction between the fixation and the absorption probabilities. The former is the probability to reach a homogeneous state *for the first time* after the first mutation takes place. We denote by π_f^A , π_f^B , and π_f^W the fixation probabilities of each type. The time it takes from the first mutation until fixation is called the *fixation time*. Complementarily, we define the *absorption probability* as the probability to reach an absorbing state (all- A or all- B). We denote by π_a^A , π_a^B , and π_a^W the absorption probabilities of each type, with $\pi_a^W = 0$ and $\pi_a^A + \pi_a^B = 1$. Table I lists the symbols used in this work.

We study this model via simulation, and obtain results for all regimes of mutations, including rare and frequent mutations. Subsequently, we model analytically the latter two regimes, and investigate approximations based on these results.

TABLE I. Symbols for the evolutionary parameters and variables.

$\pi_f^W, \pi_f^A, \pi_f^B$	Fixation probability of types W , A , and B , respectively.
$\pi_a^W, \pi_a^A, \pi_a^B$	Absorption probability of types W , A , and B , respectively. $\pi_a^W = 0$.
t_f	Fixation time.
N	Population size.
N_W^t, N_A^t, N_B^t	Size of subpopulations W , A , and B , respectively, at time t .
p'_W, p'_A, p'_B	Probability of an individual to become type W , A , and B , respectively, at time t .
r	Reference mutation rate. Also the mutability, a common factor that multiplies all mutation rates.
r_A, r_B	Reference mutation rates from the wild type to types A and B , respectively. $r_A = r$.
ρ	Ratio of r_A to r_B , i.e., $r_B = r_A/\rho$.
θ_A, θ_B	Mutation supply of types A and B , equal to $N m r$ and $N m r/\rho$, respectively, in a wildtype population
w_W, w_A, w_B	Fitnesses of types W , A , and B , respectively.
s_A, s_B	Selection coefficients of types A and B . $w_A = 1 + s_A$ and $w_B = 1 + s_B$.
s	Reference selection coefficient. $s_A = s$.
$\omega_A, \omega_B, \omega$	Equal to e^{-2s_A} , e^{-2s_B} , and $\omega_A/\omega_B = e^{2(s_B - s_A)}$, respectively
f	Fraction of mutants that emigrate, and are replaced by wildtype individuals.

In simulations, each realization of the process begins with N wildtype individuals. The generation at time $t + 1$ is formed by sampling independently with replacement with probabilities of sampling the types W , A , and B given by Eqs. (1) (after normalization). The simulation ends when one of the absorbing states (all- A or all- B) is attained. If a type has infinite fitness, then it fixes the generation after it appears. We calculate means and uncertainties over samples of 2^{21} ($\gtrsim 2 \times 10^6$) independent realizations. Population sizes varied from a few individuals up to 10^9 . The sampling algorithm for the multinomial distribution corresponds to the command “`gsl_ran_multinomial`” from the GNU Scientific Library [34].

Last, we can show that if we define the selection coefficient of a single mutant allele of an n -ploid under additive fitness and we fix the number of chromosomes, then the fixation probability and fixation time are invariant with the ploidy [30]. Thus, our results remain valid for diploids and haploids if we make the transformation $2N \rightarrow N$, $s \rightarrow 2s$ to go from diploid to haploid organisms.

III. RESULTS AND DISCUSSION

This section is organized as follows. In Sec. III A we analyze the absorption probability and in Sec. III B the fixation probability and fixation times. Next, in Secs. III C and III D, respectively, we develop analytical approximations for two limiting regimes, rare and frequent mutations, and find the conditions under which the least-fit mutant prevails. Finally, in Sec. III E we extend the basic model to include more than two mutants or a slightly deleterious mutation.

A. Absorption probability

Simulation results illustrating the effects of the mutation rate and population size on the absorption probability, as well as the effect of fitness and genetic flux on the steady-state prevalence of the least-fit mutant are shown in Fig. 2. The absorption probability of a type with higher mutation rate and smaller fitness (in this case, type A) is typically close to unity if the mutation rate is low but tends to zero as the mutation

rate attains higher values. The reason is that, at high mutation rates, the two types can coexist and the advantage of the fitter type may dominate evolution, depending on how close are the fitnesses of the mutants. Interestingly, even if the fittest type B has an infinite fitness value, the type A with greater mutation rate and lower fitness still may have an absorption probability close to unity for low mutation rates, as shown in Fig. 2(a). On the other hand, type A no longer has an advantage if the population size is large. In large populations, the number of individuals of the fitter type B that appear due to mutation is higher and, consequently, they take over the population with higher probability, as shown in Fig. 2(b). Figure 2(c) illustrates that increasing the fitness of A (increasing s_A) in the rare-mutation regime increases its absorption probability, π_a^A , as expected.

In the presence of migration, there is a continuous flux of W types and hence no absorbing states. The dominant mutant can be defined as the more populous one, more specifically, the one that first reaches more than half of the total population size. Figure 2(d) shows that the less-fit type is benefited by an increase in the migration rate. The frequent gene flow removal of mutants is compensated by the higher mutation rate of type A .

To better understand the simulation results, let us develop an analytical approximation for the fixation probabilities of the mutants. This approximation can then be inserted in the relation,

$$\pi_a^i = \frac{\pi_f^i}{\pi_f^A + \pi_f^B}, \quad (2)$$

where $i \in \{A, B\}$, to obtain the absorption probabilities. Suppose that all mutations that may happen take place at a single generation, an assumption that simplifies the analysis. The probability of a set of subpopulations, (N_W, N_A, N_B) , in the presence of mutation and drift in this same generation is

$$p(N_W, N_A, N_B) = \frac{N!}{N_W!N_A!N_B!} (1 - r - r/\rho)^{N_W} r^{N_A} (r/\rho)^{N_B}. \quad (3)$$

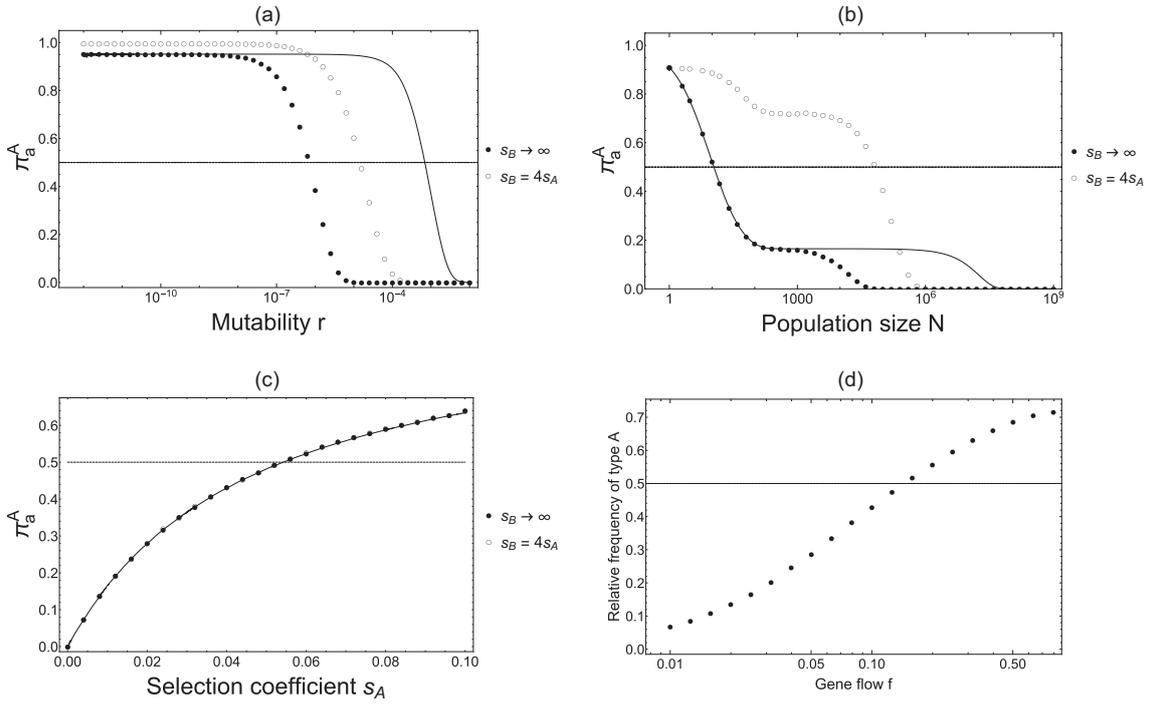


FIG. 2. Dependence of the prevalence of type A on the mutability r , the population size N , the fitness (in terms of the selection coefficient s_A) and the gene flow f . The disks and circles represent results from simulations. The solid disks (\bullet) represent an infinite fitness for type B and the empty circles (\circ) represent a selection coefficient of B four times higher than type A . The continuous line ($—$) represents the analytical approximation, Eq. (5). In panel (d), “relative frequency” means the frequency of A relative to both mutants in the stationary state. The selection coefficients s_A and s_B are given by $s_A = w_A - 1$ and $s_B = w_B - 1$. Parameters are (a) $N = 10^6$, $s_A = 0.01$, $\rho = 1000$; (b) $r = 1 \times 10^{-6}$, $s_A = 0.01$, $\rho = 10$; (c) $N = 10^6$, $r = 1 \times 10^{-12}$, $\rho = 10$; and (d) $N = 100$, $\rho = 1.2$, $w_A = 1.01$, $w_B = 1.04$.

When type B has an infinite fitness, type A will fix only if type B does not arise, that is, $N_B = 0$. In the presence of selection only, the probability of fixation of N_A individuals in a population with $N - N_A$ wild individuals is given by the classic result [30,35]

$$\Pi_A(N_A) = \frac{1 - e^{-2N_A s_A}}{1 - e^{-2N s_A}} = \frac{1 - \omega_A^{N_A}}{1 - \omega_A^N}, \quad (4)$$

where $w_A = 1 + s_A$ (s_A is called the selection coefficient) and $\omega_A = e^{-2s_A}$.

Thus, the probability of fixation of A can be approximated as the product of the probability of a configuration ($N_W, N_A, 0$) and the probability of fixation departing from that configuration, summed over all configurations, that is,

$$\begin{aligned} \pi_f^A &= \sum_{N_A=1}^N \frac{N!}{N_W! N_A!} (1 - r - r/\rho)^{N_W} r^{N_A} \frac{1 - \omega_A^{N_A}}{1 - \omega_A^N} \\ &= \frac{\left[1 - \frac{(\rho+1)r}{\rho}\right]^N}{\omega_A^N - 1} \left\{ \left[1 - \frac{\rho r \omega_A}{\rho(r-1) + r}\right]^N \right. \\ &\quad \left. - \left[\frac{r - \rho}{\rho(r-1) + r}\right]^N \right\}. \end{aligned} \quad (5)$$

For a very small mutability r , the term with $N_A = 1$ dominates the sum, and the expression reduces to

$$\pi_f^A = \theta_A \frac{1 - \omega_A}{1 - \omega_A^N},$$

where $\theta_A = Nr$ is the mutation supply of type A in a wildtype population, and the second factor is the probability of a single mutant A to fix. The restriction to $r \ll 1/N$ indicates that the absorption of type A is strongly affected by the mutants arising in the first generations, that is, by the initial mutation supply.

Similarly, the probability of fixation of B is given by the probability that at least one mutant B arises ($N_B > 0$) and does not depend on the presence of type A , that is,

$$\begin{aligned} \pi_f^B &= \sum_{N_A=0}^{N-1} \sum_{N_B=1}^{N-N_A} \frac{N!}{N_W! N_A! N_B!} (1 - r - r/\rho)^{N_W} r^{N_A} (r/\rho)^{N_B} \\ &= \left[1 - \frac{(\rho+1)r}{\rho}\right]^N \left\{ \left(\frac{1}{1-r}\right)^N \left[\frac{\rho(r-1)}{\rho(r-1) + r}\right]^N \right. \\ &\quad \left. - \left[\frac{r - \rho}{\rho(r-1) + r}\right]^N \right\}. \end{aligned}$$

Again, for a very small mutability r , this formula simplifies to

$$\pi_f^B = \theta_B,$$

where $\theta_B = \theta_A/\rho$ is the mutation supply of type B in a wildtype population. Comparison with the small- r limit of the

absorption probability of type *A*, shows that the mutation supply is a determinant factor. However, for type *B*, it is as if there were a fixation probability term equal to unity multiplying the mutation supply. In other words, the small-*r* approximation shows that although type *A* has a small fixation probability, this can be compensated by a greater mutation supply.

The expressions of the absorption probabilities are plotted as solid lines in Figs. 2(a)–2(c). They work well in the low-mutability regime, but at high mutability they overestimate π_a^A . This is explained by the more important role of recurrent mutations at intermediate mutabilities, facilitating absorption by *B*. Naturally, the higher the fitness of *B*, the stronger the effect. For sufficiently high mutabilities, the continuous curve coincides with the solid points: In this case essentially all mutants appear at the first generation of the fixation process.

The assumption used to derive the analytical approximation sheds light on the mechanism driving absorption. For example, the good agreement of the approximation for small *r* indicates that the first generation of mutants is determinant for the fixation success of type *A*. Considering mutational supply, that is, the product of population size times total mutation rate, we see that with a small mutation supply ($Nr \ll 1$) the dynamics is well described as “one mutant at a time.” By contrast, the disagreement between the approximation and the simulation at high levels of mutation supply indicates the crucial role of recurrent mutations, which occur only because there are enough wild types that can mutate, even after the first generation of mutants. The plateau observed in Fig. 2(b) suggests that recurrent mutations have no role in the absorption process. However, at some point, for sufficiently large population sizes, recurring mutations are responsible for the abrupt decrease of π_a^A .

B. Fixation probability and fixation times

The route to absorption is not always direct. Mutants can arise and go extinct before they spread. The fixation probability of the wild type is a measure of the likelihood of the mutations to disappear before they can fix. The fixation time of type *A* or *B* is the number of generations between the occurrence of the first mutation of that type and its fixation. It is also possible to define the fixation time of the wild type as the number of generations between the occurrence of the first mutation, either *A* or *B*, and the fixation of *W*. Figure 3 shows the dependence of the fixation probability and fixation times on the mutation rate.

At low mutation rates, the wild type fixes more than at other mutabilities because of its numerical superiority over the rare mutant. Although the fixation probability of the mutants seems to be low for lower mutation rates, notice that, in the end, no wild type remains. Because the fixation time of *W* measures the time to return to the homogeneous state after it was modified by a single mutant, one expects its fixation time to be very small compared to the other mutants, as shown in Fig. 3. Indeed, fixation of the wild type can be seen as a renewal process [36] in which the population can return to the all-*W* state after the first mutation. The probability that the first mutation is of type *A* or type *B* is equal to $r_A/(r_B + r_A)$ or $r_B/(r_B + r_A)$, respectively. Thus, in the rare mutation regime, the fixation probability of the wild type is

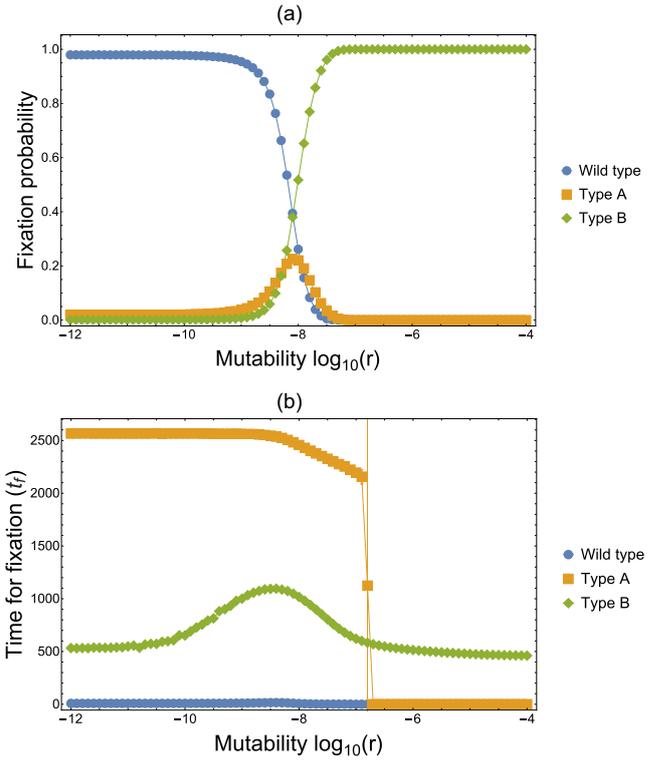


FIG. 3. Fixation probability of each type (a) and the fixation time (b) as a function of mutability. The abrupt fall of the fixation time of type *A* and its standard deviation is due to the finite sample size. Parameters are $N = 10^7$, $r = 1 \times 10^{-9}$, $s_A = .01$, $s_B = 0.056$, and $\rho = 100$.

given by

$$\pi_f^W = \frac{\rho}{\rho + 1} \frac{1 - \left(\frac{1}{\omega_A}\right)^{N-1}}{1 - \left(\frac{1}{\omega_A}\right)^N} + \frac{1}{\rho + 1} \frac{1 - \left(\frac{1}{\omega_B}\right)^{N-1}}{1 - \left(\frac{1}{\omega_B}\right)^N}.$$

Notice that if N is large, then $\pi_f^W \approx 1$, that is, most of the time the mutations are lost and the population returns to the homogenous wild state.

Figure 3 shows that in the regime of low mutation rates, although type *A* fixes more often than *B*, type *A* takes more time to fix. Interestingly, the fixation time for the mutant *A* is reduced as mutability is increased, although competition from *B* is also enhanced. The fixation time of mutant *B* shows a peak at mutation rate close to the value at which the probability of fixation of mutant *A* is higher than other variants. The fixation time increases because there are more individuals of type *A* to compete with. As mutability is further increased, the number of both mutants increase, and fluctuations no longer favor a greater competitiveness of type *A*, so that the fixation time of *B* decreases.

To have a better picture of the mechanisms of fixation of each type, we analyze a population composed of only the wild type and N_e initial mutants of the same type with effective selection coefficient s_e , without any further mutation in the evolutionary process. Then, we can ask what are the effective values N_e and s_e that yield the same fixation probability and fixation times as the full model. To calculate these effective values, we need two equations. The first is Eq. (4); for the

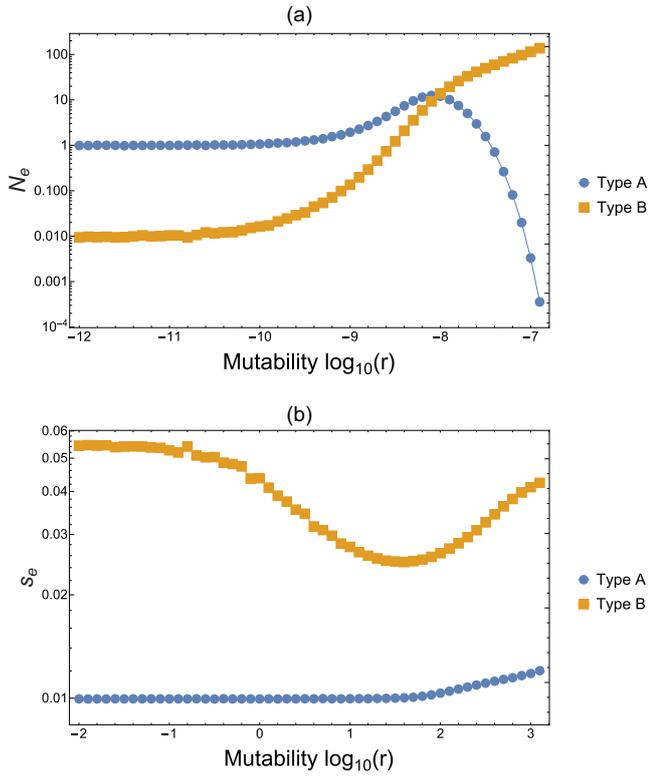


FIG. 4. Effective size and selection coefficients. The two graphs, (a) and (b), must be read in pairs: For each r , the values of N_e and s_e correspond to the solution of a system composed with the wild type and N_e mutants of a single type that yields the same fixation probability and fixation time as the full model. The original data come from simulation. Parameters are $N = 10^7$, $s_A = 0.01$, $s_B = 0.056$, and $\rho = 100$.

second we can use the fixation time in the diffusive limit [37] for our conditions (high $\alpha = 2Ns$), given by,

$$\tau = \int_q^1 \frac{[1 - e^{-2(1-q)Ns}](1 - e^{-2qNs})}{(1-q)qs(1 - e^{-2Ns})} dq, \quad (6)$$

where q is the initial frequency of the mutant type. For our $\alpha = 10^5$ the integrand between 0 and p is essentially null. The initial number of mutants is then Nq . The numerical results shown in Fig. 3 are used as input to Eqs. (4) and (6). The resulting system of equations is solved using the integration algorithm QAGS (quadrature routine, adaptive integrand, general function, singularities can be more readily integrated) and the `gls_multiroot_fsolver` algorithm, both from the GNU Scientific Library.

Figure 4 shows the effective subpopulation size of mutants and the effective selection coefficient for both types of mutants. The greater the effective number of mutants, the more readily fixation occurs in the equivalent population. At small mutabilities, the initial effective number of mutants is close to the probability of having one mutation in the first step of the full model, $\rho/(\rho + 1)$ and $1/(\rho + 1)$ for types A and B, respectively. This result corroborates the regime of “one mutant at a time.” As the mutability value increases, the same qualitative behavior found in Fig. 3 is observed, with a peak of N_e near the r value associated with the peak in the fixation

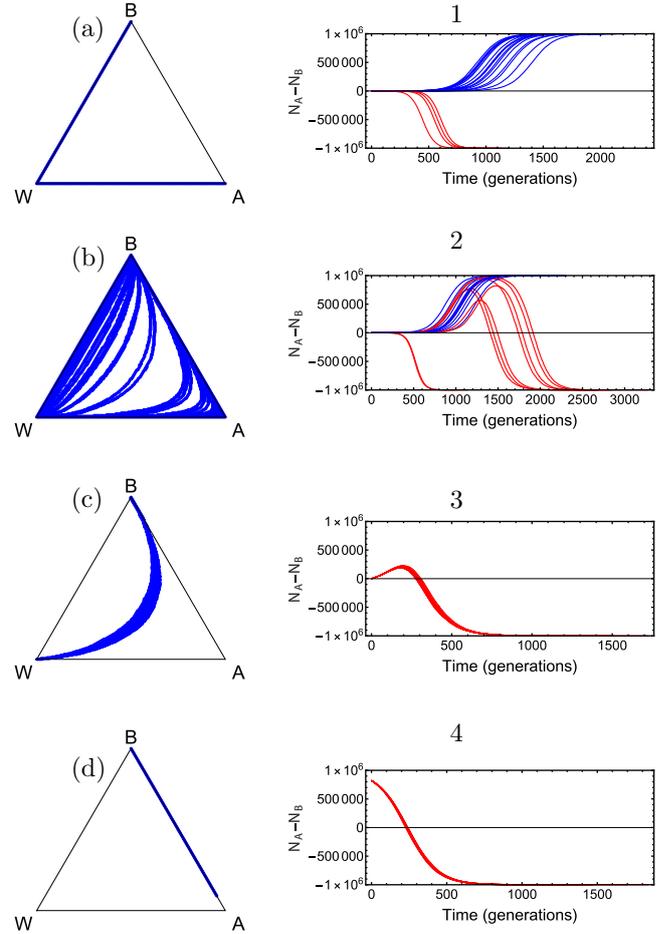


FIG. 5. Typical simulated histories of population composition, ending in absorption by A or B. The left panels [(a)–(d)] show the evolution on the simplex $N_W + N_A + N_B = N$, with the vertices corresponding to single-species states and edges to two-species states. The right panels (1–4) show the difference $N_A - N_B$, (regardless the value of N_W) versus time. Histories ending in absorption by A (B) are shown in blue (red). Rows correspond to mutabilities (from top) $r = 10^{-11}$, $r = 10^{-7}$, $r = 10^{-3}$, and $r = .91$. Parameters are $N = 10^6$, $s_A = .01$, $s_B = 0.02$, and $\rho = 10$.

probability. In contrast to the large variation in the effective size, the effective selection is roughly constant for type A and has small variations for type B. In the diffusive limit, the expected change of the number of mutants in one step is proportional to the selection coefficient and the variance of that change is inversely proportional to the size. Thus, for type B, the variance decreases steadily with the increase of mutability, while its mean step has a minimum close to where competition has a maximum. So, at this mutability, type B varies in small steps while the steps of type A vary in a less-regular manner. This may be the reason for the relatively higher competitiveness type B faces in the vicinity of this mutability value.

The temporal evolution of the abundances of types A and B, shown in Fig. 5, exhibits several interesting features. At low mutation rates, there is no A-B coexistence at any point in the evolution, but as we leave the rare-mutation

regime, coexistence does occur. Interestingly, the evolution in cases featuring fixation of type *A* is always monotonic, whereas those in which type *B* fixes can be nonmonotonic. The reason for nonmonotonicity in the latter case is simple: Since type *B* is fitter, it can recover after a delayed emergence. If the mutation rate is high, then type *B* easily dominates.

Finally, the evolution predicted by the approximation used in the previous subsection, where it is supposed that all mutations take place at once, is shown in the lowest row of Fig. 5. The trajectories occur mostly at the segment *AB* (except for the vertex all-*W*) beginning around $N_A = N\rho/(1 + \rho)$ and $N_B = N - N_A$, depending on the parameters, sometimes shifted toward vertex all-*A* or all-*B*. For such parameters, the fate of fixation is almost exclusively to *B*.

C. Rare mutations

In the regime in which mutations occur more slowly than selection, fixation takes place before the next mutation appears. This is the “strong selection-weak mutation” (SSWM) regime, as defined by Gillespie (1983) [38]. In this regime, the analysis is greatly simplified and provides analytical support for the simulations in the limit of low mutation supply.

To determine the absorption probabilities we note that, in the SSWM regime, only one mutant coexists with the wild type. Thus, the next equation furnishes the fixation probabilities of both *A* and *B*, that is,

$$\pi_f^i = \frac{1 - \omega_i}{1 - \omega_i^N}. \tag{7}$$

The absorption probability of type *i* is then given by

$$\pi_a^i = \frac{r_i \pi_f^i}{r_A \pi_f^A + r_B \pi_f^B}, \tag{8}$$

which, after substitution, becomes

$$\pi_a^A = \frac{1}{1 + \frac{1 - \omega_B}{\rho} \frac{1 - \omega_A^N}{1 - \omega_A}}, \tag{9}$$

where $\omega_A = e^{-2s_A}$ and $\omega_B = e^{-2s_B}$.

To check this formula, the case $\omega_A = \omega_B$ is helpful. In this case there is no competitive advantage between the types and the absorption probability should depend only on the relative mutation rates. Indeed, Eq. (9) simplifies to

$$\pi_a^A = \frac{\rho}{1 + \rho}. \tag{10}$$

In our initial problem, type *A* is less fit than type *B*, but arises through mutation more often than type *B*, so that fixation of *A* remains possible. To quantify the prevalence of *A* over *B*, we determine the condition to have $\pi_a^A > 1/2$. Rearranging the terms in Eq. (9), we find

$$\rho > \frac{1 - \omega_B}{1 - \omega_B^N} \frac{1 - \omega_A^N}{1 - \omega_A}. \tag{11}$$

Since we are analyzing adaptive mutations, we have that $0 < \omega_B < \omega_A < 1$. In the regime $N \rightarrow \infty$, the condition sim-

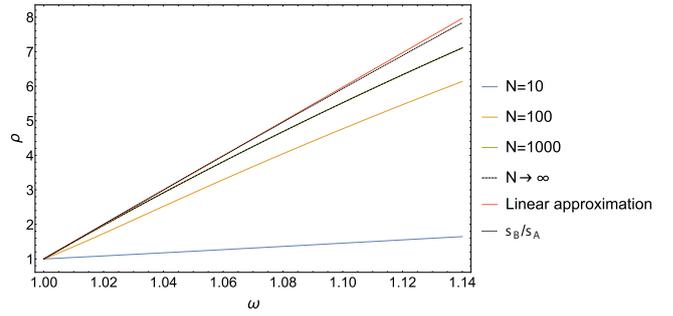


FIG. 6. Threshold for ρ above which type *A* prevails in the regime of rare mutations from Eq. (11) for $N = 10$ (blue), $N = 100$ (yellow), and $N = 1000$ (green). The limit $N \rightarrow \infty$, Eq. (12) (black, dashed) and the approximation of Eq. (13) (red) are also shown. Notice that the $N \rightarrow \infty$ approximation fits the $N = 1000$ data well. The greater the advantage of *B* over *A* or the larger the population size, the larger the ratio ρ required for *A* to prevail.

plifies to

$$\rho \gtrsim \frac{1 - \omega_B}{1 - \omega_A}. \tag{12}$$

Because $\omega_B < \omega_A$, it follows that ρ must be always greater than 1. For weak selection, $e^{s_A} \approx w_A = 1 + s_A$ and $w_B = e^{s_B} \approx 1 + s_B$, and Eq. (12) can be further simplified to

$$\rho \gtrsim s_B/s_A. \tag{13}$$

The condition in Eq. (11), or the simplification in Eq. (13), shows that, if type *A* is to have an evolutionary advantage over the fitter type, then the mutation rate to type *A* should be higher than that to type *B* at least as much as determined by Eq. (11). Notice that, the greater the fitness advantage of *B*, the greater should be ρ , as shown in Fig. 6.

D. Frequent mutations

In the regime of frequent mutation, all mutations take place nearly at once. Thus, the population is composed of types *A* and *B* only and the equivalent of Eq. (4) (replacing w_W , which is equal to 1, by w_B) is

$$\pi_f^A = \frac{1 - \omega^{N_A}}{1 - \omega^N},$$

where $\omega = \omega_A/\omega_B$. The probability of having N_A mutants in the first generation is given by the binomial distribution $B(N, \frac{\rho}{1+\rho})$. Thus, the probability of absorption of *A* is

$$\begin{aligned} \Pi_a^A &= \sum_{N_A=1}^N B\left(N, \frac{\rho}{1+\rho}\right) \frac{1 - \omega^{N_A}}{1 - \omega^N} \\ &= \frac{1 - \left(\frac{\rho \omega + 1}{\rho + 1}\right)^N}{1 - \omega^N}. \end{aligned} \tag{14}$$

This equation also simplifies to Eq. (10) if $s_A = s_B$, which is again a consistent result.

The condition for $\pi_a^A > 1/2$ is

$$\rho > \frac{T - 1}{\omega - T}, \tag{15}$$

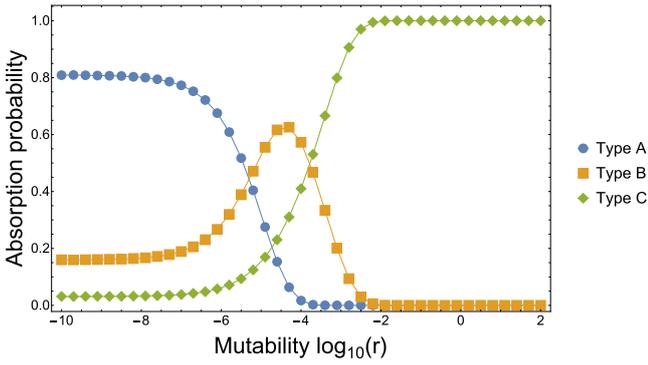


FIG. 7. System with three adaptive mutants, with increasing fitness and decreasing mutation rate. Parameters are $N = 10^6$, $s_A = .01$, $s_B = 0.02$, $s_C = 0.04$, and $\rho = 10$ from A to B and from B to C .

where

$$T = \left[\frac{1}{2}(\omega^N + 1) \right]^{1/N} \\ = \omega \left(1 - \frac{\log 2}{N} + \frac{1}{N\omega^N} - \frac{\log 2}{N^2\omega^N} + \dots \right).$$

For $N \gg 1$, because $\omega > 1$, T is well approximated by the leading terms,

$$T \approx \omega \left(1 - \frac{\log 2}{N} \right),$$

and Eq. (15) becomes

$$\rho > N \frac{(\omega - 1)}{\omega \log 2}. \quad (16)$$

Thus, the greater the size or the greater the selective advantage of B over A ($\omega = e^{2(s_B - s_A)}$), the harder it is for type A to prevail. It is still possible that type A prevail, but the mutation rate to this type must be very high.

E. A further result

The results obtained in our model with two adaptive mutants are also observed in systems that have three adaptive mutants, as shown in Fig. 7. In this last model all three mutants are generated by mutation from the W type. The selection coefficients of types A , B , and C are $s_A = 0.01$, $s_B = 0.02$, and $s_C = 0.04$. Mutation rates are $r_A = r$, $r_B = r_A/\rho$, and $r_C = r_B/\rho = r_A/\rho^2$, with $\rho = 10$. Population size is $N = 10^6$. Sampling of new individuals is performed according to

$$p_W^{t+1} \propto N_W^t (1 - r_A - r_B - r_C), \\ p_A^{t+1} \propto w_A N_A^t + N_W^t r_A, \\ p_B^{t+1} \propto w_B N_B^t + N_W^t r_B, \\ p_C^{t+1} \propto w_C N_C^t + N_W^t r_C,$$

The overall qualitative behavior persists: As we increase the mutability, mutants with decreasing mutation rates and increasing fitness prevail.

For small r , the most frequent mutant is absorbed more often because the effectiveness of fixation itself, given that at least one mutant already exists, is similar among the types.

So the mutation rate alone decides which type absorbs. As we increase the mutation rate, we favor the coexistence of types A and B while the mutation rate is too low for type C to appear. Because type B is fitter than A , type B wins the competition in this regime. Finally, with higher mutabilities, type C coexists with the other types and wins the competition since it is the fittest.

IV. CONCLUSIONS

The traditional understanding since the modern synthesis of the role of mutations is that in the most typical cases, the majority of the variability of a population, after some rounds of mutation, is already available and the role of mutations is merely to shift the frequency of preexisting variants [2–4]. However, Yampolsky and Stoltzfus (2001) [6] showed that, under competition between mutation rate and selection, mutation bias can change the prevailing allele. We present a thorough examination of the conditions for which bias in mutation changes the prevailing type.

Our analysis employs a Wright-Fisher model starting with a wildtype population which can mutate one-way to two other mutants, A and B , with B fitter than A and the mutation rate to A higher than that to B . We verify that a sufficiently strong mutation bias can change the type that would prevail if selection were the only mechanism in operation. We observe this under varying mutability, drift (created by fluctuation in small populations), selective advantage and gene flow parameters.

Our results imply that the conditions under which the fate of evolution changes are much broader than commonly acknowledged. A focus on the effect of mutability shows that distinct kinetics are possible, but that there are two basic regimes: one with rare mutations and the other with frequent mutations. In the rare regime, the less-fit type can prevail, whereas in the frequent mutation regime this is nearly impossible.

In addition, we show that increasing mutability may determine the order which mutant among three. In addition, we show that increasing mutability may determine the order which mutant among three is more likely to prevail.

We expect that this work will underpin explanations for the dynamics in more complex fitness landscapes, revealing why one path prevails over others. This leads to questions related to the predictability and repeatability of evolutionary processes [39–44]. In future work we plan to investigate the fate of evolution in simple, completely smooth fitness landscapes while varying population size, fitness, and mutability. This should yield further insights to understanding the kinetics and evolutionary fate in realistic and experimental landscapes [45–51].

ACKNOWLEDGMENTS

A.A.S. was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); L.W. was supported by CNPq; L.B.K. was supported by CNPq, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); R.D. acknowledges support from CNPq, Brazil, through Project No. 303766/2016-6.

- [1] B. Shorrocks, *A Origem da Diversidade* (Edusp, São Paulo, 1978).
- [2] J. Haldane, The part played by recurrent mutation in evolution, *Am. Nat.* **67**, 5 (1933).
- [3] J. Maynard-Smith, R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert, Developmental constraints and evolution: A perspective from the mountain lake conference on development and evolution, *Q. Rev. Biol.* **60**, 265 (1985).
- [4] J. Bull, M. Badgett, H. Wichman, J. Huelsenbeck, D. Hillis, A. Gulati, C. Ho, and I. Molineux, Exceptional convergent evolution in a virus, *Genetics* **147**, 1497 (1997).
- [5] D. Futuyma, *Biologia Evolutiva* (SBG: CNPq, 1992).
- [6] L. Yampolsky and A. Stoltzfus, Bias in the introduction of variation as an orienting factor in evolution, *Evol. Dev.* **3**, 73 (2001).
- [7] A. Stoltzfus and D. M. McCandlish, Mutational biases influence parallel adaptation, *Mol. Biol. Evol.* **34**, 2163 (2017).
- [8] A. Stoltzfus, Mutation-biased adaptation in a protein nk model, *Mol. Biol. Evol.* **23**, 1852 (2006).
- [9] J. F. Storz, C. Natarajan, A. V. Signore, C. C. Witt, D. M. McCandlish, and A. Stoltzfus, The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function, *Philos. Trans. R. Soc. London B* **374**, 20180238 (2019).
- [10] K. Jain, J. Krug, and S.-C. Park, Evolutionary advantage of small populations on complex fitness landscapes, *Evolution* **65**, 1945 (2011).
- [11] A. Stoltzfus and R. W. Norris, On the causes of evolutionary transition: Transversion bias, *Mol. Biol. Evol.* **33**, 595 (2016).
- [12] P. Higgs, Frequency distributions in population genetics parallel those in statistical physics, *Phys. Rev. E* **51**, 95 (1995).
- [13] L. Peliti, Introduction to the statistical theory of darwinian evolution (1997), [arXiv:cond-mat/9712027](https://arxiv.org/abs/cond-mat/9712027).
- [14] B. Drossel, Biological evolution and statistical physics, *Adv. Phys.* **50**, 209 (2001).
- [15] N. H. Barton and J. B. Coe, On the application of statistical physics to evolutionary biology, *J. Theor. Biol.* **259**, 317 (2009).
- [16] H. P. de Vladar and N. H. Barton, The contribution of statistical physics to evolutionary biology, *Trends Ecol. Evol.* **26**, 424 (2011).
- [17] H. Flyvbjerg and B. Lautrup, Evolution in a rugged fitness landscape, *Phys. Rev. A* **46**, 6714 (1992).
- [18] L. S. Tsimring, H. Levine, and D. A. Kessler, Rna Virus Evolution Via a Fitness-Space Model, *Phys. Rev. Lett.* **76**, 4440 (1996).
- [19] E. Baake, M. Baake, and H. Wagner, Ising Quantum Chain is Equivalent to a Model of Biological Evolution, *Phys. Rev. Lett.* **78**, 559 (1997).
- [20] S. Galluccio, Exact solution of the quasispecies model in a sharply peaked fitness landscape, *Phys. Rev. E* **56**, 4526 (1997).
- [21] M. Nilsson and N. Snoad, Error Thresholds for Quasispecies on Dynamic Fitness Landscapes, *Phys. Rev. Lett.* **84**, 191 (2000).
- [22] D. B. Saakian, E. Muñoz, C.-K. Hu, and M. W. Deem, Quasispecies theory for multiple-peak fitness landscapes, *Phys. Rev. E* **73**, 041913 (2006).
- [23] L. Tan, S. Serene, H. X. Chao, and J. Gore, Hidden Randomness Between Fitness Landscapes Limits Reverse Evolution, *Phys. Rev. Lett.* **106**, 198102 (2011).
- [24] T. Brotto, G. Bunin, and J. Kurchan, A model with darwinian dynamics on a rugged landscape, *J. Stat. Phys.* **166**, 1065 (2017).
- [25] Alfred Ajay Aureate R. and Vaibhav Madhok, Typicality in quasispecies evolution in high dimensions, *Phys. Rev. E* **100**, 042407 (2019).
- [26] D. B. Saakian, T. Yakushkina, and E. V. Koonin, Allele fixation probability in a Moran model with fluctuating fitness landscapes, *Phys. Rev. E* **99**, 022407 (2019).
- [27] B. Ottino-Löffler and M. Kardar, Population extinction on a random fitness seascape, *Phys. Rev. E* **102**, 052106 (2020).
- [28] P. Gerrish and R. Lenski, The fate of competing beneficial mutations in an asexual population, *Genetica* **102-3**, 127 (1998).
- [29] R. Miralles, P. Gerrish, A. Moya, and S. Elena, Clonal interference and the evolution of rna viruses, *Science* **285**, 1745 (1999).
- [30] W. Ewens, *Mathematical Population Genetics I: Theoretical Introduction*, Interdisciplinary Applied Mathematics (Springer, New York, 2004).
- [31] T. Ohta, Slightly deleterious mutant substitutions in evolution, *Nature (London)* **246**, 96 (1973).
- [32] T. Ohta, The nearly neutral theory of molecular evolution, *Annu. Rev. Ecol. Evol. Syst.* **23**, 263 (1992).
- [33] T. Ohta and J. H. Gillespie, Development of neutral and nearly neutral theories, *Theor. Pop. Biol.* **49**, 128 (1996).
- [34] M. Galassi *et al.*, GNU scientific library reference manual.
- [35] J. Gillespie, *Population Genetics: A Concise Guide*, Population Genetics (Johns Hopkins University Press, Baltimore, MD, USA, 2004).
- [36] N. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library (Elsevier Science, Amsterdam, The Netherlands, 2011).
- [37] B. Charlesworth, How long does it take to fix a favorable mutation, and why should we care?, *Am. Nat.* **195**, 753 (2020).
- [38] J. H. Gillespie, Some properties of finite populations experiencing strong selection and weak mutation, *Am. Nat.* **121**, 691 (1983).
- [39] J. Meyer, D. Dobias, J. Weitz, J. Barrick, R. Quick, and R. Lenski, Repeatability and contingency in the evolution of a key innovation in phage lambda, *Science (New York, N.Y.)* **335**, 428 (2012).
- [40] M. K. Burke, G. Liti, and A. D. Long, Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *saccharomyces cerevisiae*, *Mol. Biol. Evol.* **31**, 3228 (2014).
- [41] J. Lachapelle, J. Reid, and N. Colegrave, Repeatability of adaptation in experimental populations of different sizes, *Proc. R. Soc. B* **282**, 20143033 (2015).
- [42] B. Bauer and C. S. Gokhale, Repeatability of evolution on epistatic landscapes, *Sci. Rep.* **5**, 9607 (2015).
- [43] T. van Dijk, S. Hwang, J. Krug, J. A. de Visser, and M. Zwart, Mutation supply and the repeatability of selection for antibiotic resistance, *Phys. Biol.* **14**, 055005 (2017).
- [44] D. Nichol, J. Rutter, C. Bryant, A. Hujer, S. Lek, M. Adams, P. Jeavons, A. Anderson, R. Bonomo, and J. Scott, Antibiotic collateral sensitivity is contingent on the repeatability of evolution, *Nat. Commun.* **10**, 334 (2019).
- [45] M. Lunzer, S. Milner, R. Felsheim, and A. Dean, The biochemical architecture of an ancient adaptive landscape, *Science* **310**, 499 (2005).

- [46] C. P. Goulart, M. Mahmudi, K. A. Crona, S. D. Jacobs, M. Kallmann, B. G. Hall, D. C. Greene, and M. Barlow, Designing antibiotic cycling strategies by determining and understanding local adaptive landscapes, *PLoS One* **8**, e56040 (2013).
- [47] P. M. Mira, J. C. Meza, A. Nandipati, and M. Barlow, Adaptive landscapes of resistance genes change as antibiotic concentrations change, *Mol. Biol. Evol.* **32**, 2707 (2015).
- [48] A. C. Palmer, E. Toprak, M. Baym, S. Kim, A. Veres, S. Bershtein, and R. Kishony, Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes, *Nat. Commun.* **6**, 7385 (2015).
- [49] J. V. Rodrigues, S. Bershtein, A. Li, E. R. Lozovsky, D. L. Hartl, and E. I. Shakhnovich, Biophysical principles predict fitness landscapes of drug resistance, *Proc. Natl. Acad. Sci. USA* **113**, E1470 (2016).
- [50] Y. T. Tamer, I. K. Gaszek, H. Abdizadeh, T. A. Batur, K. A. Reynolds, A. R. Atilgan, C. Atilgan, and E. Toprak, High-order epistasis in catalytic power of dihydrofolate reductase gives rise to a rugged fitness landscape in the presence of trimethoprim selection, *Mol. Biol. Evol.* **36**, 1533 (2019).
- [51] M. Knopp and D. Andersson, Predictable phenotypes of antibiotic resistance mutations, *mBio* **9**, e00770 (2018).