

01/

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
DEPARTAMENTO DE FÍSICA E INFORMÁTICA**

“DINÂMICA E GENEALOGIA DE MODELOS DE EVOLUÇÃO”

MILTON TAIDI SONODA

Dissertação apresentada ao Instituto de Física de São Carlos, Universidade de São Paulo, para obtenção do título de Mestre em Ciências “Física Básica”

**Orientador:
Prof. Dr. José Fernando Fontanari**

**Comissão Julgadora:
Prof. Dr. José Fernando Fontanari (IFSC-USP)
Prof. Dr. Otávio Henrique Thiemann (IFSC-USP)
Profa. Dra. Tânia Tomé Martins de Castro (IF-USP)**

01/

USP/IFSC/SBI



8-2-001373

**São Carlos - São Paulo
2001**

IFSC-USP SERVIÇO DE BIBLIOTECA
INFORMAÇÃO

Sonoda, Milton Taidi

“Dinâmica e genealogia de modelos de evolução”/ Milton Taidi Sonoda – São Carlos, 2001

Dissertação (Mestrado) – Instituto de Física de São Carlos da Universidade de São Paulo, 2001
páginas: 65

Área: Física Básica

Prof. Dr. José Fernando Fontanari

1. Simulações numéricas, 2. Genealogia

I. Título



MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE Mestrado DE MILTON
TAIDI SONODA APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, DA
UNIVERSIDADE DE SÃO PAULO, EM 21 DE FEVEREIRO DE 2001

COMISSÃO JULGADORA:

José F. Fontanari

Prof. Dr. José Fernando Fontanari/IFSC-FFI-USP

Otávio Henrique Thiemann

Prof. Dr. Otávio Henrique Thiemann/IFSC-FFI-USP

Tânia Tomé Martins de Castro

Profa. Dra. Tânia Tomé Martins de Castro/IF-USP

*Aos meus pais, Taizo e Haruco,
e irmãos, Moriharu e Thaís.*

*Este trabalho teve o apoio
financeiro da CAPES.*

Agradecimentos

Gostaria de agradecer à algumas pessoas que estiveram ao meu lado durante a realização desse trabalho.

Gradeço primeiramente ao professor Fontanari pela sua dedicação na orientação desse projeto. Gradeço também pela sua enorme contribuição na minha formação profissional.

Ao pessoal do grupo: Viviane, Fernandos (Ferreira e Crepaldi), Ana Tereza, Cláudia e Domingos pelo companheirismo. Um agradecimento especial ao Paulo, com quem eu sempre pude contar para discussões e questionamentos sobre o trabalho, e ao Alexandre Colato, pela leitura cuidadosa dessa dissertação.

Aos amigos Rômulo (Rão), Alexandre (B1), Armando (Arn), Rafael (Faer), David e Oliver (Gooliver). Levarei comigo várias lembranças de companheirismo e amizade verdadeira compartilhadas com cada um deles.

Por último, quero agradecer a Deus.

Resumo

Nesse trabalho investigamos através de simulações numéricas a evolução da composição genética de uma população, dando atenção especial ao processo dinâmico conhecido como catraca de Müller, que é responsável pela degradação da população devido ao acúmulo de mutações deletérias em populações finitas. Consideramos também a genealogia dos indivíduos em uma população sob a ação da catraca de Müller. Ainda, investigamos analiticamente o limite determinístico do modelo, no qual o tamanho da população é infinito, onde o processo da catraca não atua. O relevo replicativo, ou seja, a função que mapeia a carga genética de um indivíduo com a sua probabilidade de reprodução utilizado nesse trabalho é uma generalização do relevo originalmente proposto por Müller para ilustrar o processo da catraca. Adicionamos a esse relevo um parâmetro de epistase que simula a interação entre os sítios das seqüências dos indivíduos. A escolha desse parâmetro determina três tipos possíveis de epistase: (i) sinérgica, no qual as mutações ficam cada vez mais deletérias com o número de mutações já existentes; (ii) atenuante, no qual o efeito deletério de uma nova mutação é atenuado; e (iii) multiplicativa, no qual as novas mutações causam danos idênticos, independentemente do número anterior de mutações.

Abstract

In this work we investigate through numerical simulations the evolution of the genetic composition of a population, giving emphasis to the dynamic process termed Müller's ratchet, which is responsible for the degradation of the population due to the accumulation of deleterious mutations in finite populations. We consider also the genealogy of the individuals evolving in a population under the effect of the Müller's ratchet. In addition, we investigate analytically the deterministic limit of the model, in which the population size is infinite, where ratchet process does not act. The replication landscape, i.e., the function that maps the genetic load of an individual on its probability of reproduction used in this work is a generalization of that originally considered by Müller to illustrate the process of the ratchet. In particular, we add to that landscape a parameter of epistasis that models the interactions among the sites of the sequences of the individuals. The tuning of this parameter determines three different types of epistasis: (i) synergistic, where the mutations become more deleterious with the number of mutations already present; (ii) diminishing, where the deleterious effect of a new mutation is attenuated; and (iii) multiplicative, where the new mutations cause identical damages, independently of the previous number of mutations.

Conteúdo

1	Introdução	1
2	O Modelo	5
3	O limite determinístico: população infinita	12
3.1	Formulação matemática	13
3.2	Resultados numéricos	18
3.3	Limiar do erro de replicação	27
4	A catraca de Muller: população finita	30
4.1	O modelo de Wright-Fisher	30
4.2	A catraca de Müller	34
5	Genealogias	49
5.1	Alguns resultados analíticos simples	50
5.2	Alguns Resultados Numéricos	54
6	Conclusão	60
	Bibliografia	63

Lista de Figuras

- 3.1 Freqüência de indivíduos com k mutações para $s = 0.5$ e $U = 0.5$ e diversos valores do parâmetro de epistase α para o caso $C_0 > 0$ 18
- 3.2 Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.1$ 20
- 3.3 Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.3$ 21
- 3.4 Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.5$ 22
- 3.5 Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.7$ 23

- 3.6 Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.9$ 24
- 3.7 Frequência C_k dos indivíduos com k mutações contra o número de mutações k , onde cada curva colorida é o estado de equilíbrio para um valor de classe mais adaptada na população. Os parâmetros utilizados são $S = 0.5$ $U = 0.5$ $\alpha = 0.5$ (topo) ,1.0(meio) e 1.5(base). A curva tracejada é o gráfico da frequência do indivíduos mais adaptado para cada estado de equilíbrio. 26
- 4.1 Frequência C_A do alelo A contra geração t para simulações do modelo de Wright-Fisher, com tamanhos de população $N=6$ (topo), $N=20$ (meio) e $N=100$ (base). Cada cor diferente corresponde a evolução de uma população independente. 32
- 4.2 Frequência de populações F_{pop} contra estado de frequência C_A do alelo A para simulações com 3000 populações com 50 indivíduos. São mostrados gráficos para tempos de simulação a) $t=3$, b) $t=10$, c) $t=25$, d) $t=50$, e) $t=100$ e f) $t=200$. Inicialmente todas as populações possuem metade dos seus alelos no estado A. Com o passar do tempo a distribuição alarga-se e, para tempos muito longos, metade das populações se fixa em A e a outra metade em B. 33
- 4.3 Gráfico do número N_k de indivíduos com k mutações (número de indivíduos na classe k) contra o número de mutações para diferentes gerações t ; a) $t=2000$, b) $t=4000$, c) $t=6000$ e d) $t=8000$. A catraca aparece como uma onda propagando-se para a direita com velocidade e forma constantes. Os parâmetros utilizados foram $N = 300$, $s = 0.1$, $U = 0.3$ e $\alpha = 1.0$ 38

- 4.4 Gráficos de k_{med} , k_{fix} , k_{min} e k_{max} contra o tempo mostrando a acumulação de mutações na população. Os parâmetros utilizados foram $N = 100$, $s = 0.1$, $U = 0.5$ e $\alpha = 1.0$. O gráfico superior apresenta os resultados para uma única população, enquanto que o inferior apresenta resultados mediados sobre 50 populações independentes. 41
- 4.5 A Figura mostra uma genealogia possível para uma população haplóide sem recombinação. A linha pontilhada representa a geração atual, e as pontas das linhas da genealogia que terminam na linha pontilhada representam os oito indivíduos da geração atual. Seguindo a história passada de todos esses indivíduos podemos encontrar seus ancestrais comuns. Os indivíduos indicados por A , B e C são alguns dos ancestrais da população, sendo A a ancestral comum mais recente de toda a população. 42
- 4.6 Gráficos superiores: número médio de mutações $\overline{\langle k \rangle}$ contra a geração t . Gráficos inferiores: velocidade da catraca R contra a geração t . Gráficos à esquerda: valores de $0.1 \leq \alpha \leq 1$. Gráficos à direita: valores de $\alpha > 1.0$. O tamanho da população é $N = 30$, o parâmetro seletivo é $s = 0.2$ e a taxa de mutação é $U = 1.0$. Os dados representam médias sobre 150 populações. 43
- 4.7 Gráficos superiores: número médio de mutações $\overline{\langle k \rangle}$ mutações contra a geração t . Gráficos inferiores: velocidade da catraca R contra a geração t . Gráficos à esquerda: valores de $0.1 \leq \alpha \leq 1$. Gráficos à direita: valores de $\alpha > 1.0$. O tamanho da população é $N = 30$, o parâmetro seletivo é $s = 0.4$ e a taxa de mutação é $U = 1.0$. Os dados representam médias sobre 150 populações. 44

- 4.8 Gráficos superiores: número médio de mutações $\overline{\langle k \rangle}$ mutações contra a geração t . Gráficos inferiores: velocidade da catraca R contra a geração t . Gráficos à esquerda: valores de $0.1 \leq \alpha \leq 1$. Gráficos à direita: valores de $\alpha > 1.0$. O tamanho da população é $N = 30$, o parâmetro seletivo é $s = 0.6$ e a taxa de mutação é $U = 1.0$. Os dados representam médias sobre 150 populações. 45
- 4.9 Velocidade R da catraca contra parâmetro seletivo s para parâmetros de epistase α entre 0 e 0.1. Utilizamos taxa de mutação U igual a 0.2 no topo, 0.4 no meio e 0.6 na base. Os dados representam médias sobre 200 populações de 30 indivíduos cada. 47
- 4.10 Velocidade R da catraca contra a taxa de mutação U para parâmetros de epistase α entre 0 e 1.0. Os dados representam médias sobre 300 populações de 30 indivíduos cada e o parâmetro seletivo utilizado é $s = 0.5$ 48
- 5.1 Duas genealogias possíveis para uma amostra de 5 indivíduos escolhidos de uma população de tamanho $N \gg 5$, com tempos de coalescência médios dados pela equação (5.1). O número total de genealogias possíveis é dado pela equação (5.2) com $n = 5$ 51
- 5.2 Gráficos superiores: coeficiente de epistase $\alpha = 1.0$. Gráficos inferiores: $\alpha = 1.2$. Gráficos à esquerda: tempo médio de coalescência dividido pelo tamanho da população $\overline{\langle T \rangle}/N$ contra taxa de mutação U . Gráficos à direita: distância de Hamming média $\overline{\langle d \rangle}$ contra U . Foram simuladas 2000 populações de tamanho $N = 30$, para valores de coeficiente seletivo s variando de 0.0 à 0.9. O tempo de simulação t empregado foi da ordem de 10^4 56

- 5.3 Gráficos superiores: média de 2000 populações de tamanho $N = 10$. Gráficos inferiores: média de 2000 populações de tamanho $N = 50$. Gráficos à esquerda: tempo médio de coalescência dividido pelo tamanho da população $\langle T \rangle / N$ contra a taxa de mutação U . Gráficos à direita: distância de Hamming média $\langle d \rangle$ contra U . Em todos os casos foi utilizado $\alpha = 0.5$ e $s = 0.5$. Cada curva mostra o resultado para diferentes tempos de simulação indicados na legenda. 57
- 5.4 Gráfico da dispersão relativa da adaptação r_f contra taxa de mutação U . Cada curva representa um tempo de simulação diferente. Foram realizadas médias em 300 populações de tamanho 50, com coeficiente de epistase $\alpha = 0.5$ e de seleção $s = 0.5$ 59

Capítulo 1

Introdução

Nesta dissertação investigamos um modelo de evolução de população, procurando compreender como a composição genética da população varia de acordo com os diversos parâmetros que definem o modelo. A evolução de uma população ocorre, basicamente, devido à ação de duas forças: mutação e seleção. Há uma pequena sutileza na definição de mutação: não são consideradas como mutações as alterações genéticas em um indivíduo que não sejam transmissíveis aos seus descendentes. Em outras palavras, se um indivíduo nasce com uma alteração genética que o torna estéril, então essa alteração não pode ser considerada uma mutação, e o traço dessa alteração é definitivamente eliminado da população. A seleção é, meramente, a sobrevivência e reprodução mais eficiente de algumas variantes genéticas em função das condições ambientais em que eles vivem. Portanto, somente as características hereditárias são o alvo da seleção.

Assim, os processos evolutivos aos quais uma população está sujeita determinam sua composição genética. Além da seleção, eventos históricos como expansões demográficas, especiações, migração de parte da população ou catástrofes, que possam causar alterações bruscas na composição de uma população, deixam sua marca no material genético dos indivíduos que formam a população.

A partir de meados da década de 70, quando as técnicas de seqüenciamento de ácidos nucleicos foram inventadas tornou-se possível obter com exatidão a composição

genética de uma população, possibilitando assim o estudo da sua evolução ao nível molecular. Antes desse período, os estudos experimentais eram realizados utilizando a técnica da frequência de eletromorfos, na qual se mede a mobilidade dos genes em gel eletroforético. Essa técnica limita-se a determinar se os alelos¹ dos genes são iguais ou diferentes. Ao entrar na era de biologia molecular, o estudo da genética de populações tornou-se mais fácil, pois atualmente dispõe-se de meios para se reproduzir a evolução de uma população natural em laboratório, como, por exemplo, populações virais ou de bactérias, de modo que informações sobre a diversidade genética ou a evolução de uma característica particular dos indivíduos podem ser obtidas diretamente [1, 2].

A disponibilidade desses dados experimentais motiva a formulação de modelos, como o estudado nesta dissertação, tornando possível a confrontação entre resultados experimentais e teóricos. Com isso podemos determinar as variáveis responsáveis pelas mudanças genéticas na população, aumentando nosso conhecimento sobre o processo evolutivo. Esses modelos simplificados de evolução servem como ponto de partida para o estudo de modelos mais complexos, que envolvem uma dinâmica mais rica. Na formulação de um modelo de genética de populações é necessário especificar *a priori* o mapeamento entre o espaço de genótipos, no qual estão representadas todas as possíveis configurações genéticas de um indivíduo, e a função adaptação. Esse mapa é denominado de *relevo de replicação*. É necessário também especificar o tipo de amostragem ou procedimento que determina a passagem de uma geração para a seguinte. Esses fatores – mapeamento e amostragem – são determinantes para a adequação do modelo ao dado experimental observado.

Outra forma de se investigar a evolução de uma população é através do seqüenciamento de genes homólogos (que desempenham a mesma função) em uma amostragem

¹Vamos utilizar a nomenclatura usual de genética: qualquer segmento de DNA sob investigação é denominado *gene*, enquanto que uma seqüência particular de bases para o segmento é denominada de *alelo*.

de indivíduos da população. Comparando os genes de indivíduos diferentes, dois a dois, pode-se observar padrões de semelhança e dissimilaridades. No caso de uma população de indivíduos haplóides sem recombinação, sabemos que todos os indivíduos possuem um ancestral na geração imediatamente anterior, porém nem todos os indivíduos geram descendentes e, considerando que todas as populações reais são finitas, se seguirmos a descendência de qualquer par de indivíduos em direção ao passado, em algum momento iremos encontrar um ancestral comum. Assim, as diferenças observadas entre dois indivíduos em uma mesma geração devem-se às mutações que ocorreram desde a geração em que o ancestral comum existiu. Portanto, a diversidade genética da amostragem pode ser relacionada com a duração desse período de tempo em que ocorreu a divergência. Fazendo isso para todos os pares de indivíduos da amostragem, podemos descobrir quais são os pares mais próximos geneticamente e dispor esses dados na forma de uma árvore genealógica (no caso de indivíduos de uma mesma espécie), ou de uma árvore filogenética (no caso de indivíduos de espécies diferentes), e através dessas árvores, inferir a história evolutiva dos indivíduos. A construção de árvores genealógicas e filogenéticas foi muito beneficiada pelas técnicas de seqüenciamento de DNA. De fato, antes da revolução da biologia molecular a construção dessas árvores baseava-se unicamente nas características morfológicas e comportamentais das espécies, sendo que atualmente o fator mais importante (mas não o único) é a diferença genética entre as espécies.

Do ponto de vista teórico, muita atenção foi dada a processos genealógicos em ambientes neutros (não seletivos) nas últimas duas décadas, culminando na *Teoria de Coalescência* de Kingman [3] (para revisão [4, 5]), porém pouco se sabe sobre esses processos na presença de forças seletivas. Testes estatísticos têm sido desenvolvidos para determinar quais os modelos que melhor explicam os dados observados, e também quais são os melhores valores dos parâmetros de um determinado modelo frente aos mesmos dados (para revisão [6]).

Nesse contexto geral, o objetivo específico deste trabalho é o estudo de um modelo de evolução de uma população haplóide em um meio seletivo, onde o relevo de replicação é uma generalização daquele utilizado por Müller em seu trabalho de 1964 [7]. Em particular, introduzimos nesse relevo um parâmetro que simula a interação entre os genes dos indivíduos. Modelos de populações haplóides aplicados à evolução servem para entendermos, por exemplo, a evolução de populações virais e assim definir estratégias para combater as enfermidades causadas por elas [8]. Outra aplicação é o estudo da dinâmica de mutações deletérias envolvidas no cromossomo Y , que apenas os machos transportam, bem como no DNA-mitocondrial, herdado apenas pelas fêmeas [4, 9]. Além de poderem ser usadas no estudo da história evolutiva dos humanos, mutações recorrentes no cromossomo Y e no DNA-mitocondrial estão relacionadas com diversas doenças que afligem a humanidade.

O restante desta dissertação está organizado da seguinte forma. No capítulo 2 descrevemos o modelo, discutimos e justificamos algumas de suas simplificações. No capítulo 3 investigamos o regime determinístico, onde consideramos o tamanho da população infinito. O efeito da finitude do tamanho da população é estudado no capítulo 4. Em particular, investigamos o comportamento da catraca de Müller (perda progressiva dos indivíduos mais aptos da população via mutações deletérias) e seu efeito na adaptação total da população. As propriedades estatísticas das árvores genealógicas da população são discutidas no capítulo 5. Por fim, as principais conclusões são apresentadas no capítulo 6.

Capítulo 2

O Modelo

Nessa dissertação apresentamos o estudo de um modelo de evolução de populações compostas por indivíduos (ou organismos) haplóides que obedecem regras probabilísticas de reprodução e mutação. Nesse capítulo vamos introduzir o modelo considerado e as grandezas observáveis de interesse nesse tipo de sistema.

Os indivíduos que constituem a população são descritos por sua carga genética, ou seqüência específica do seu genoma, de modo que nesta dissertação usaremos indistintamente os termos indivíduo e seqüência para a unidade de seleção. A carga genética é transmitida pelos indivíduos de geração em geração através do processo de reprodução. O tamanho da população N é mantido constante, modelando assim uma capacidade limitada de indivíduos que podem ser suportados pelo ambiente e impondo uma competição efetiva entre eles.

Um dos requisitos essenciais para que ocorra evolução é a possibilidade de variação na capacidade de adaptação dos indivíduos. Nesse trabalho vamos supor que essa capacidade dependa única e exclusivamente da seqüência que descreve o genótipo do indivíduo, e as variações sejam devidas às mutações nessas seqüências que ocorrem durante o processo de reprodução. Embora estejamos cientes de que essa escolha não é completamente apropriada, usaremos o termo *adaptação* como tradução do termo inglês '*fitness*', cuja definição, aliás, ainda não é consensual [10] na literatura.

Nosso estudo restringe-se apenas às mutações pontuais, ou seja, alterações que ocorrem trocando-se o estado (nucleotídeo) de um sítio da seqüência. Lembremos que cada sítio pode assumir quatro estados: A (adenina), G (guanina), C (citosina) e U (uracil, no caso de RNA) ou T (timina, no caso de DNA). Vamos supor que na população inicial todos os indivíduos sejam iguais, ou seja, todos os indivíduos possuam os mesmos nucleotídeos em sítios homólogos (sítios com a mesma posição na seqüência), e também que os comprimentos das seqüências sejam infinitos, de forma que as mutações não ocorram no mesmo sítio duas vezes¹.

Dessas suposições decorre que há apenas duas possibilidades para um sítio homólogo na população: ou o nucleotídeo original é idêntico ao do ancestral da população inicial, ou é um eventual mutante. Assim, representaremos a população inicial atribuindo o valor 0 a todos os sítios de todos os indivíduos. Ocorrendo uma mutação em um dado sítio, trocamos o valor 0 por 1, de modo que o valor 1 em um sítio de uma seqüência indica simplesmente que aquele nucleotídeo é diferente de seu homólogo no ancestral da população inicial. O modelo apresentado é nada menos que o clássico modelo de infinitos sítios, o qual foi proposto por Motoo Kimura para estudar o efeito da deriva genética na evolução molecular [11]. Em uma população de tamanho finito, é possível que todos os indivíduos da população apresentem o estado mutante 1 em um dado sítio e, neste caso, dizemos que ocorreu uma fixação. Por outro lado, dizemos que um sítio é segregante se os dois estados, 0 e 1, co-existirem na população. É fundamental observar que uma fixação no, digamos, sítio i , não ocorre porque todos os sítios i de todas as seqüências da população sofreram mutação, já que a probabilidade de que duas seqüências tenham sofrido mutação no mesmo sítio é nula. A fixação ocorre simplesmente porque todos os indivíduos que não apresentavam mutação no sítio i foram extintos.

¹A probabilidade de que duas mutações ocorram no mesmo sítio tende a zero com o inverso do tamanho da seqüência.

A probabilidade \mathcal{M}_k de termos k novas mutações em uma seqüência é dada por uma distribuição de Poisson

$$\mathcal{M}_k = e^{-U} \frac{U^k}{k!}, \quad (2.1)$$

onde U é o número médio de mutações por seqüência. Essa distribuição decorre de supormos que o tamanho da seqüência é infinito. De fato, no caso de tamanho de seqüência finito L e probabilidade u de mutação por nucleotídeo, a probabilidade de k mutações obedece a uma distribuição binomial, que reduz-se a distribuição de Poisson no limite em que $L \rightarrow \infty$ e $u \rightarrow 0$ com $U = uL$ constante.

Em sistemas biológicos as mutações que ocorrem nos indivíduos podem ser benéficas ou prejudiciais, no entanto a proporção de mutações prejudiciais (ou deletérias) é muito maior do que a de mutações benéficas. Para simplificar o modelo, vamos supor que todas as mutações que ocorrem em um indivíduo sejam deletérias, diminuindo portanto as suas chances de reprodução. Daí, quanto menor o número de mutações em um indivíduo, maior será a sua adaptação. A função que relaciona o número de mutações k em um indivíduo com sua adaptação é a seguinte

$$f_k = (1 - s)^{k^\alpha}, \quad (2.2)$$

onde $s \in [0, 1)$ e $\alpha \in (0, \infty)$ são os parâmetros seletivo e de epistase, respectivamente. A grandeza relevante para a determinação do destino de um dado indivíduo, digamos indivíduo i , na população é a sua adaptação relativa W_i , definida como a razão entre a adaptação do indivíduo f_{k_i} e a adaptação total da população T_t na geração t , isto é,

$$W_i = \frac{1}{T_t} f_{k_i} \quad (2.3)$$

com

$$T_t = \sum_{j=1}^N (1 - s)^{k_j^\alpha}, \quad (2.4)$$

onde k_j representa o número de mutações no indivíduo $j = 1, 2, \dots, N$ da população. É importante observar que o valor de W_i depende da geração (tempo) pois T depende da

composição da população que, por sua vez, evolui no tempo. Da forma como são definidos os parâmetros α e s , observamos que a adaptação f_k é uma função monótona decrescente de k .

Os coeficientes seletivo s e de epistase α controlam o grau de dano que as mutações causam nos indivíduos. Por exemplo, à medida que s aproxima-se de seu valor máximo ($s = 1$), o efeito deletério das mutações torna-se cada vez maior, enquanto que para s pequeno, as mutações praticamente não afetam a adaptação do indivíduo. O caso limite $s = 0$ é denominado de regime de seleção neutra, no qual as mutações não produzem efeito algum na capacidade reprodutiva dos indivíduos, de modo que todos têm probabilidades iguais de se reproduzir. Já o limite $s \rightarrow 1$ leva à situação conhecida como regime de seleção extrema, no qual apenas os indivíduos com o menor número de mutações conseguem se reproduzir, sendo que as demais seqüências surgem apenas através das mutações dessa classe de indivíduos.

O coeficiente de epistase α modela a interação entre os sítios (nucleotídeos), causando diferentes decréscimos na adaptação de um indivíduo de acordo com o número de mutações já existentes. No caso $\alpha > 1$, a epistase é dita sinérgica, e o decréscimo na adaptação será maior quanto maior for o número de mutações prévias no indivíduo, enquanto que no caso $0 < \alpha < 1$, denominado epistase atenuante, este decréscimo é menor. O caso $\alpha = 1$ representa o regime não epistático, para o qual o efeito de uma nova mutação sobre a adaptação de um indivíduo não depende do número de mutações prévias. O efeito total das mutações é simplesmente o produto do efeito de cada mutação. Por essa razão, esse caso é também chamado de relevo de replicação multiplicativo.

Para ilustrar melhor o efeito do parâmetro α , consideremos dois indivíduos, 1 e 2, com k_1 e k_2 ($k_1 > k_2$) mutações, respectivamente, e definamos $\Delta \equiv k_1 - k_2$. A razão

entre as suas probabilidades de reprodução é escrita como

$$\frac{W_1}{W_2} = \frac{(1-s)^{k_1^\alpha}}{(1-s)^{k_2^\alpha}} = \exp \{[(k_2 + \Delta)^\alpha - k_2^\alpha] \ln(1-s)\}, \quad (2.5)$$

Vamos supor que a diferença Δ seja constante, de forma que W_1/W_2 dependa apenas do número de mutações de um dos indivíduos, por exemplo, k_2 . Para $\alpha = 1$, o valor de W_1/W_2 não depende do valor de k_2 , dependendo apenas da diferença Δ . Para $0 \leq \alpha \leq 1$, o valor do termo $(k_2 + \Delta)^\alpha - k_2^\alpha$ diminui à medida que k_2 aumenta, de modo que W_1/W_2 aproxime-se de 1. Isso indica que a superioridade de reprodução do indivíduo 2, comparado com a do indivíduo 1, diminui com o aumento do número de mutações e, no limite $k_2 \rightarrow \infty$ tal que $W_1/W_2 \rightarrow 1$, a população se encontra no regime de seleção neutra. Já para $\alpha \geq 1$, teremos o efeito contrário, ou seja, $W_1/W_2 \rightarrow 0$ quando $k_2 \rightarrow \infty$ de forma que a população entra no regime de seleção extrema.

Definidas a função adaptação dos indivíduos e a probabilidade de mutação, pasamos agora a discutir a dinâmica de uma população composta por N indivíduos. As gerações são discretas e a população na geração $t + 1$ é obtida da população na geração t através de N sorteios com reposição entre os indivíduos da geração t . Conforme discutido anteriormente, a probabilidade de que um indivíduo com k mutações seja escolhido em um desses sorteios é dada pela sua adaptação relativa W .

Esse procedimento de obter as populações a cada geração por amostragem e reposição é conhecido como modelo de Wright-Fisher na literatura de genética de populações [14], sendo também bastante utilizado na literatura de algoritmos genéticos [15]. Deve-se notar que a cada geração ocorre a renovação total da população, pois os pais dos indivíduos não convivem com os filhos, não havendo portanto, superposição de gerações. Após a escolha dos N indivíduos que passarão para a próxima geração, determina-se quantas mutações ocorrerão em cada indivíduo. Como já foi observado, se as seqüências têm tamanho infinito, duas mutações não podem ocorrer num mesmo sítio. Assim o número

de mutações de um indivíduo é igual a soma do número de mutações herdadas de seu pai com as novas mutações geradas pela distribuição de Poisson dada na equação (2.1).

Para concluir vamos a seguir discutir algumas limitações do modelo proposto e justificar certas simplificações. Primeiramente, a suposição de que as mutações são sempre deletérias (isto é, diminuem a adaptação dos indivíduos) é razoável, sendo corroborada por alguns resultados experimentais [2, 12, 13]. Porém um modelo mais realístico deveria considerar o efeito das mutações (deletéria ou benéfica) de uma forma estocástica. O estudo da função adaptação considerado neste trabalho decorre historicamente das pesquisas realizadas por Müller [7] e Crow [16]. Em particular Müller admitia a hipótese de que os efeitos das mutações fossem independentes ($\alpha = 1$), enquanto que Crow acreditava que as mutações interagissem sinergisticamente ($\alpha > 1$). Tais interações são relevantes para o estudo do processo chamado *Catraca de Müller* que ocorre em populações haplóides de tamanho finito [17]. Esse processo consiste na perda estocástica sucessiva e irreparável dos indivíduos com menor número de mutações, ou seja, do mais adaptados na população. Como a grande maioria das mutações são deletérias, a Catraca de Müller pode levar à degradação e eventual extinção da população. De fato, uma das teorias mais populares sobre a evolução do sexo baseia-se justamente no fato de que em uma população sexuada não ocorre acumulação de mutações deletérias e, portanto, a Catraca de Müller deixa de atuar [19, 20].

Um dos objetivos desse trabalho é estudar o efeito do parâmetro de epistase α em todo o intervalo de valores possíveis, para tamanho de população infinito (capítulo 3) e finito (capítulo 4). Deve-se notar que um limite extremode epistase atenuante, $\alpha \rightarrow 0$, corresponde ao relevo de replicação de um pico, freqüentemente associado à teoria de quase-espécies de evolução molecular de Eigen [22]. Outro aspecto importante da função adaptação adotada é a forma simples com que a epistase é modelada. Obviamente, a epistase entre os genes de um organismo tende a ser bem mais complexa e a interação

entre os sítios deve ser levada em conta de forma explícita como, por exemplo, no modelo *NK* de Kauffman [21].

Uma população real que satisfaz muitas das características do modelo são os vírus de RNA, que por serem organismos simples, possuindo entre 3 a 30 mil nucleotídeos em seu genoma, tornam o mapeamento entre o genótipo e a função adaptação mais simples que em organismos superiores. Outras características dos vírus de RNA são o tempo curto de geração e alta taxa de mutação, produzindo grandes populações com grande variabilidade dentro dessas populações. Por esses motivos os vírus de RNA constituem bons modelos experimentais para o estudo da evolução sob a perspectiva de genética de população, fornecendo grandezas observáveis que podem ser comparadas com os resultados de modelos teóricos [2].

Capítulo 3

O limite determinístico: população infinita

Considerando o modelo descrito no capítulo anterior, vamos primeiramente investigar a situação em que o tamanho da população é infinito. Nesse limite, a mutação, que gera novos tipos de indivíduos, e a seleção, que elimina os menos aptos, balanceiam-se de modo que a dinâmica atinge um estado estacionário, no qual a distribuição de probabilidade de um indivíduo sorteado aleatoriamente da população ter k mutações não depende do tempo. O objetivo da seção inicial desse capítulo é determinar essa distribuição para valores quaisquer do parâmetro seletivo s , da taxa de mutação por seqüência U e do parâmetro de espistase α . Iniciaremos com a apresentação de um tratamento analítico geral, porém, num certo estágio dos cálculos seremos obrigados a nos restringir ao caso $\alpha = 1$. Apresentaremos o procedimento numérico para obter os resultados para s , U e α quaisquer na seção 3.2. Na seção final estudaremos, ainda no limite determinístico, um caso particular importante do modelo, a saber, o limite $\alpha \rightarrow 0$ no qual apenas os indivíduos da classe 0 (isto é, que não sofreram nenhuma mutação) possuem vantagem seletiva, tendo as demais classes adaptações iguais e menor que a da classe 0. Como já mencionado, esse relevo replicativo é freqüentemente usado no modelo de quase-espécies de evolução molecular proposto por Eigen [22].

3.1 Formulação matemática

Vamos primeiramente separar os indivíduos em infinitas classes $k = 0, 1, 2, \dots$ de acordo com o número de mutações que possuam. Para isso definamos o vetor $\mathbf{X}(t) = (X_0(t), X_1(t), X_2(t), \dots)$, onde $X_k(t)$ representa o número de indivíduos na classe k na geração t . É conveniente considerarmos a população como sendo finita, de tamanho N , e no momento apropriado tomarmos o limite $N \rightarrow \infty$. Assim $\sum_{i=0}^{\infty} X_i(t) = N$ para todo t . Definamos ainda o vetor das freqüências das classes, $\mathbf{C}(t) = (C_0(t), C_1(t), \dots)$ com $C_k(t) = X_k(t)/N$ de modo que $\sum_{i=0}^{\infty} C_i(t) = 1$ para todo t .

Nosso problema consiste em determinar $\mathbf{X}(t+1)$ dado $\mathbf{X}(t)$. Lembremos que os indivíduos da geração $t+1$ são escolhidos em N sorteios entre os indivíduos da geração t . Seja $q_{m/k}$ a probabilidade de que um indivíduo da classe k seja escolhido em um sorteio, e seu descendente pertença à classe $m \geq k$, isto é, sofra $j = m - k$ mutações. Temos

$$q_{m/k} = \mathcal{M}_j \frac{X_k(t) f_k}{T_t} \quad (3.1)$$

onde $f_k = (1-s)^{k\alpha}$ e T_t é a adaptação total da população que, na nova notação, é escrita como

$$T_t = \sum_{i=0}^{\infty} X_i(t) f_i. \quad (3.2)$$

Daí, a probabilidade de se escolher em um sorteio um indivíduo de qualquer classe na geração t que será representante da classe m na geração $t+1$, denotada por $p_m(t)$, é simplesmente

$$\begin{aligned} p_m(t) &= \sum_{k=0}^m q_{m/k} \\ &= \sum_{k=0}^m \mathcal{M}_{m-k} \frac{X_k(t) f_k}{T_t}. \end{aligned} \quad (3.3)$$

Nesta equação podemos ver claramente que a classe m é populada pelos indivíduos da própria classe via reprodução exata, ou pelos indivíduos das classes com menor número de mutações e que, via reprodução errônea, mutam para a classe m .

Em N sorteios entre os indivíduos da geração t , a probabilidade de serem produzidos $X_m(t+1)$ indivíduos da classe m na geração $t+1$ é dada por uma binomial com parâmetros N e p_m ,

$$B[X_m(t+1)|\mathbf{X}(t)] = \binom{N}{X_m(t+1)} [p_m(t)]^{X_m(t+1)} [1 - p_m(t)]^{N - X_m(t+1)}. \quad (3.4)$$

Daí o valor esperado (média) do número de indivíduos na classe m na geração $t+1$ condicionado ao vetor $\mathbf{X}(t)$ é dado por

$$\begin{aligned} \langle X_m(t+1) \rangle_{\mathbf{X}} &= N p_m(t) \\ &= N \sum_{k=0}^m \mathcal{M}_{m-k} \frac{X_k(t) f_k}{\sum_{i=0}^{\infty} X_i(t) f_i} \end{aligned} \quad (3.5)$$

ou, em termos das frequências das classes,

$$\langle C_m(t+1) \rangle_{\mathbf{C}} = \sum_{k=0}^m \mathcal{M}_{m-k} \frac{C_k(t) f_k}{\sum_{i=0}^{\infty} C_i(t) f_i}. \quad (3.6)$$

É importante observar que essa equação é exata, sendo válida para qualquer valor de N . Para obtermos uma relação de recorrência devemos mostrar que o valor esperado no lado esquerdo de (3.6) pode ser igualado a $C_m(t+1)$. De fato, tomando o limite $N \rightarrow \infty$ na distribuição binomial (3.4) vemos que

$$X_m(t+1) = \langle X_m(t+1) \rangle_{\mathbf{X}} + \mathcal{O}(N^{1/2}) \quad (3.7)$$

ou

$$C_m(t+1) = \langle C_m(t+1) \rangle_{\mathbf{C}} + \mathcal{O}(N^{-1/2}). \quad (3.8)$$

Portanto, no limite determinístico $N \rightarrow \infty$ as frequências das classes obedecem a seguinte equação de recorrência

$$C_m(t+1) = \sum_{k=0}^m \mathcal{M}_{m-k} \frac{C_k(t) f_k}{\sum_{i=0}^{\infty} C_i(t) f_i} \quad (3.9)$$

para $m = 0, 1, \dots, \infty$.

Estamos interessados apenas no estado estacionário dessa recorrência, obtido fazendo-se $C_m(t+1) = C_m(t) = C_m, \forall m$. Nesse caso, reescrevemos (3.9) como

$$C_m = \frac{1}{T_e} \sum_{k=0}^m \mathcal{M}_{m-k} C_k f_k, \quad (3.10)$$

onde $T_e = \sum_{k=0}^{\infty} C_k f_k$ é a adaptação *média* da população no estado estacionário. Fazendo $m = 0$ em (3.10) obtemos

$$C_0 = \frac{1}{T_e} \mathcal{M}_0 C_0 f_0 = \frac{1}{T_e} e^{-U} C_0 \quad (3.11)$$

o que leva ao resultado surpreendente de que a adaptação média da população no equilíbrio é independente dos parâmetros s e α , ou seja,

$$T_e = e^{-U}, \quad (3.12)$$

desde que C_0 seja não nulo. Mais importante, já que a obtenção de T_e não envolve o conhecimento *a priori* de todas as frequências, como seria o caso se fôssemos calculá-lo através de sua definição, podemos obter uma solução explícita para C_m em termos das frequências das classes de maior adaptação

$$C_m = \frac{1}{1 - (1-s)^{m\alpha}} \sum_{k=0}^{m-1} C_k (1-s)^{k\alpha} \frac{U^{m-k}}{(m-k)!} \quad m = 1, 2, \dots \quad (3.13)$$

Substituindo seqüencialmente $m = 1, 2, \dots$ em (3.13), observamos que a frequência de qualquer classe é proporcional à frequência da classe zero C_0 unicamente. Podemos escrever $\mathbf{C} = C_0 \mathbf{C}'$, onde as componentes de $\mathbf{C}' = (1, C'_1, C'_2, \dots)$ também obedecem a equação (3.13) mas com $C'_0 = 1$. Essa equação permite-nos encontrar os valores de C'_1, C'_2, \dots . Finalmente, com o auxílio da condição

$$\sum_{k=0}^{\infty} C_k = 1 \quad (3.14)$$

obtemos C_0

$$C_0 = \frac{1}{\sum_{k=0}^{\infty} C'_k} \quad (3.15)$$

e, conseqüentemente C_1, C_2, \dots . Com esse procedimento podemos facilmente obter as freqüências de equilíbrio para qualquer conjunto dos parâmetros U, s e α . É interessante notar que esse procedimento é diferente daquele utilizado em geral na literatura que baseia-se na iteração direta da equação de recorrência (3.9) até que um ponto fixo seja alcançado [23, 24]. Observamos também que ambas as formas de se obter a distribuição de equilíbrio são gerais e exatas no limite determinístico, podendo ser aplicadas para qualquer escolha da função adaptação e distribuição de probabilidade do número de novas mutações. O procedimento descrito aqui, baseado no trabalho de Haigh [26], mostrou-se muito mais rápido e preciso, pois fornece as freqüências de equilíbrio diretamente.

Para $\alpha = 1$ (relevo multiplicativo) podemos encontrar uma solução analítica bastante simples para a equação (3.13). Substituindo seqüencialmente $m = 1, 2, \dots$ determinamos que

$$C_k = C_0 \frac{\theta^k}{k!} \quad (3.16)$$

onde $\theta = U/s$ e supomos que $C_0 \neq 0$. Usando a condição de normalização (3.14) obtemos $C_0 = e^{-\theta}$ e portanto

$$C_k = e^{-\theta} \frac{\theta^k}{k!}, \quad (3.17)$$

ou seja, para o relevo multiplicativo a probabilidade de que um indivíduo escolhido aleatoriamente da população tenha k mutações é dada por uma distribuição de Poisson de média (e variância) θ . Com esta distribuição calculamos facilmente a média \bar{f} e a variância $\sigma^2(f)$ da adaptação da população:

$$\bar{f} = \sum_{k=0}^{\infty} f_k C_k = e^{-U} \quad (3.18)$$

$$\bar{f}^2 = \sum_{k=0}^{\infty} f_k^2 C_k = e^{-2U+Us} \quad (3.19)$$

$$\sigma^2(f) = \bar{f}^2 - \bar{f}^2 = e^{-2U} (e^{Us} - 1). \quad (3.20)$$

Naturalmente, o resultado (3.18) já era conhecido e, conforme discutido anteriormente,

permanece válido para $\alpha \neq 1$ também. Como esperado, a variância da adaptação da população anula-se para $s = 0$ já que neste caso todos os indivíduos possuem a mesma adaptação $f_k = 1, \forall k$.

Uma grandeza particularmente importante para se determinar a intensidade do efeito da seleção é a dispersão relativa da adaptação r_f , definida como

$$\begin{aligned} r_f &= \frac{\sqrt{\sigma^2(f)}}{\bar{f}} \\ &= e^U \sqrt{\sigma^2(f)}. \end{aligned} \quad (3.21)$$

Para $\alpha = 1$ obtemos $r_f^2 = e^{Us} - 1$, indicando que a diversidade da população, em termos de valores de adaptação, cresce exponencialmente com a taxa de mutação.

Por completeza, vamos agora considerar a situação em que os indivíduos com o menor número de mutações pertençam à classe $J \geq 0$, isto é, $C_0 = C_1 = C_2 = \dots = C_{J-1} = 0$ e $C_J \neq 0$. Deste modo, todos somatórios sobre índices de classes começam em $k = J$, ao invés de em $k = 0$. A análise é completamente análoga à descrita para $C_0 \neq 0$ e, em particular, a adaptação média da população no equilíbrio é $T_e = f_J e^{-U}$, sendo portanto dependente de s, α e U . Explicitamente, a equação análoga a (3.13) é

$$C_m = \frac{1}{(1-s)^{J\alpha} - (1-s)^{m\alpha}} \sum_{k=J}^{m-1} \frac{(1-s)^{k\alpha} U^{m-n}}{(m-n)!}. \quad (3.22)$$

para $m = J, J+1, \dots$. Como antes, todas as frequências são proporcionais a C_J de modo que podemos escrever $\mathbf{C} = C_J \mathbf{C}'$ com $C'_J = 1$. A frequência C_J é obtida da condição de normalização, e as demais frequências através da relação $C_k = C_J C'_k$. Para $\alpha = 1$, a distribuição no equilíbrio é uma distribuição de Poisson truncada com parâmetro $\theta = U/s$:

$$C_k = \begin{cases} 0 & k < J \\ \frac{e^{-\theta} \theta^{k-J}}{(k-J)!} & k \geq J \end{cases} \quad (3.23)$$

Os resultados para α geral devem ser obtidos numericamente e serão o tema da seção seguinte. Apesar desses resultados serem válidos para populações infinitas somente,

no capítulo 4 mostraremos como eles podem ser úteis para compreender melhor alguns aspectos da dinâmica de populações finitas.

3.2 Resultados numéricos

Nesta seção vamos apresentar os resultados obtidos utilizando-se a equação (3.13) no caso de $C_0 > 0$ ou a equação (3.22) no caso de $C_k = 0$ para $k < J$ e $C_J > 0$. Na figura 3.1 mostramos as frequências das classes para algumas combinações de α , s e U no caso em que $C_0 > 0$.

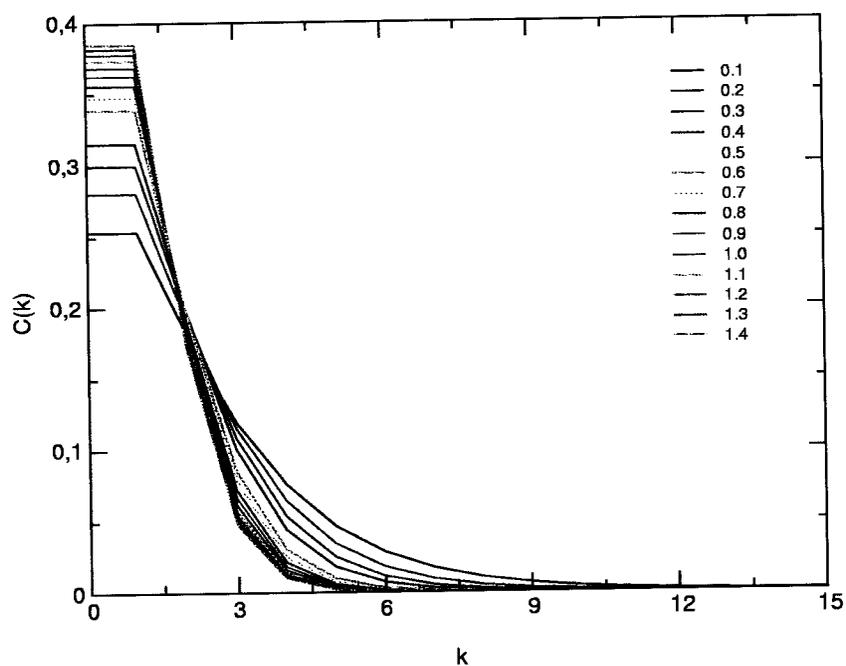


Figura 3.1: Frequência de indivíduos com k mutações para $s = 0.5$ e $U = 0.5$ e diversos valores do parâmetro de epistase α para o caso $C_0 > 0$.

Observamos que quanto maior for o valor do parâmetro de epistase α , maior será a frequência da classe mais apta, isto é, $k = 0$. Esse resultado é esperado, uma vez que valores grandes de α penalizam mais fortemente as classes mutantes. Também por essa

razão classes com muitas mutações ($k = 5$, por exemplo) aparecem com menor frequência no caso de epistase sinérgica, $\alpha > 1$.

Conhecidas as frequências C_k podemos obter o número médio de mutações \bar{k} , a variância do número de mutações $\sigma^2(k)$, e a variância da adaptação da população $\sigma^2(f)$, de onde calculamos a dispersão relativa da adaptação r_f utilizando a equação (3.21). No que segue, graficamos essas grandezas como função da taxa de mutação U , para vários valores dos parâmetros α e s . Uma vez que para $\alpha = 1$ C_k é uma distribuição de Poisson, é interessante graficarmos também \bar{k} contra $\sigma^2(k)$ para cada U , mantendo α e s fixos. Lembramos que no caso de uma distribuição de Poisson temos $\bar{k} = \sigma^2(k)$. Esses gráficos são apresentados nas figuras 3.2, 3.3, 3.4, 3.5 e 3.6.

Observamos dos gráficos que para todos os valores de s e α , tanto \bar{k} como $\sigma^2(k)$ aumentam com U . Para s fixo, essas grandezas crescem muito rapidamente para $\alpha < 1$ e mais lentamente para $\alpha > 1$. Como mostrado na primeira seção deste capítulo, o crescimento é linear para $\alpha = 1$. Esse comportamento deve-se apenas a forma com que a função adaptação dos indivíduos diminui com o número de mutações. Para $\alpha > 1$ a seleção age mais drasticamente quanto maior for o número de mutações do indivíduo, eliminando-o da população, de forma que novas mutações são incluídas com dificuldade, necessitando para isso de uma taxa de mutação elevada. Já para $\alpha < 1$, o efeito eliminatório da seleção sobre os indivíduos que carregam mais mutações é atenuado.

Os gráficos da dispersão relativa da adaptação r_f contra U mostram que, para uma taxa de mutação fixa, a dispersão é menor para menores valores de α , apesar da média e das variâncias no número de mutações serem maiores. Portanto, quanto menos sinérgica e mais atenuante a epistase entre os nucleotídeos das seqüências dos indivíduos da população, maior sua variabilidade genética, no entanto as diferenças de adaptação entre as classes serão menores.

Ainda, esses gráficos indicam comportamentos diferentes para $\alpha < 1$ e $\alpha > 1$.

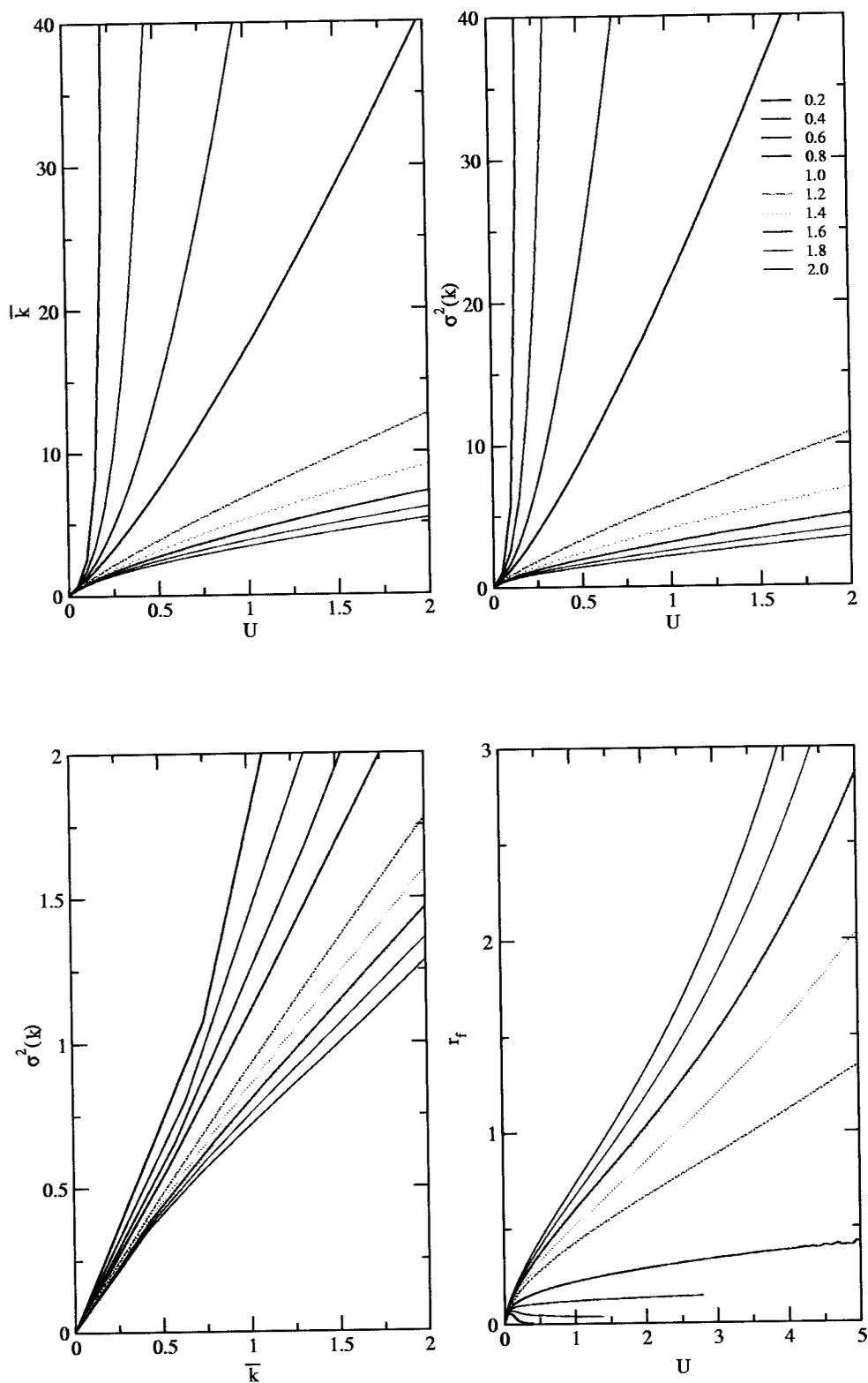


Figura 3.2: Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.1$.

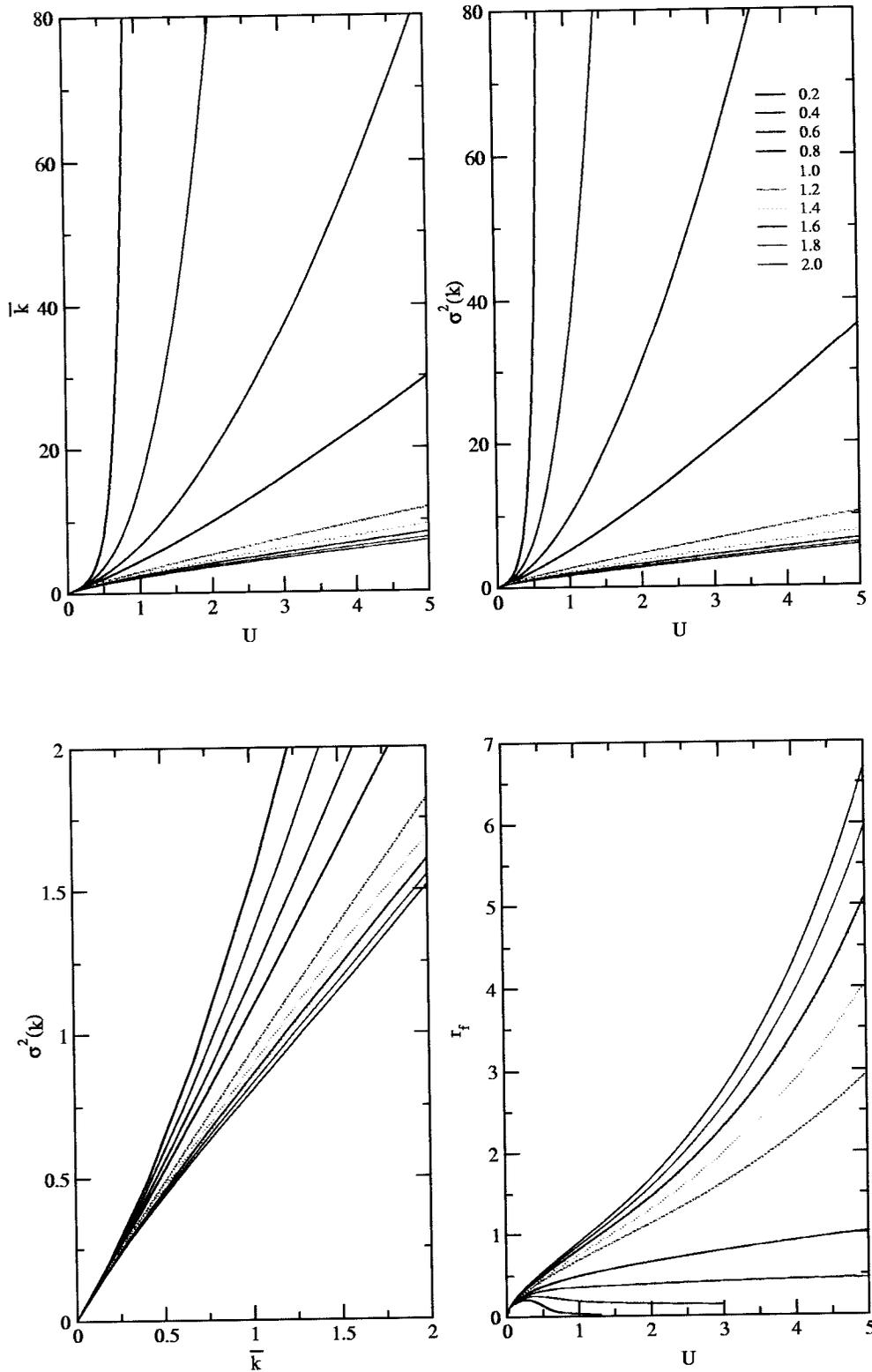


Figura 3.3: Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.3$.

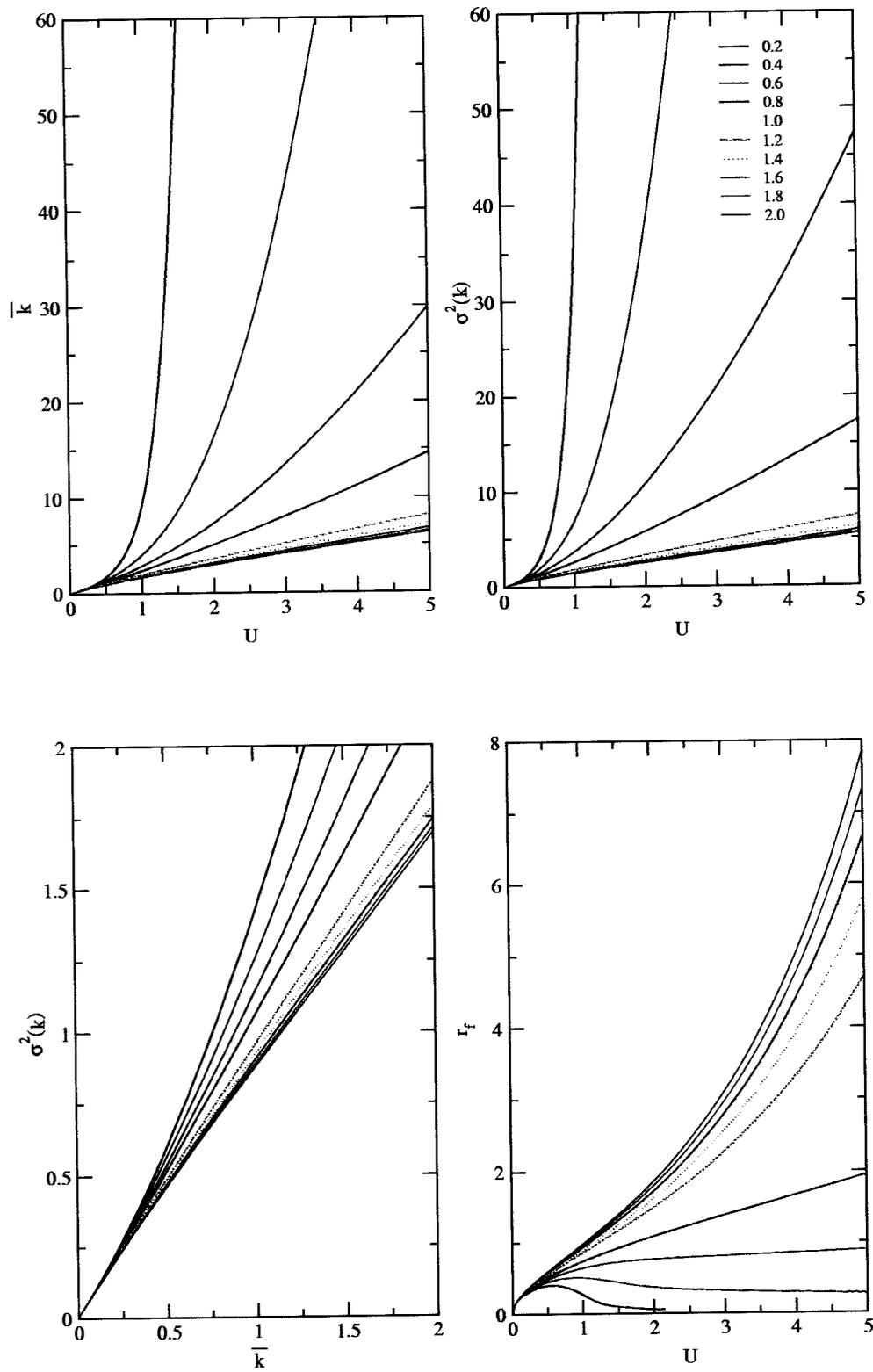


Figura 3.4: Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.5$.

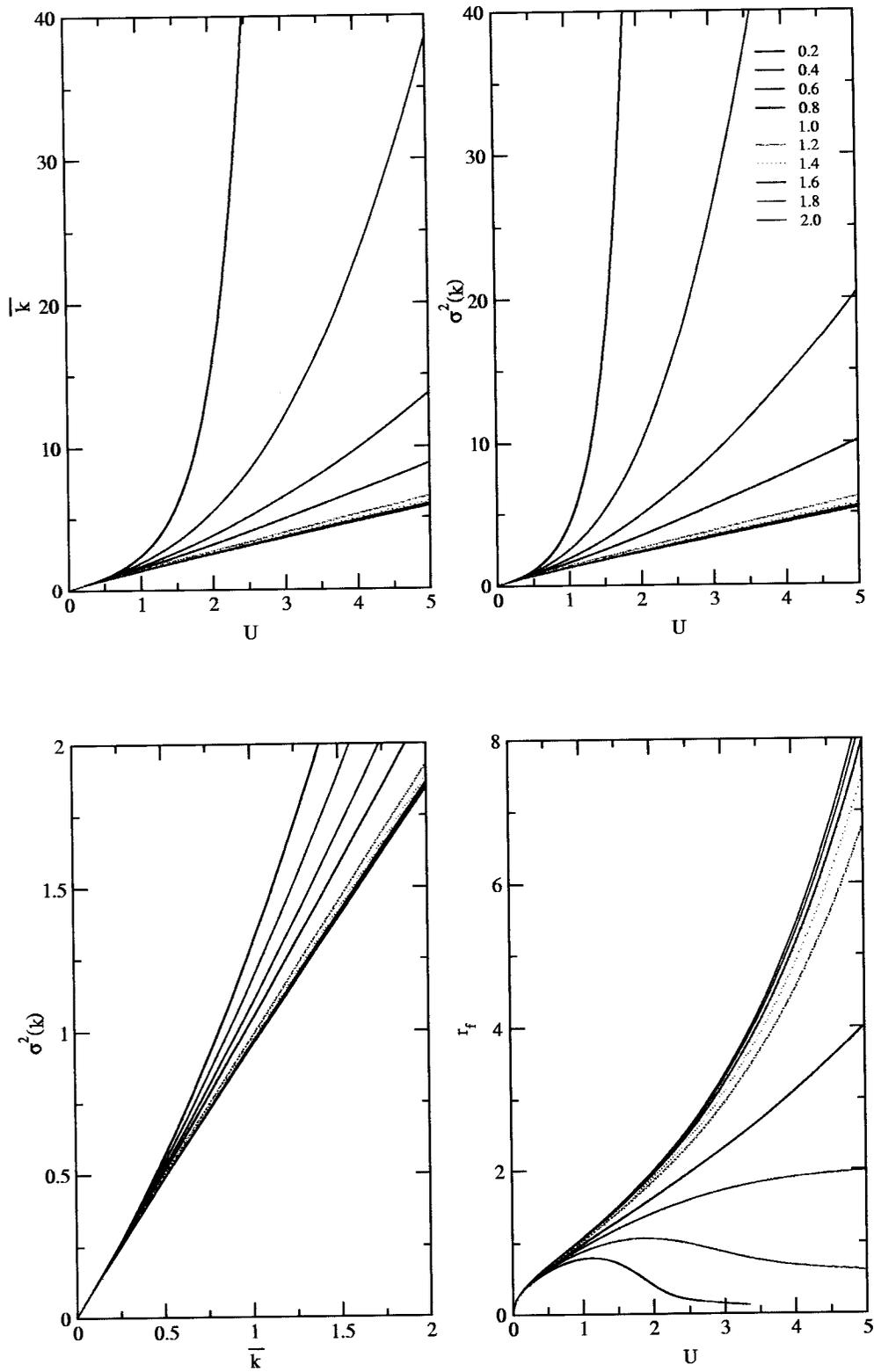


Figura 3.5: Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.7$.

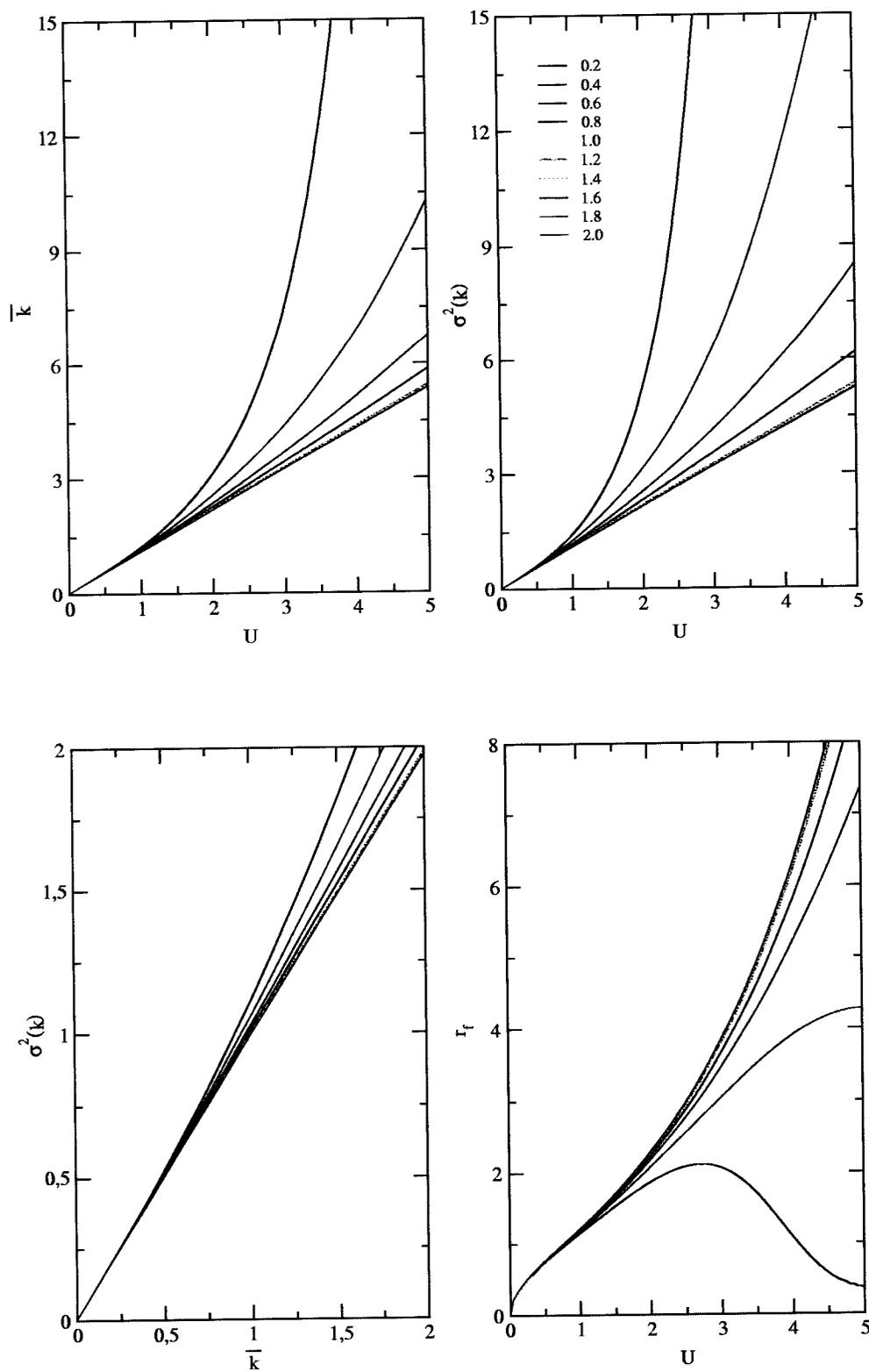


Figura 3.6: Gráficos superiores: número médio de mutações \bar{k} e variância do número de mutações $\sigma^2(k)$ como funções da taxa de mutação U . Gráficos inferiores: $\sigma^2(k)$ versus \bar{k} para U variando entre 0 e $2s$ e dispersão relativa da adaptação da população r_f contra U . O parâmetro seletivo é $s = 0.9$.

Para o caso $\alpha < 1$, $r_f \rightarrow 0$ quando $U \rightarrow \infty$, havendo um valor $U = U_0$ para o qual r_f é máximo. Isso é mais evidente nos gráficos com coeficiente de seleção menores. Infelizmente não foi possível ir além na taxa de mutação U devido à problemas com precisão numérica. Esse comportamento pode ser explicado da seguinte forma: partindo de $U = 0$, a medida que U aumenta, as classes de menor adaptação começam a aparecer fazendo com que $\sigma^2(f)$, e conseqüentemente r_f , aumente; já para U suficientemente grande, \bar{k} é grande e a distribuição de indivíduos nas classes é máxima onde as diferenças de adaptação entre as classes de maior freqüência na população é baixa. Quando U é muito grande, o máximo da distribuição está numa região onde a diferença de adaptação entre os indivíduos na população é nula, realizando um cenário de seleção neutra. Já os gráficos para $\alpha > 1$ ficam acima do gráfico com $\alpha = 1$, cuja expressão analítica, apresentada no texto após à equação (3.21), cresce exponencialmente. Para esses valores de α o efeito da seleção é reforçado fazendo com que exista muita diferença reprodutiva entre as classes, porém havendo pouca variabilidade, ou seja, a maioria da população pertence algumas poucas classes, como vemos dos gráficos da média e da variância no número de mutações.

Aumentando-se o coeficiente seletivo s , as mutações tornam-se mais deletérias, diminuindo portanto os valores de \bar{k} e $\sigma^2(k)$ da população no equilíbrio. Não há entretanto nenhuma mudança qualitativa no comportamento do sistema a medida que s aumenta. Quando $s \rightarrow 1$ as curvas para diferentes valores de α colapsam entre si, indicando que no limite de seleção extrema a distribuição não depende de α . De fato, nesse limite apenas os indivíduos com o menor número de mutações, isto é, aqueles pertencentes à classe C_0 , podem se reproduzir, sendo as demais classes populadas devido somente aos mutantes da classe $k = 0$, tornando a função adaptação independente de α . Assim podemos tomar o limite $s \rightarrow 1$ nas fórmulas para o caso $\alpha = 1$ obtendo $\bar{k} = U$, $\sigma^2(k) = U$ e $\sigma^2(f)/\bar{f}^2 = e^U - 1$, resultados que são válidos para α qualquer.

Finalizando esta seção, vamos apresentar agora os resultados para as distribuições

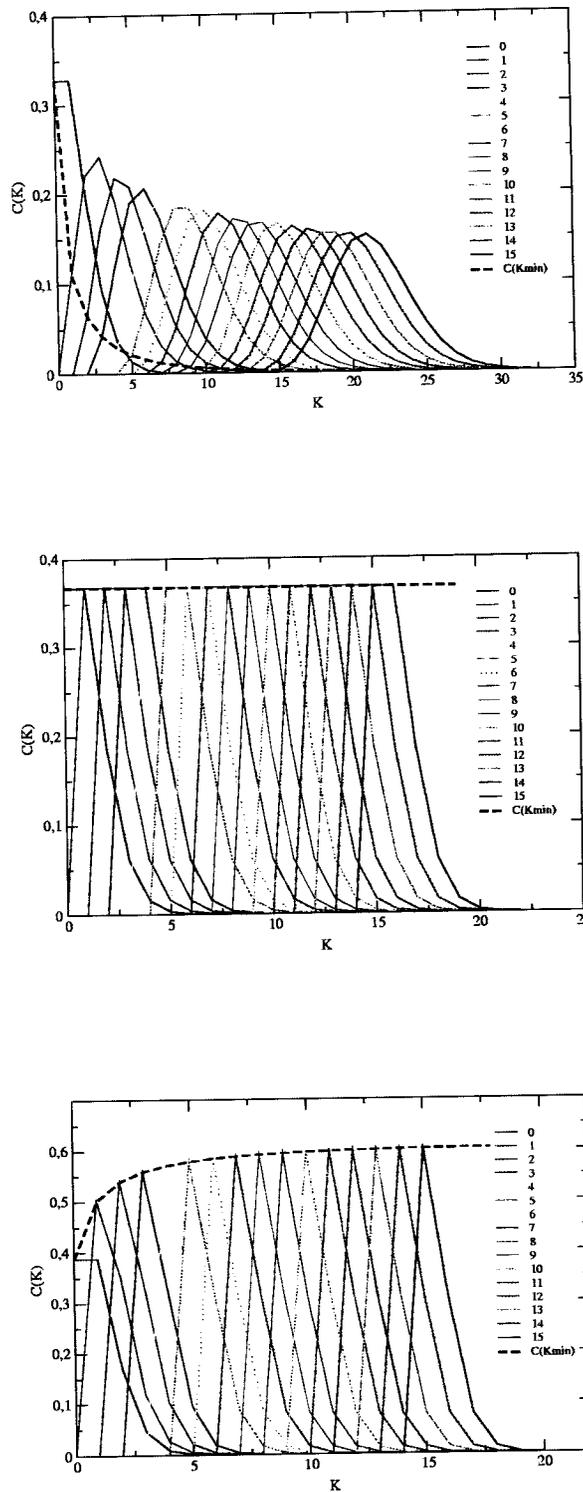


Figura 3.7: Frequência C_k dos indivíduos com k mutações contra o número de mutações k , onde cada curva colorida é o estado de equilíbrio para um valor de classe mais adaptada na população. Os parâmetros utilizados são $S = 0.5$ $U = 0.5$ $\alpha = 0.5$ (topo) ,1.0(meio) e 1.5(base). A curva tracejada é o gráfico da frequência do indivíduos mais adaptado para cada estado de equilíbrio.

de frequências com diferentes valores para o número mínimo de mutações na população, ou seja, a classe com menos mutações e portanto com maior valor de adaptação é $k = J > 0$. Neste caso utilizamos a equação (3.22) para obter as frequências de equilíbrio. Na figura 3.7 mostramos os gráficos dessas distribuições para diferentes valores de J e alguns valores de α . Nesses gráficos, as curvas tracejadas mostram a frequência de equilíbrio C_J da classe com menor número de mutações contra J .

Novamente podemos observar comportamentos bem distintos para $0 < \alpha < 1$, $\alpha = 1$ e $\alpha > 1$. Para $0 < \alpha < 1$ (gráfico do topo da página) ocorre um alargamento das distribuições com o aumento de J , e C_J tende rapidamente a zero a medida que J aumenta. Para $\alpha = 1$ (gráfico do meio) temos o resultado previsto pela equação (4.5), onde as frequências são distribuições de Poisson para qualquer valor de J e $C_J = e^{-U/s}$. Já para $\alpha > 1$, a frequência da classe de maior adaptação cresce com J , estreitando as distribuições.

3.3 Limiar do erro de replicação

Nessa seção estudaremos o limite $\alpha \rightarrow 0$. Neste caso os indivíduos da classe $k = 0$ terão adaptação $f_0 = 1$ e os das demais classes $f_k = 1 - s$ ($k > 1$). Esse relevo de adaptação é chamado de relevo de um pico, e é frequentemente associado à teoria de quase-espécies proposta por Manfred Eigen na década de 70, que tenta explicar a origem e evolução de macromoléculas que seriam as precursoras daquilo que se chama de organismo vivo [27]. Na formulação original daquela teoria, a população também é infinita, porém as seqüências que representam as moléculas são de tamanho finito L (portanto o número de seqüências diferentes também é finito), de modo que faz sentido definir uma taxa de mutação μ por sítio. As características do modelo de quase-espécie mais marcantes são:

- a coexistência da seqüência de maior adaptação com suas variantes mutantes mais

próximas no estado de equilíbrio, produzindo uma distribuição de frequências de classes batizada de quase-espécie;

- a transição de fase entre o regime de quase-espécie e o regime homogêneo, no qual todas as espécies (classes) aparecem com igual frequência. A taxa de mutação crítica para a qual ocorre a transição de fase é denominada de limiar de erro e é função do tamanho L das moléculas e da vantagem seletiva s da molécula com maior eficiência de replicação (molécula mestra).

Considerando o nosso modelo, onde as seqüências possuem infinitos sítios ($L \rightarrow \infty$) e a taxa de mutação U é definida por seqüência, vamos agora determinar a taxa de mutação crítica U_c , acima da qual a seqüência mais eficiente desaparece da população. Seja $C_0(t)$ a frequência da classe zero e $\Pi(t) = \sum_{k=1}^{\infty} C_k(t)$ a frequência total das demais classes. Não é difícil se convencer de que a frequência $C_0(t)$ deve obedecer a seguinte equação de recorrência:

$$\begin{aligned} C_0(t+1) &= \frac{C_0(t)f_0}{C_0(t)f_0 + \Pi(t)f_{k \neq 0}} \mathcal{M}_0 \\ &= \frac{C_0(t)}{C_0(t) + \Pi(t)(1-s)} e^{-U}. \end{aligned} \quad (3.24)$$

Utilizando a condição de normalização $C_0(t) + \Pi(t) = 1$ e considerando a situação de equilíbrio com $C_0(t) = C_0(t+1) = C_0$ obtemos

$$C_0 = \frac{e^{-U} + s - 1}{s}. \quad (3.25)$$

Essa equação descreve a frequência da classe zero enquanto a molécula mestra existe. Na taxa de mutação crítica U_c essa frequência anula-se, daí

$$U_c = -\ln(1-s). \quad (3.26)$$

Pode-se recuperar o resultado obtido por Eigen [22] em termos da taxa de mutação por sítio u e do comprimento das seqüências L usando-se o fato de que a probabilidade de que

não ocorra mutação em uma seqüência é dada por $\mathcal{M}_0 = e^{-U}$ em nosso modelo, sendo dado por $(1 - u)^L$ no modelo de Eigen. Igualando-se essas expressões obtemos

$$u_c = 1 - (1 - s)^{1/L}. \quad (3.27)$$

Capítulo 4

A catraca de Muller: população finita

No capítulo 3 vimos que para populações infinitas a distribuição dos indivíduos em classes tende a uma distribuição de equilíbrio devido ao balanço seleção-mutação. Nesse capítulo vamos estudar o efeito da finitude da população na sua composição genética. No caso de uma população finita o estado da população também se deve a processos estocásticos que fazem flutuar as frequências das classes na população de geração em geração, em contraposição ao cenário determinístico do equilíbrio entre seleção e mutação em populações infinitas. Às flutuações nas frequências de indivíduos ou classes em uma população finita dá-se o nome de *deriva genética*.

4.1 O modelo de Wright-Fisher

Inicialmente vamos apresentar o modelo Wright-Fisher para deriva genética [15] a fim de ilustrar esse fenômeno e também o fenômeno da fixação. Considere uma população de N indivíduos haplóides cujas características fenotípicas são descritas por um único gene que, por sua vez, possui apenas dois alelos segregantes, A e B . A adaptação dos indivíduos é igual para os dois alelos (seleção neutra). Não existe mutação de um alelo para outro e, como no modelo proposto nessa dissertação, para formar a geração posterior são realizados N sorteios com reposição entre os indivíduos da geração atual. O estado

da população pode ser descrito pelo número N_A de alelos do tipo A , de forma que os possíveis estados são $N_A = 0, 1, 2, \dots, N$. Alternativamente podemos fazer a descrição do estado do sistema em termos da frequência $C_A = N_A/N$ do alelo A na população. Os estados $N_A = 0$ e $N_A = N$ são absorventes pois, estando neles, o sistema não sairá mais, e em genética de populações dizemos que ocorreu uma fixação. Apresentamos na figura 4.1 os gráficos do estado da população C_A através das gerações t para algumas simulações realizadas desse modelo para diferentes tamanhos de população. Inicialmente metade dos alelos da população são do tipo A e metade do tipo B . Observamos que o estado da população varia erráticamente com o passar do tempo, ou seja, ocorre a deriva genética do alelo A na população, e não sabemos *a priori* qual será o estado final do sistema: fixação do alelo A ($N_A = N$) ou B ($N_A = 0$). Também observamos que para populações maiores o tempo para que ocorra a fixação aumenta. Se o tamanho da população fosse infinito não haveria deriva genética e as frequências de alelos seriam constantes e iguais ao estado inicial, o que corresponde ao princípio de Hardy-Weinberg [15]. Realizamos essa simulação para um número grande de populações independentes, acompanhando o estado dessas populações no tempo. Na figura 4.2 apresentamos a frequência de populações F_{pop} pela frequência C_A do alelo A em tais populações para diferentes valores de tempo de simulação.

O modelo de Wright-Fisher pode ser formulado matematicamente em termos de cadeias de Markov [15]. Seja $C^A(t)$ um vetor de dimensão $N + 1$, tal que o i -ésimo elemento $C_i^A(t)$ represente a probabilidade de uma população ter i alelos A na geração t . Naturalmente, os elementos desse vetor estão sujeitos à condição de normalização

$$\sum_{i=0}^N C_i^A = 1. \quad (4.1)$$

Suponha que na geração 0 tenhamos $N_A = N_A^0$, ou seja, $C_n^A(0) = \delta_{nN_A^0}$ com $n = 0, 1, 2, \dots, N$. A probabilidade de transição T_{ij} de que $N_A = j$ na geração $t + 1$, da-

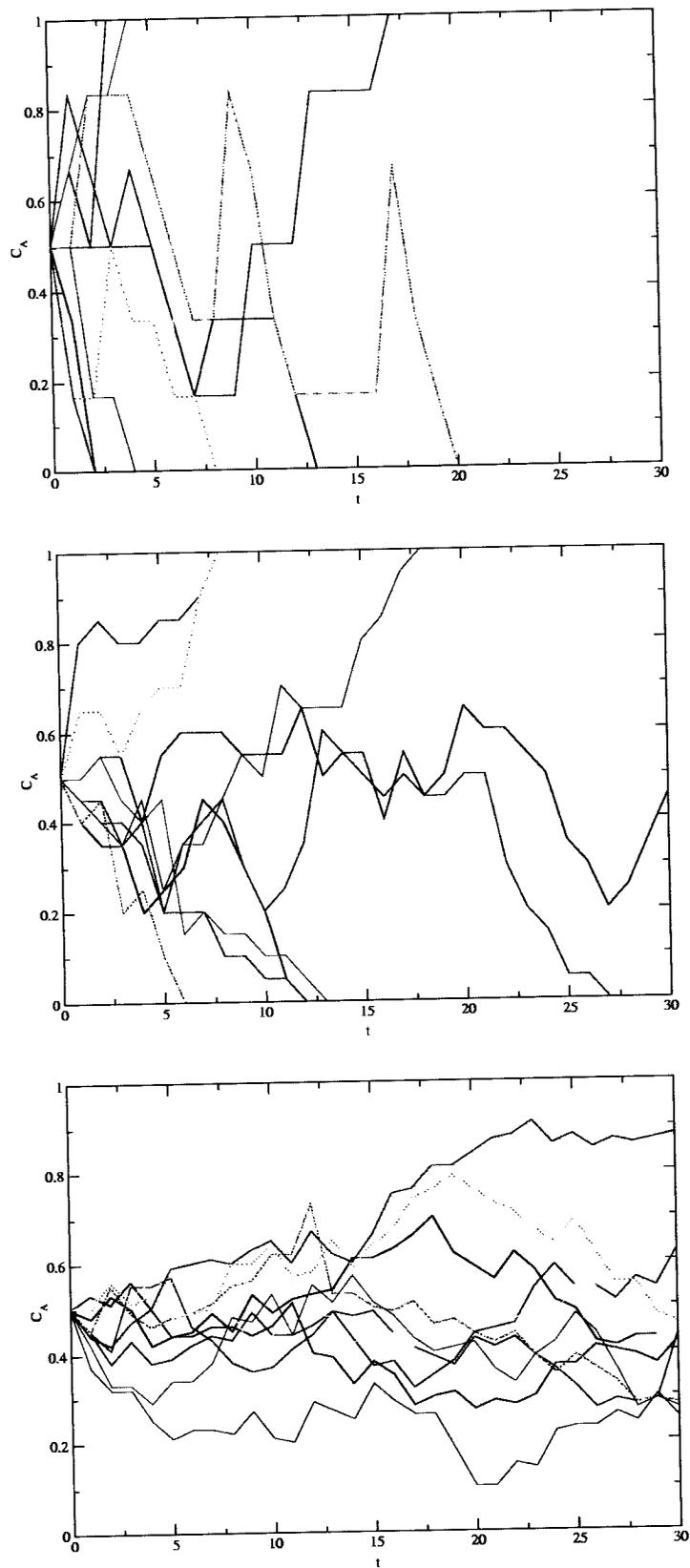


Figura 4.1: Frequência C_A do alelo A contra geração t para simulações do modelo de Wright-Fisher, com tamanhos de população $N=6$ (topo), $N=20$ (meio) e $N=100$ (base). Cada cor diferente corresponde a evolução de uma população independente.

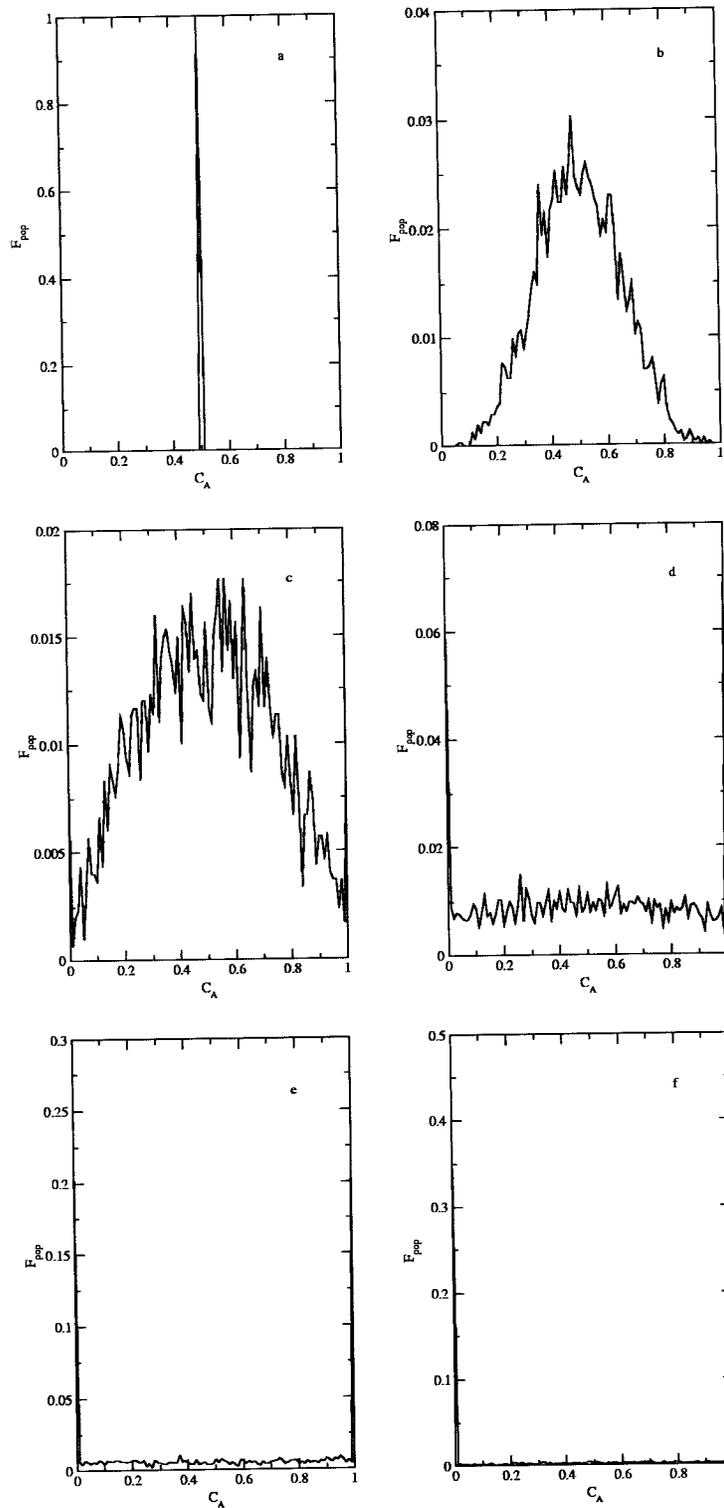


Figura 4.2: Frequência de populações F_{pop} contra estado de frequência C_A do alelo A para simulações com 3000 populações com 50 indivíduos. São mostrados gráficos para tempos de simulação a) $t=3$, b) $t=10$, c) $t=25$, d) $t=50$, e) $t=100$ e f) $t=200$. Inicialmente todas as populações possuem metade dos seus alelos no estado A. Com o passar do tempo a distribuição alarga-se e, para tempos muito longos, metade das populações se fixa em A e a outra metade em B.

do que $N_A = i$ na geração t é dada pela binomial

$$T_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}. \quad (4.2)$$

De fato, essa probabilidade de transição define o modelo de Wright-Fisher. Conhecida a matriz \mathbf{T} , formada pelos elementos T_{ij} , podemos determinar a distribuição de probabilidade do número N_A de alelos A na população para qualquer geração t dada a distribuição de probabilidade do estado inicial, descrita pelo vetor $\mathbf{C}^A(0)$. A distribuição de probabilidade do estado da população na geração $t + 1$ é obtida pela multiplicação da matriz \mathbf{T} pela distribuição na geração t , isto é,

$$\mathbf{C}^A(t+1) = \mathbf{T}\mathbf{C}^A(t), \quad (4.3)$$

caracterizando então uma cadeia de Markov. A distribuição de equilíbrio, denotada por $\overline{\mathbf{C}}^A$, satisfaz a seguinte equação

$$\overline{\mathbf{C}}^A = \mathbf{T}\overline{\mathbf{C}}^A. \quad (4.4)$$

Dado o estado inicial $\mathbf{C}^A(0) = \delta_{nN_A^0}$, verificamos que a solução dessa equação é

$$\overline{C}_n^A = \begin{cases} \frac{N_A^0}{N} & n = 1 \\ 1 - \frac{N_A^0}{N} & n = 0 \\ 0 & \text{outro caso} \end{cases} \quad (4.5)$$

que dá, portanto, as probabilidades de fixação dos alelos A ($n = 1$) e B ($n = 0$).

4.2 A catraca de Müller

Ilustrados os fenômenos da deriva genética e da fixação devido à finitude da população, nesta seção voltamos a tratar o modelo de evolução de populações proposto e discutido nos capítulos anteriores. Lembremos que, naquele modelo, uma população é definida pelos parâmetros tamanho de população N , coeficiente de seleção s , coeficiente de espistase α e taxa de mutação U , e que os indivíduos são descritos por seqüências binárias infinitas.

O conceito de fixação é um pouco diferente do utilizado no modelo de Wright-Fisher, no qual dizia-se que os dois alelos eram segregantes quando ambos estavam presentes na população, e o evento de fixação significava que toda a população era composta por um deles. No modelo analisado nesta dissertação, porém, esse conceito é aplicado aos sítios mutantes. Desta forma, dizemos que um sítio homólogo é segregante quando a população exhibe os dois estados possíveis para ele – o original 0 e o mutante 1 – e dizemos que uma fixação ocorre quando todos os indivíduos da população apresentam o estado mutante 1 para aquele sítio. Assim sendo, o modelo supõe que inicialmente não haja fixações na população (todos os sítios no estado 0). Quando ocorre uma mutação em um sítio de um indivíduo, ela começa a segregar, e o número de indivíduos que possuem aquela mutação, ou seja, os descendentes do indivíduo que sofreu a mutação pela primeira vez, varia à cada geração devido tanto a deriva genética quanto a seleção. A longo prazo, há apenas duas possibilidades para essa mutação: ou (i) o número de indivíduos que apresentam a mutação começa a crescer e, em um dado instante, todos os indivíduos da população apresentam aquela mutação, ocorrendo então uma fixação, ou (ii) o número de indivíduos que apresentam essa mutação se mantêm em níveis reduzidos e, em um dado instante, são extintos da população, de forma que todos os indivíduos da população apresentem novamente o estado original naquele sítio.

Posto isso, passamos a descrever o fenômeno denominado *catraca de Müller*, devido também à finitude da população. Como no capítulo 3, continuaremos a representar a população pelo vetor $\mathbf{X}(t)$ de dimensão infinita, onde cada componente X_i dá o número de indivíduos na população com i mutações (ditos pertencentes à classe i). A catraca de Müller é a perda sucessiva e irreparável dos indivíduos da classe mais baixa (mais adaptados) da população devido às flutuações decorrentes da deriva genética. Segundo Kondrashov [17], esse fenômeno pode ser descrito da seguinte forma: iniciando a dinâmica do modelo temos que o estado atual $\mathbf{X}(t)$ do sistema tende a um estado dado pelo vetor

$\bar{\mathbf{X}}^0$ cujas componentes \bar{X}_i^0 são dadas pelo inteiro mais próximo do produto NC_i^0 , onde vale $\sum_{i=0}^{\infty} \bar{X}_i^0 = N$, e C_i^0 é a componente i do vetor de frequências que descreve o estado estacionário no limite determinístico ($N \rightarrow \infty$), dado pela equação (3.17), com os mesmos valores para os parâmetros s , α e U para a população com tamanho finito. Os super-índices 0 em \mathbf{C}^0 e $\bar{\mathbf{X}}^0$ são utilizados para indicar que nas equações de recorrência determinísticas a frequência inicial de indivíduos da classe $k = 0$ é não nula e, conseqüentemente, permanecerá não-nula para todas as gerações posteriores. Porém, como já foi mencionado, a finitude da população provoca flutuações no estado atual $\mathbf{X}(t)$ em torno de $\bar{\mathbf{X}}^0$. Se a amplitude da flutuação em um dado instante t_0 for tal que o número de indivíduos na classe zero atinja o valor zero ($X_0(t_0) = 0$), todos os indivíduos da população terão ao menos uma mutação e, como desconsideramos mutações reversas e recombinação que poderiam produzir indivíduos com menor número de mutações, a classe zero está perdida e não pode mais ser recuperada. A partir desse instante, o estado da população flutua em torno de um estado $\bar{\mathbf{X}}^1 = N\mathbf{C}^1$ onde $\bar{X}_0^1 = 0$, e a classe com menor número de mutações é a classe $k = 1$ (o super-índice 1 indica que a recorrência determinística é iniciada com $C_0 = 0$ e $C_1 > 0$) com \mathbf{C}^1 dado pela equação (4.5) com $J = 1$ se $\alpha = 1$, ou pelo método de recorrência descrito no capítulo 3 para α qualquer, baseado na equação (3.22). Novamente as flutuações podem eliminar definitivamente a classe $k = 1$ e depois disso a classe de menor número de mutações pertencente à população é $k = 2$, e assim por diante.

Portanto no fenômeno da catraca de Müller estão em jogo três forças evolucionárias: mutação, seleção e a deriva genética. Populações com taxas de mutação altas e tamanho pequeno, resultando em uma grande deriva genética, acumularão mutações deletérias rapidamente, quase que independentemente do grau do dano causado pelas mutações. Já para populações de grande tamanho, a deriva genética atua com menor intensidade, e a população permanece num estado *quase-estacionário* por um tempo muito

grande, resultando numa acumulação de mutações mais lenta. Considerando esse último caso, aumentando-se a taxa de mutação, diminuimos o número de indivíduos na classe mais baixa, facilitando a sua perda. Esse processo de perdas sucessivas de classes de mutação e a conseqüente acumulação de mutações deletérias foi denominado catraca de Müller em reconhecimento a H. J. Müller que descreveu o fenômeno em seu artigo de 1964 [7]. Se esse processo ocorresse indefinidamente e o tamanho da população pudesse variar, por exemplo como função da adaptação média da população, haveria a degradação e eventual extinção dessa população [18], de modo que a catraca de Müller seria uma explicação para a vantagem seletiva da recombinação e do sexo como uma forma de obter indivíduos com menor número de mutações. Existe ainda uma certa controvérsia sobre esse resultado, pois como veremos logo adiante, ajustando o parâmetro de epistase convenientemente podemos parar a catraca. Além disso, o tamanho usual de populações naturais haplóides é relativamente grande e alguns autores afirmam que níveis de recombinação muito baixos já seriam suficientes para evitar o efeito da catraca [13].

O estado da catraca na população é dado pelo número de mutações do indivíduo mais apto na população. É importante ressaltar que a catraca move-se devido a dois processos: (i) surgimento e fixação de uma única mutação na população, e (ii) surgimento simultâneo de mutações diferentes em todos os indivíduos. Assim, fica evidente a diferença entre o número mínimo de mutações e o número de mutações fixadas. Desenvolvemos um programa de computador que simula o modelo estocástico (população finita) descrito nos capítulos anteriores e realizamos um estudo sistemático da dinâmica de mutações deletérias na população. Na figura 4.3 mostramos um gráfico do número N_k de indivíduos na classe k como função da classe k para uma única população em diferentes gerações. Nesse gráfico podemos observar o efeito da catraca nas classes de indivíduos como um pacote de onda propagando-se para direita com velocidade e formato médio constantes. O fato da velocidade de propagação do pacote ser constante deve-se, como

veremos logo adiante, à escolha do valor $\alpha = 1$ para o coeficiente de epistase. Neste caso, a forma do pacote de onda é aproximadamente o de uma distribuição de Poisson, como no caso determinístico dado pela equação (3.17). Nas figuras 4.4 apresentamos outros

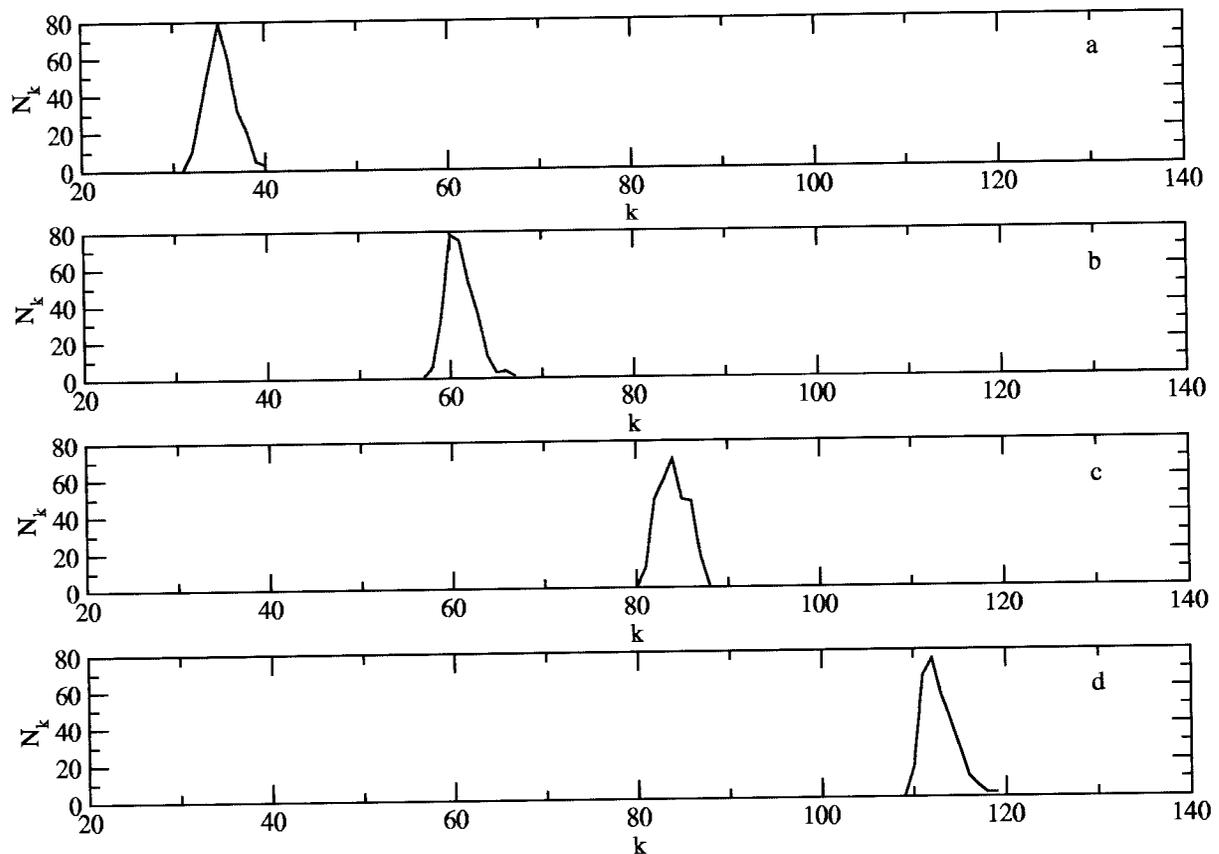


Figura 4.3: Gráfico do número N_k de indivíduos com k mutações (número de indivíduos na classe k) contra o número de mutações para diferentes gerações t ; a) $t=2000$, b) $t=4000$, c) $t=6000$ e d) $t=8000$. A catraca aparece como uma onda propagando-se para a direita com velocidade e forma constantes. Os parâmetros utilizados foram $N = 300$, $s = 0.1$, $U = 0.3$ e $\alpha = 1.0$.

dois gráficos mostrando o efeito da catraca. Graficamos o número médio de mutações $\langle k \rangle$, o número mínimo de mutações k_{min} , o número máximo de mutações k_{max} , e o número de mutações fixados k_{fix} contra a geração. No gráfico superior, apresentamos essas grandezas para apenas uma população, enquanto que no gráfico inferior mostramos essas mesmas grandezas mediadas em um número grande de populações independentes.

Para distinguirmos esses dois tipos de média usaremos a notação $\langle \dots \rangle$ para representar a média em uma única população, e a notação $\overline{\dots}$ para representar a média sobre várias populações independentes. Desta maneira, $\overline{\langle k \rangle}$ significa a média do número médio de mutações em várias populações independentes. No gráfico para uma única população podemos observar claramente que k_{fix} e k_{min} aumentam na forma de degraus de diversas alturas, enquanto que $\langle k \rangle$ aumenta de forma relativamente suave e k_{max} flutua violentamente. Esse comportamento para as fixações pode ser explicado pela estrutura da genealogia da população, ilustrada na figura 4.5. Nessa figura vemos uma árvore genealógica possível para uma população haplóide sem recombinação. O tempo aumenta no sentido para baixo da figura e a reta tracejada representa a geração atual. Os indivíduos indicados como A , B e C são ancestrais da geração atual, sendo A o último ancestral comum de todos os oito indivíduos presentes. Devido a população ser constituída de indivíduos haplóides e estarmos supondo que o comprimento das seqüências é infinito, o número de mutações do ancestral A é igual ao número de mutações fixadas em toda a população, pois todos os oito indivíduos da população atual possuem as mesmas mutações que existiam naquele indivíduo, somadas as novas mutações segregantes na população que ocorreram entre a geração em que A viveu e a geração atual. Por outro lado, a população atual pode ser subdividida entre os indivíduos que são descendentes de B e os que são descendentes de C . Da mesma forma, todos os indivíduos de um mesmo grupo têm fixadas as mutações do último ancestral comum do grupo (B ou C). Devido à deriva genética, seguindo adiante no tempo, em um dado instante todos os indivíduos descendentes de B ou de C serão eliminados da população. Caso os descendentes de B sejam eliminados, o indivíduo C se tornará o último ancestral comum de toda a população e todas as mutações que ocorreram entre A e C são fixadas simultaneamente. O gráfico inferior da figura 4.4 mostra que se realizarmos médias das mutações fixadas e dos números mínimo, máximo e médio de mutações na população sobre várias populações independentes, essas curvas

tornam-se suaves e crescem no tempo a uma mesma taxa R . Definimos então essa taxa R como sendo a velocidade da catraca de Müller. A diferença $\overline{\langle k \rangle} - \overline{k_{fix}}$ corresponde ao número médio de mutações segregantes por indivíduo por população.

Nas figuras 4.6, 4.7 e 4.8 mostramos a dependência do número médio $\overline{\langle k \rangle}$ de mutações (gráfico superior) e da taxa R de aumento dessa grandeza na população (gráfico inferior) com o tempo t , para alguns conjuntos de parâmetros de controle do modelo. O valor de R é obtido através da derivação numérica das curvas de $\overline{\langle k \rangle}$ versus t . Os gráficos à esquerda são para $\alpha \leq 1$, enquanto que os da direita são para $\alpha > 1$. Desses gráficos observamos que para t suficientemente grande, a taxa R tende para um valor constante não nulo, ou seja, as mutações acumulam-se linearmente com o tempo transcorrido, com valores de R que variam de forma decrescente desde $R = U$ para o caso neutro ou para $\alpha \rightarrow 0$, até $R = 0$ correspondente ao caso $\alpha \rightarrow \infty$. Vemos que para $\alpha \leq 1$, $\overline{\langle k \rangle}$ é uma função côncava de t , enquanto que para $\alpha > 1$ esta função torna-se convexa. Isso dificulta a obtenção de uma conclusão definitiva, através da análise dos dados das simulações, se R de fato tende para um limite não nulo no caso de epistase sinérgica ($\alpha > 1$) mas, pelo menos, os resultados mostram claramente que R diminui com o aumento de α . Essa diminuição deve-se ao aumento da adaptação relativa dos indivíduos da classe com menor número de mutações, dificultando assim a perda desses indivíduos. Naturalmente, esse efeito é mais pronunciado quanto maior for o valor do coeficiente seletivo s . No gráfico da base da figura 3.7 do capítulo 3 (limite determinístico) vemos que a frequência do indivíduo mais adaptado aumenta quando o número mínimo de mutações na população aumenta. O mesmo ocorre para populações de tamanho finito com $\alpha > 1.0$: o número de indivíduos na classe mais adaptada aumenta com o número mínimo de mutações na população, tornando-se cada vez mais difícil por deriva genética eliminá-la da população, retardando, assim, o efeito da catraca. As oscilações observadas nas curvas de R versus t para $\alpha > 1$ são devidas ao procedimento numérico utilizado para efetuar a derivação:

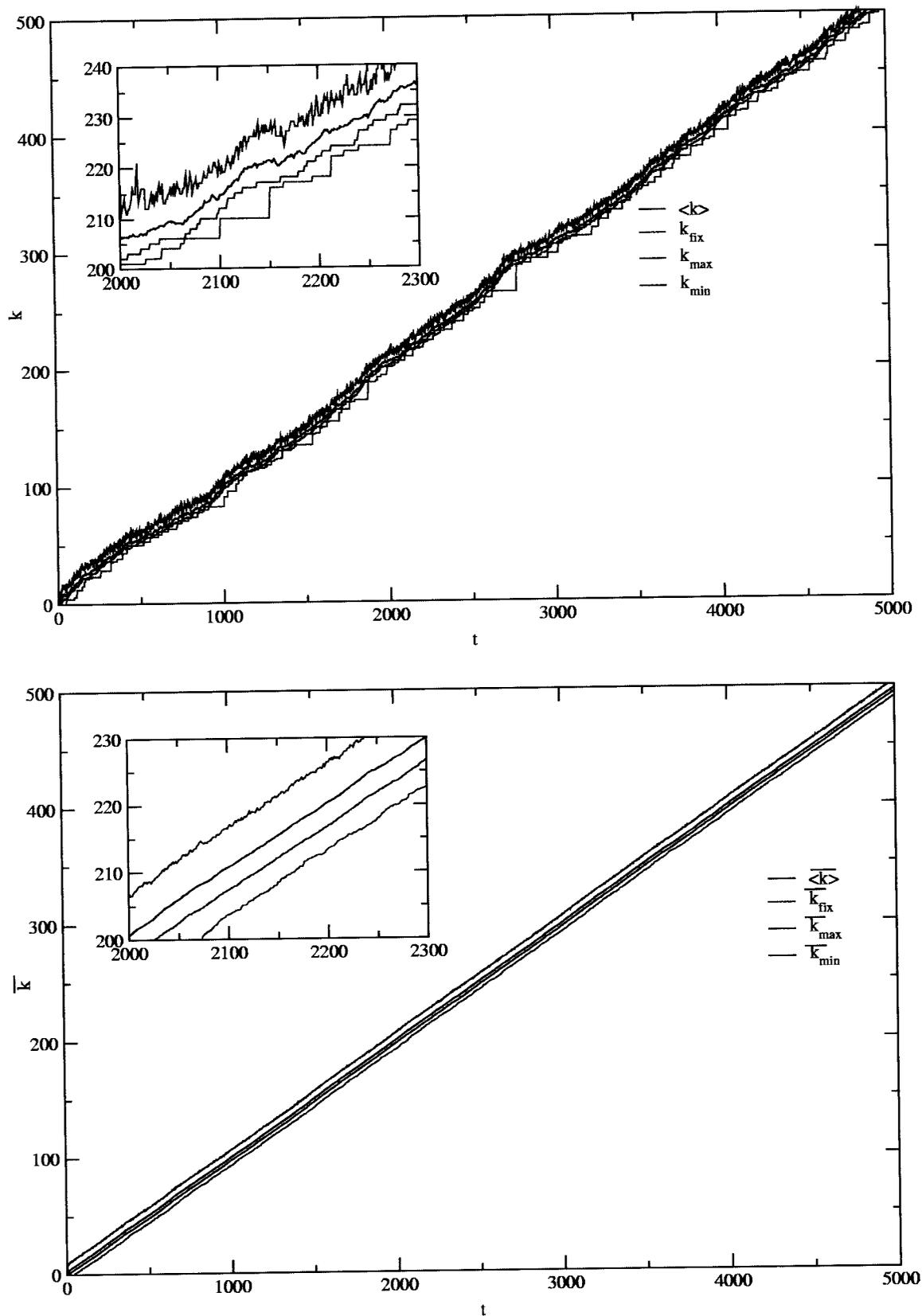


Figura 4.4: Gráficos de k_{med} , k_{fix} , k_{min} e k_{max} contra o tempo mostrando a acumulação de mutações na população. Os parâmetros utilizados foram $N = 100$, $s = 0.1$, $U = 0.5$ e $\alpha = 1.0$. O gráfico superior apresenta os resultados para uma única população, enquanto que o inferior apresenta resultados mediados sobre 50 populações independentes.

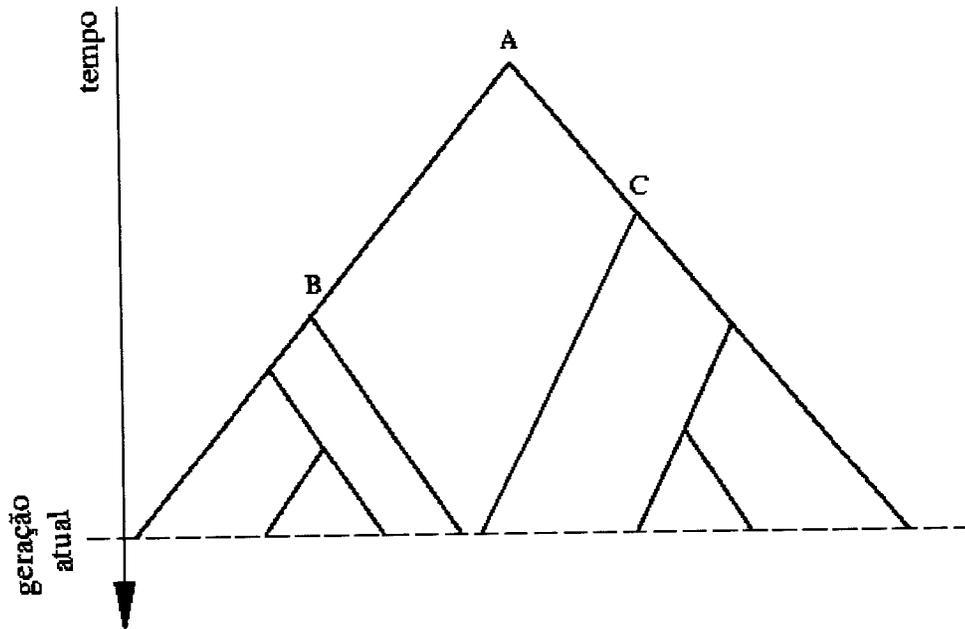


Figura 4.5: A Figura mostra uma genealogia possível para uma população haplóide sem recombinação. A linha pontilhada representa a geração atual, e as pontas das linhas da genealogia que terminam na linha pontilhada representam os oito indivíduos da geração atual. Seguindo a história passada de todos esses indivíduos podemos encontrar seus ancestrais comuns. Os indivíduos indicados por *A*, *B* e *C* são alguns dos ancestrais da população, sendo *A* a ancestral comum mais recente de toda a população.

primeiramente fazemos a interpolação da curva $\overline{\langle k \rangle}$ versus t por um polinômio de ordem 10 e então efetuamos a derivada. Assim, embora a interação sinérgica das mutações ($\alpha > 1$) é seja muito mais prejudicial aos indivíduos portando novas mutações, ela é benéfica para a população como um todo, pois retarda a degradação da adaptação média da população. Visto que, pelo menos no caso em que o coeficiente de epistase α está entre 0 e 1, a velocidade da catraca R tende para um valor constante não nulo para valores de t da ordem de 10^4 , podemos neste caso estimar R para diversos conjuntos de parâmetros. Em particular, mostramos nos gráficos das figuras 4.9 a velocidade da catraca R contra o coeficiente de seleção s para vários valores de α nesse intervalo. Já na figura 4.10 mostramos um gráfico com a velocidade R contra a taxa de mutação U para coeficiente de seleção $s = 0.5$ constante. Destes gráficos podemos observar que, como visto em outros gráficos, a velocidade da catraca diminui à medida que α aumenta. Vemos também que a

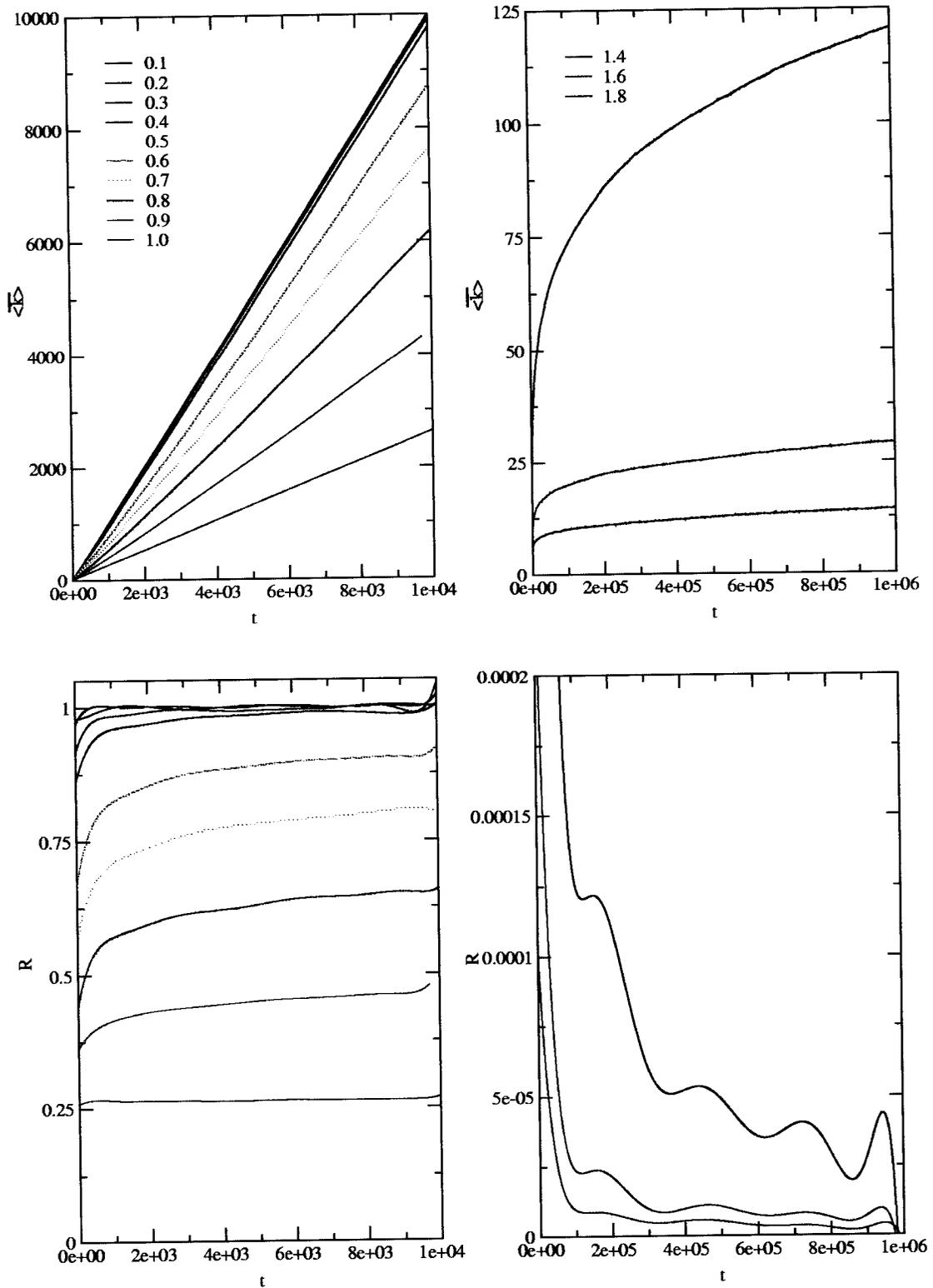


Figura 4.6: Gráficos superiores: número médio de mutações \overline{k} contra a geração t . Gráficos inferiores: velocidade da catraca R contra a geração t . Gráficos à esquerda: valores de $0.1 \leq \alpha \leq 1$. Gráficos à direita: valores de $\alpha > 1.0$. O tamanho da população é $N = 30$, o parâmetro seletivo é $s = 0.2$ e a taxa de mutação é $U = 1.0$. Os dados representam médias sobre 150 populações.

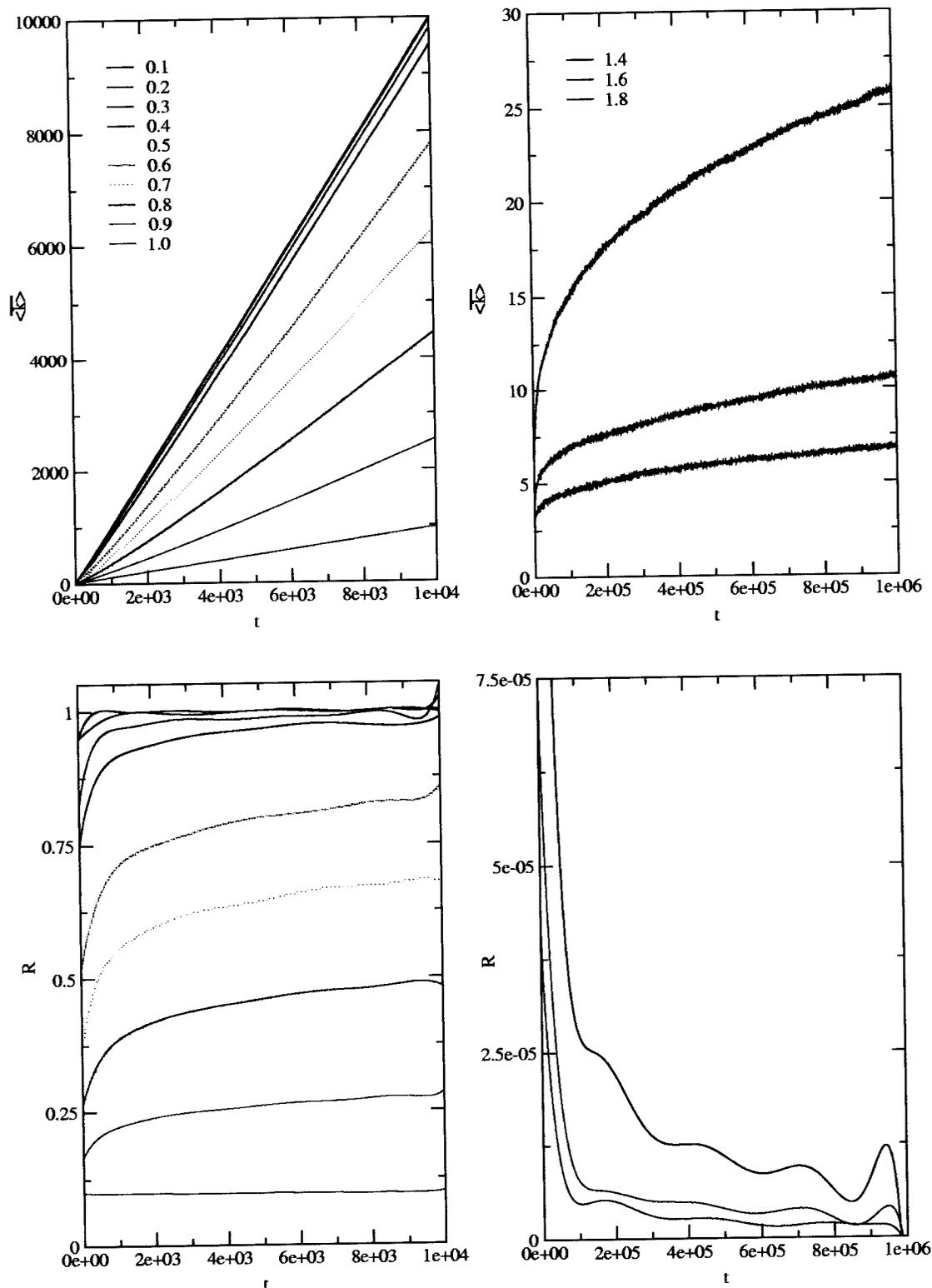


Figura 4.7: Gráficos superiores: número médio de mutações \overline{k} mutações contra a geração t . Gráficos inferiores: velocidade da catraca R contra a geração t . Gráficos à esquerda: valores de $0.1 \leq \alpha \leq 1.0$. Gráficos à direita: valores de $\alpha > 1.0$. O tamanho da população é $N = 30$, o parâmetro seletivo é $s = 0.4$ e a taxa de mutação é $U = 1.0$. Os dados representam médias sobre 150 populações.

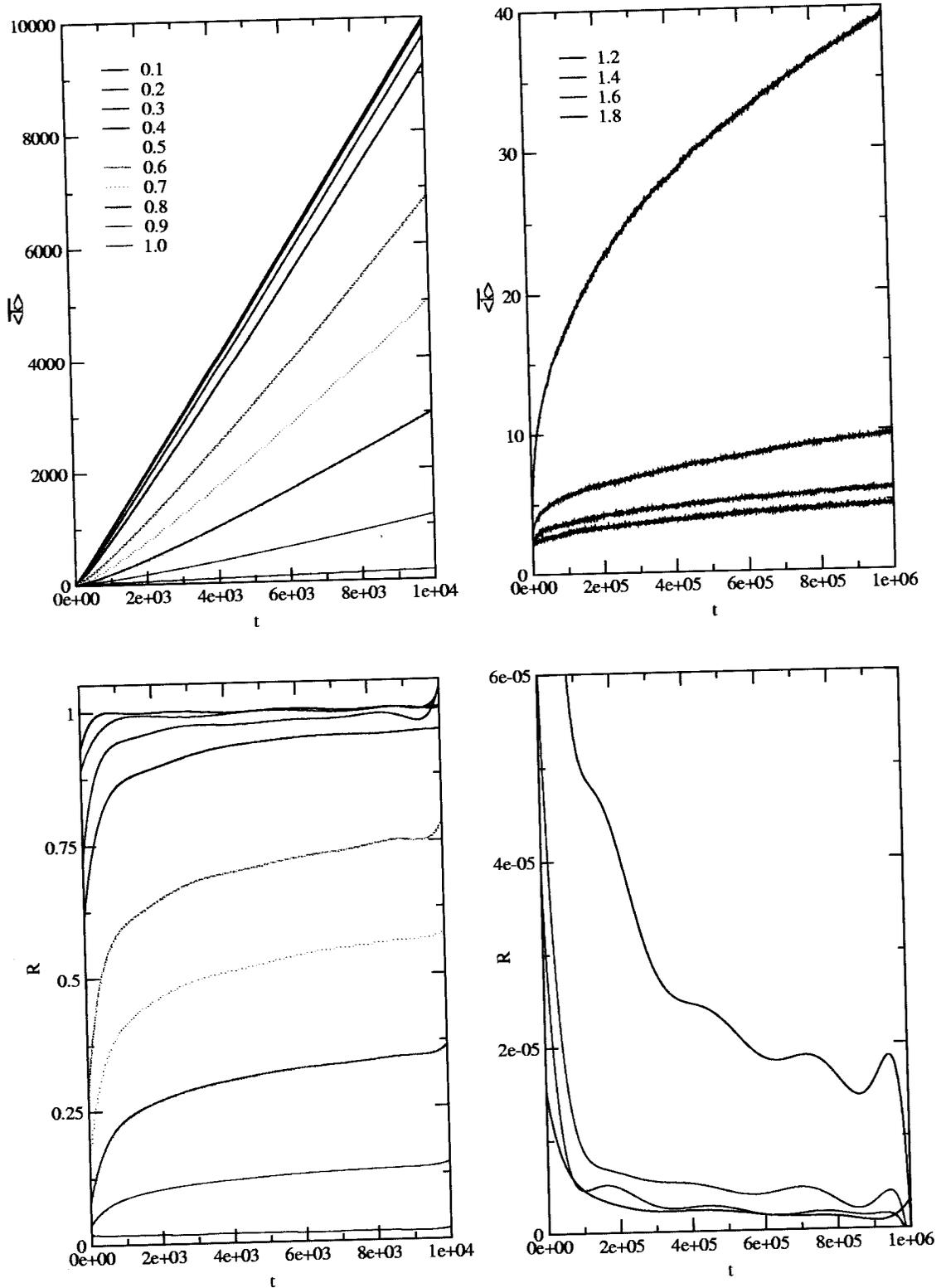


Figura 4.8: Gráficos superiores: número médio de mutações \overline{k} mutações contra a geração t . Gráficos inferiores: velocidade da catraca R contra a geração t . Gráficos à esquerda: valores de $0.1 \leq \alpha \leq 1$. Gráficos à direita: valores de $\alpha > 1.0$. O tamanho da população é $N = 30$, o parâmetro seletivo é $s = 0.6$ e a taxa de mutação é $U = 1.0$. Os dados representam médias sobre 150 populações.

velocidade da catraca diminui continuamente com s , com $R \rightarrow 0$ para $s \rightarrow 1$. Vemos pelo gráfico da figura 4.10 que α próximo de 1, a velocidade da catraca aumenta lentamente para U pequeno, devido à seletividade da população, crescendo linearmente com U , como no caso neutro, quando este assume valores suficientemente altos, porém a uma taxa menor, pois praticamente todos os indivíduos nascerão com novas mutações.

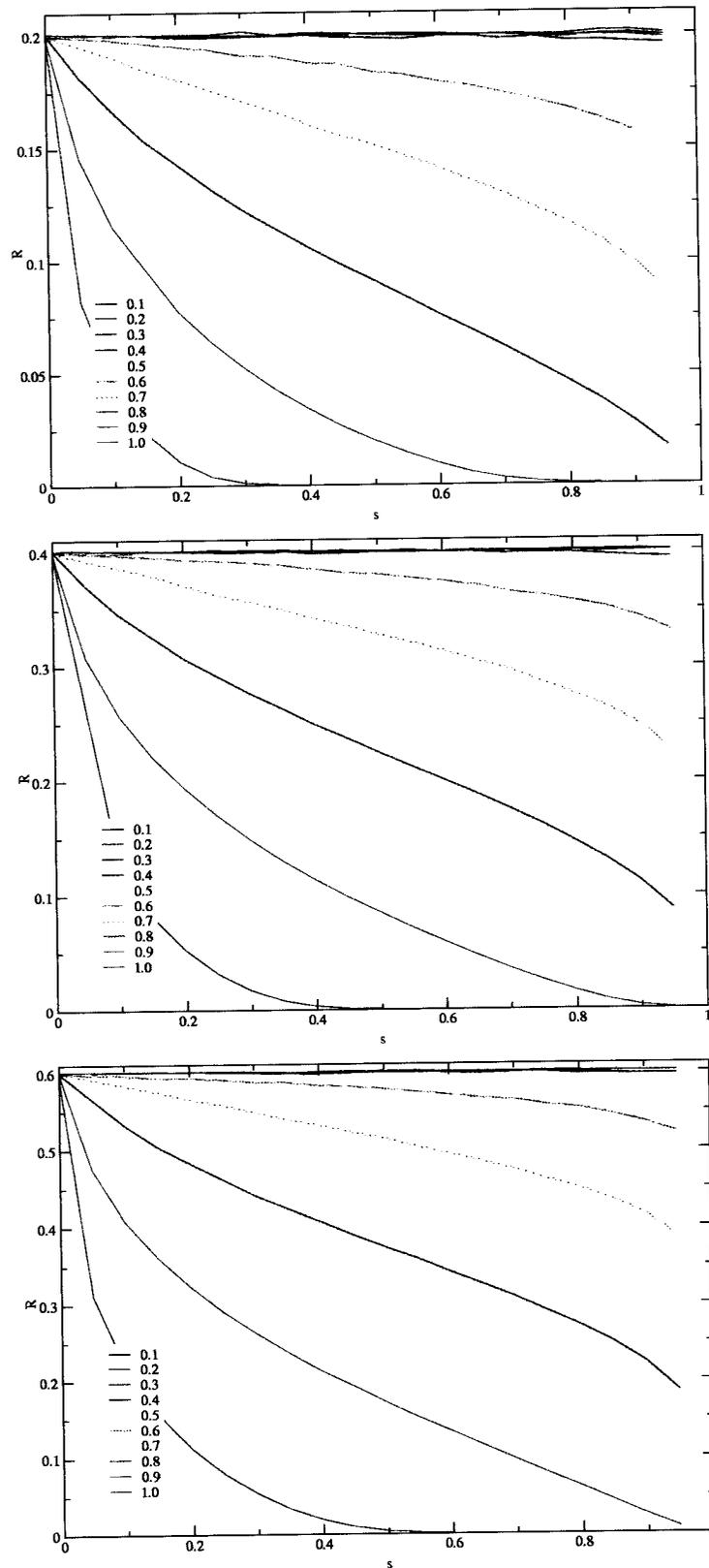


Figura 4.9: Velocidade R da catraca contra parâmetro seletivo s para parâmetros de epistase α entre 0 e 0.1. Utilizamos taxa de mutação U igual a 0.2 no topo, 0.4 no meio e 0.6 na base. Os dados representam médias sobre 200 populações de 30 indivíduos cada.

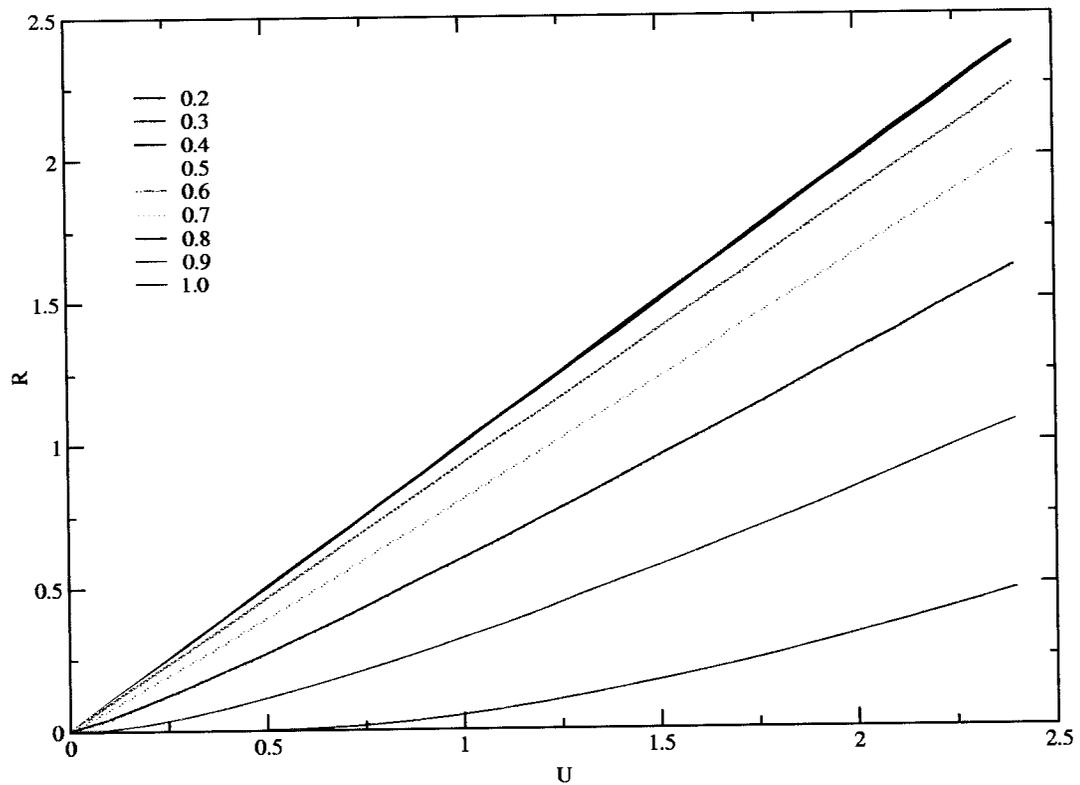


Figura 4.10: Velocidade R da catraca contra a taxa de mutação U para parâmetros de epistase α entre 0 e 1.0. Os dados representam médias sobre 300 populações de 30 indivíduos cada e o parâmetro seletivo utilizado é $s = 0.5$

Capítulo 5

Genealogias

Este capítulo é voltado à apresentação das investigações realizadas sobre a genealogia dos indivíduos componentes das populações geradas pelo modelo estudado nos capítulos anteriores. Como já foi exposto na introdução, a motivação do estudo da genealogia de uma população ou de um grupo (amostra) de indivíduos é que a história evolutiva (padrões de crescimento demográfico, extinções, migrações, etc ...) deixa marcas na composição genética da população e, portanto, essa história poderia em princípio ser recuperada através do estudo da genealogia da população. No entanto, a genealogia de uma população natural não é diretamente acessível, mas pode ser inferida dos padrões de similaridade e diferenças observadas quando comparamos, duas a duas, as seqüências dos indivíduos (DNA) de uma amostragem da população.

No modelo considerado, quando comparamos as seqüências de dois indivíduos da população, observamos que elas possuem todos os sítios originais (obviamente) iguais, alguns sítios mutantes em comum e alguns sítios mutantes distintos. Seguindo-se a linhagem dos dois indivíduos amostrados no passado, devido à finitude da população, encontraremos seu último ancestral comum, e nele suas mutações em comum. As mutações diferentes são aquelas que ocorreram entre a geração em que existiu esse último ancestral comum e a geração atual. Podemos estender esse raciocínio para uma amostragem¹ de qualquer

¹conjunto de indivíduos escolhidos aleatoriamente da população

tamanho, encontrando o último ancestral comum da amostragem e, em alguns casos, de toda a população (como o indivíduo A da figura 4.5). A variabilidade genética observada nesses indivíduos deve-se às mutações que ocorreram entre a geração em que existiu o ancestral comum mais recente e a geração atual, pois as mutações que ocorreram antes ou foram extintas ou então foram fixadas.

O processo genealógico no caso de seleção neutra ($s = 0$) foi extensamente investigado nas décadas de 70 e 80 sendo formalizado de forma matematicamente rigorosa pela teoria de coalescência de Kingman em 1982 [3] (como revisão [5, 25]). Neste caso os processos genealógicos e mutacionais são independentes, pois as mutações não acarretam variação na adaptação dos indivíduos [5], porém quando a população está submetida à seleção, a probabilidade de reprodução de um dado indivíduo, e portanto da continuidade da sua linhagem genealógica, depende das mutações que ele carrega. Neste capítulo vamos apresentar resultados ainda preliminares de uma análise do efeito da seletividade do meio, modelada pelos parâmetros s e α , na genealogia e diversidade da população. Para isso vamos utilizar os conceitos de *tempo de coalescência* e *distância de Hamming*. Coalescência é o evento de fusão da linhagem de dois indivíduos, e o tempo de coalescência T é o tempo decorrido entre a geração atual até o evento de coalescência. A distância de Hamming d mede a distância genética entre as seqüências de dois indivíduos e é definida como o número mínimo de substituições para que duas seqüências sejam iguais.

5.1 Alguns resultados analíticos simples

Nessa seção vamos determinar algumas propriedades médias da genealogia de uma amostra de n indivíduos de uma população de tamanho $N \gg n$ no regime neutro $s = 0$. Vamos desconsiderar os eventos de coalescência de três ou mais linhagens, pois esses são eventos que ocorrem com probabilidade (frequência) da ordem de $1/N^2$, enquanto que a coalescência de duas linhagens envolve probabilidades da ordem de $1/N$ apenas.

A probabilidade de que dois indivíduos escolhidos aleatoriamente na população coalesçam, isto é, tenham um ancestral comum na geração imediatamente anterior é simplesmente $1/N$. Para uma amostra de tamanho n o número de pares diferentes de indivíduos que podemos formar é $\binom{n}{2}$, sendo que cada par tem probabilidade $1/N$ de coalescer na geração precedente. Portanto a probabilidade que haja uma coalescência na geração anterior é $\frac{1}{N} \binom{n}{2}$, e o tempo médio de coalescência T_n de um par de indivíduos de uma amostra de tamanho n é dado pelo inverso dessa expressão

$$T_n = \frac{2N}{n(n-1)}. \quad (5.1)$$

Com esse resultado podemos determinar o tempo total médio de coalescência fazendo a somatória $\sum_{i=2}^n T_i$. A figura 5.1 mostra duas genealogias possíveis para uma amostra de tamanho $n = 5$ tirada de uma população de tamanho N e com tempos de coalescência dados pela equação (5.1). Fixados os tempos de coalescência T_n para uma amostra de

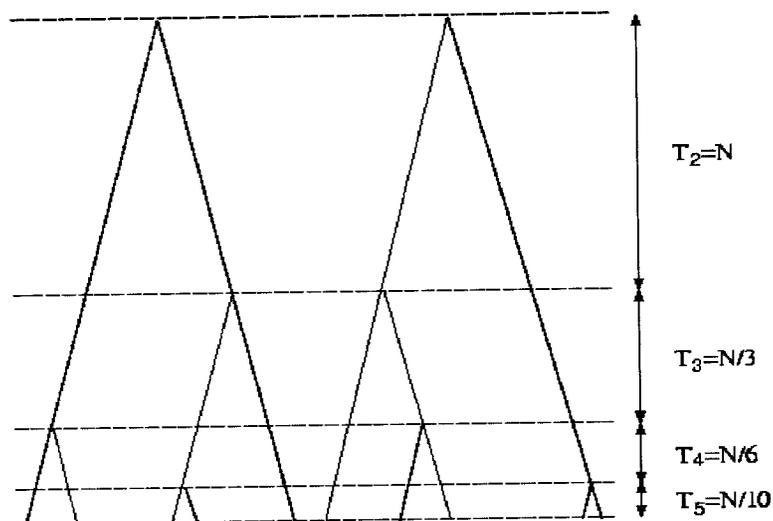


Figura 5.1: Duas genealogias possíveis para uma amostra de 5 indivíduos escolhidos de uma população de tamanho $N \gg 5$, com tempos de coalescência médios dados pela equação (5.1). O número total de genealogias possíveis é dado pela equação (5.2) com $n = 5$.

tamanho n , podemos facilmente calcular o número de genealogias possíveis, denotado por

G_n . De fato, usando a relação de recorrência $G_n = \frac{n(n-1)}{2}G_{n-1}$ obtemos

$$G_n = \frac{n!(n-1)!}{2^{n-1}}. \quad (5.2)$$

Naturalmente, na falta de informações adicionais, todas essas genealogias devem ser consideradas como igualmente prováveis.

No restante dessa seção vamos considerar amostras de tamanho $n = 2$ somente. Seja $\overline{P(T)}$ a probabilidade de que dois indivíduos escolhidos aleatoriamente dentro de uma população tenham tempo de coalescência T . A probabilidade de que estes dois indivíduos não tenham ancestral comum a t gerações no passado é simplesmente $(1 - \frac{1}{N})^t \approx e^{-t/N}$, onde a aproximação é válida para N grande. Portanto, a probabilidade de que o último ancestral comum tenha vivido exatamente a T gerações no passado é

$$\overline{P(T)} = [1 - e^{-(T+1)/N}] - [1 - e^{-T/N}] \approx \frac{e^{-T/N}}{N}, \quad (5.3)$$

de onde obtemos que $\langle T \rangle = N$ em concordância com o resultado obtido anteriormente. Alertamos que no regime neutro a genealogia da amostra não depende do processo de mutação, pois as mutações não modificam a adaptação dos indivíduos, e a equação (5.3) é válida para qualquer taxa de mutação U .

Agora vamos calcular a distribuição de probabilidade da distância de Hamming entre dois indivíduos amostrados da população. Vamos supor que durante cada intervalo dt cada indivíduo tenha probabilidade Udt de receber novas mutações. Portanto, a probabilidade $\phi_n(t)$ de que as seqüências de um indivíduo na geração atual e a de seu ancestral que viveu à t gerações no passado tenham distância de Hamming n obedece a equação

$$\frac{d\phi_n(t)}{dt} = U[\phi_{n-1}(t) - \phi_n(t)], \quad (5.4)$$

²Esta formulação de tempo contínuo produz os mesmos resultados que o modelo de tempo discreto no limite de N grande, pois a escala de tempo é da ordem de N gerações, que é muito maior que o intervalo entre gerações discretas $\Delta t = 1$.

cuja solução é uma distribuição de Poisson

$$\phi_n(t) = e^{-Ut} \frac{(Ut)^n}{n!}. \quad (5.5)$$

Portanto, a distância média entre um indivíduo e seu ancestral aumenta linearmente com o tempo $\overline{\langle n \rangle} = Ut$ (catraca de Müller no regime neutro).

Seja \overline{W}_n a probabilidade de que dois indivíduos sorteados aleatoriamente na geração atual tenham distância de Hamming n . As mutações distintas entre dois indivíduos são aquelas que ocorreram entre a geração atual e a geração em que viveu seu último ancestral comum. Portanto podemos escrever

$$\overline{W}_n = \int_0^\infty dT P(T) \phi_n(2T), \quad (5.6)$$

onde o fator 2 se deve às duas linhagens, uma para cada indivíduo. Substituindo as equações (5.3) e (5.5) em (5.6) obtemos que

$$\overline{W}_n = \frac{\lambda}{(\lambda + 1)^{n+1}}, \quad (5.7)$$

onde $\lambda = 1/2UN$. A distância de Hamming média é então $\overline{\langle d \rangle} = 1/\lambda = 2UN$.

Outro regime que permite a obtenção de alguns resultados analíticos é o regime de seleção extrema ($s \rightarrow 1$). Neste caso, apenas os indivíduos com o menor número de mutações na população, $k = k_{min}$, conseguem se reproduzir. Neste sentido, a situação é similar à do regime neutro, com N , o tamanho da população, substituído por N_0 o número de indivíduos da população atual com $k = k_{min}$, ou seja, o número de indivíduos na classe mais adaptada. Daí $\overline{\langle T \rangle} = N_0$. Como logo após o processo de reprodução toda a população passa a ser composta desses indivíduos e lembrando que no modelo o processo de mutação atua após o da reprodução, podemos escrever $N_0 \approx \overline{N}_0 = Ne^{-U}$ onde e^{-U} é a probabilidade de que um indivíduo não sofra mutação. Por outro lado, lembrando que o número médio de mutações em um indivíduo é U , e que as mutações ocorrem sempre em

sítios distintos, temos que a distância de Hamming média entre dois indivíduos no regime de seleção extrema é simplesmente $\overline{\langle d \rangle} = 2U$.

5.2 Alguns Resultados Numéricos

Nesta seção apresentamos os resultados obtidos através da simulação do modelo estocástico. O procedimento é deixar a população inicial, composta apenas por indivíduos sem mutações, evoluir até tempos t tipicamente da ordem de $10^4 - 10^5$, armazenando as informações que permitam recuperar a genealogia de cada indivíduo na população. As figuras 5.2 e 5.3 mostram os tempos de coalescência médios dividido pelo tamanho da população $\overline{\langle T \rangle}/N$ (à esquerda) e as distâncias de Hamming médias $\overline{\langle d \rangle}$ (à direita) como funções da taxa de mutação U . Para populações com $\alpha \geq 1.0$ (ausência de epistase e epistase sinérgica) essas grandezas convergem de forma relativamente rápida (isto é, não dependem muito do tempo de evolução t) para os resultados mostrados na figura 5.2. A causa dessa rapidez é que nesses casos as populações, após um tempo de evolução suficientemente grande, permanecem com o número de mutações médio inalterado durante um grande período de tempo, como podemos observar nos gráficos para $\alpha \geq 1.0$ das figuras 4.6, 4.7 e 4.8 do capítulo anterior. Encontramos, como esperado, $\overline{\langle T \rangle} = N$ e $\overline{\langle d \rangle} = 2NU$ no regime neutro ($s = 0$). Para $s > 0$, observamos que $\overline{\langle T \rangle} = N$ e $\overline{\langle d \rangle} = 0$ para $U = 0$, pois não há mutações que distingüam os indivíduos. Ainda, $\overline{\langle T \rangle}/N$ é sempre menor que 1 e decai rapidamente com U . Isso ocorre porque, com o aumento da taxa de mutação, a variância do número de indivíduos nas classes aumenta, e a seleção entre esses indivíduos com diferentes números de mutações faz diminuir o número de indivíduos que efetivamente participam do processo reprodutivo, diminuindo assim o tempo de coalescência médio. Essa análise é corroborada pelos gráficos de dispersão relativa da adaptação r_f contra U no limite determinístico, apresentados nas figuras 3.2, 3.3, 3.4, 3.5 e 3.6 do capítulo 3. Observamos naqueles gráficos que, para $\alpha \geq 1.0$, r_f sempre

aumenta com U , ou seja, o processo de reprodução torna-se cada vez mais seletivo.

Um fenômeno interessante que observamos do gráfico $\overline{\langle T \rangle} / N$ contra U é que para U pequeno (entre 0 e 1.5) e fixo, são os valores intermediários do coeficiente s que apresentam os menores tempos médios de coalescência. Esse comportamento pode ser explicado da seguinte forma: com o aumento de s a partir de 0, a diferença de adaptação entre indivíduos de classes diferentes também aumenta, fazendo com que um número cada vez menor de indivíduos participem do processo reprodutivo, diminuindo portanto o tempo médio de coalescência da população. Quando o coeficiente s é excessivamente grande, para taxas de mutação relativamente pequenas, a maior parte da população é constrangida a permanecer nas classes mais adaptadas, aumentando o número de indivíduos que participam do processo reprodutivo e, conseqüentemente, o tempo médio de coalescência entre eles. Com o aumento da taxa de mutação, a variância do número de indivíduos nas classes aumenta e o número de indivíduos nas classes mais adaptadas diminui, fazendo com que o tempo médio de coalescência diminua com o aumento do coeficiente seletivo. O efeito descrito acima não pode ser explicado baseando-se no comportamento observado no limite determinístico, como fizemos no parágrafo anterior, pois naqueles gráficos observamos que para $\alpha \geq 1.0$ a dispersão relativa de adaptações r_f sempre aumenta com s , e portanto esperaríamos tempos de coalescência menores quanto maior fosse o coeficiente seletivo. Isso de fato ocorre para valores suficientemente grandes de U . Ainda para $\alpha \geq 1.0$ (figura 5.2), vemos que a distância de Hamming aumenta com a taxa de mutação, mesmo com a diminuição do tempo médio de coalescência.

Diferentemente do observado para $\alpha \geq 1.0$, os tempos de coalescência e a distância de Hamming para os casos de epistase atenuante ($\alpha < 1.0$), com s e U mantidos fixos, aumentam com o tempo de simulação t e, como podemos observar dos gráficos da figura 5.3, parecem tender para uma situação que corresponde ao regime neutro: $\overline{\langle T \rangle} / N \rightarrow 1$ e $\overline{\langle d \rangle} \rightarrow 2UN$, onde $2UN$ é a distância de Hamming média no caso neutro, e é repre-

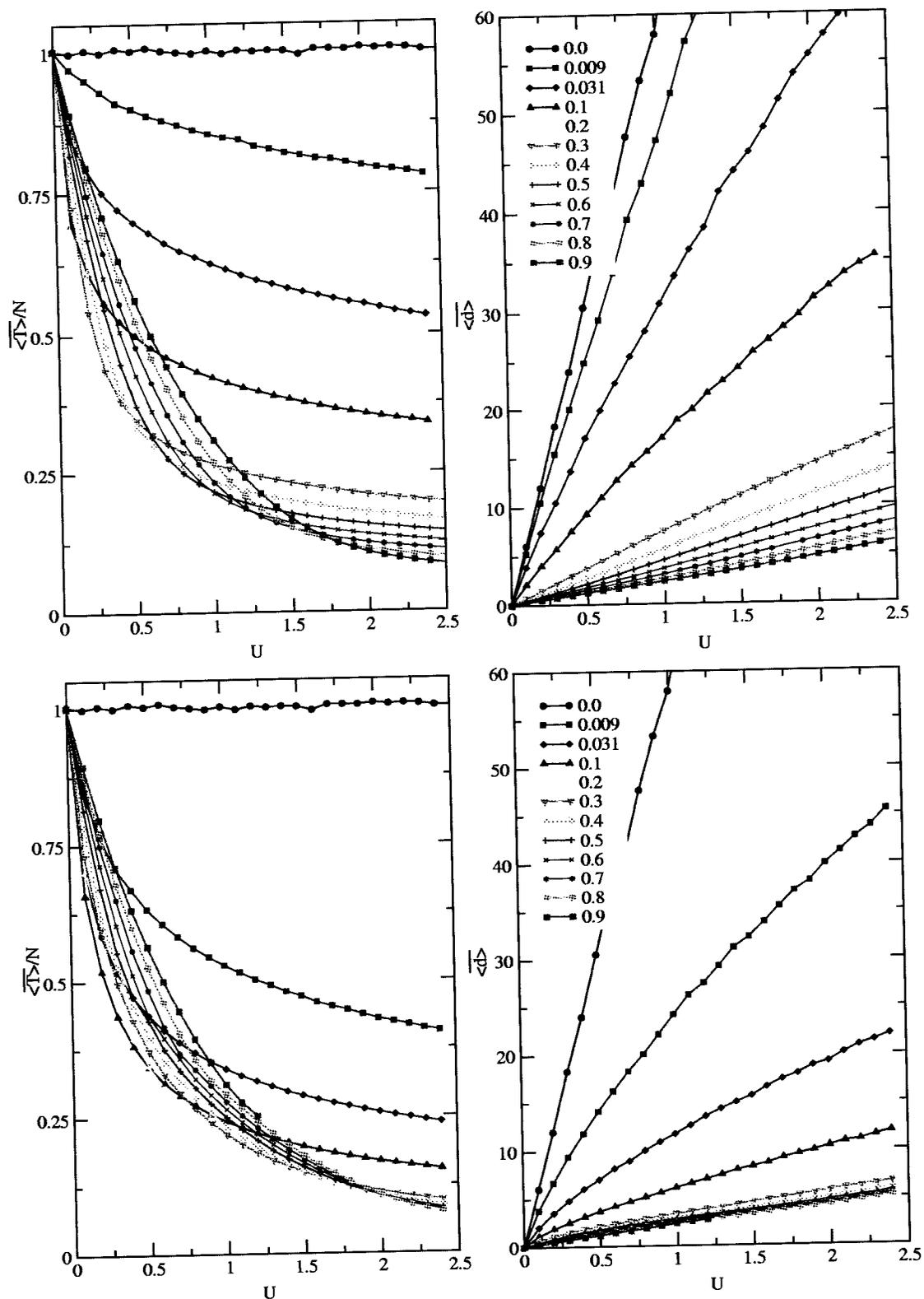


Figura 5.2: Gráficos superiores: coeficiente de epistase $\alpha = 1.0$. Gráficos inferiores: $\alpha = 1.2$. Gráficos à esquerda: tempo médio de coalescência dividido pelo tamanho da população $\langle T \rangle / N$ contra taxa de mutação U . Gráficos à direita: distância de Hamming média $\langle d \rangle$ contra U . Foram simuladas 2000 populações de tamanho $N = 30$, para valores de coeficiente seletivo s variando de 0.0 à 0.9. O tempo de simulação t empregado foi da ordem de 10^4 .

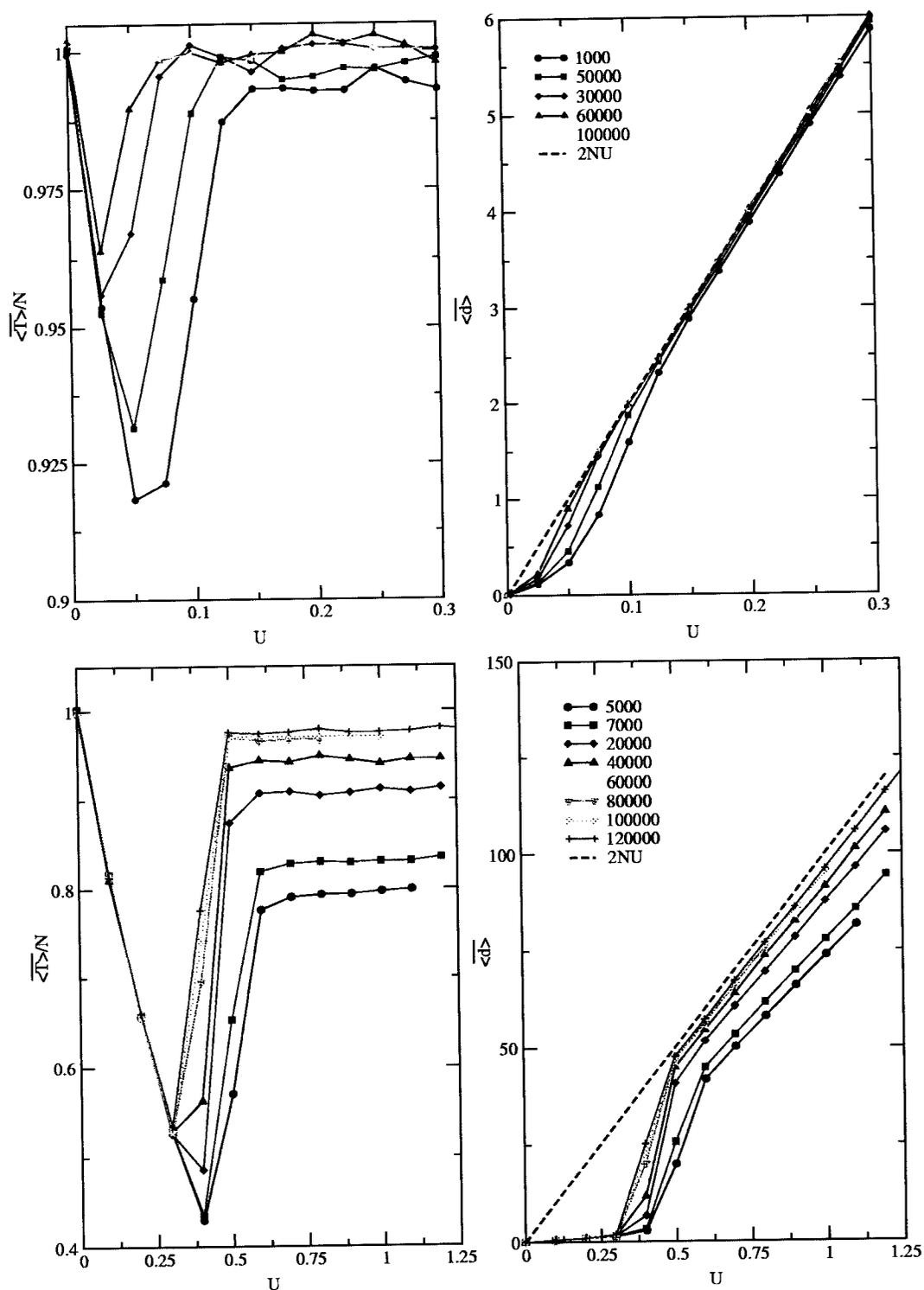


Figura 5.3: Gráficos superiores: média de 2000 populações de tamanho $N = 10$. Gráficos inferiores: média de 2000 populações de tamanho $N = 50$. Gráficos à esquerda: tempo médio de coalescência dividido pelo tamanho da população $\langle T \rangle / N$ contra a taxa de mutação U . Gráficos à direita: distância de Hamming média $\langle d \rangle$ contra U . Em todos os casos foi utilizado $\alpha = 0.5$ e $s = 0.5$. Cada curva mostra o resultado para diferentes tempos de simulação indicados na legenda.

sentada nos gráficos à direita da figura 5.3 com uma linha tracejada. Cada ponto destes gráficos foi obtido simulando o modelo até a geração indicada pela legenda e mediando o tempo de coalescência e distância de Hamming nas 1000 gerações seguintes.

Para taxas de mutação U suficientemente pequenas, independentemente do tempo de simulação t , o tempo médio de coalescência cai rapidamente com U , enquanto a distância de Hamming média cresce a uma taxa muito reduzida. Já para taxas de mutação maiores, o tempo médio de coalescência aumenta e, com um valor suficientemente alto do tempo de simulação t , a população atinge um regime semelhante ao da seleção neutra, no qual o tempo médio de coalescência independe de U e a distância de Hamming aumenta linearmente com U . Para explicar esse comportamento apresentamos, na figura 5.4, o gráfico da dispersão relativa da adaptação r_f contra a taxa de mutação U , onde cada curva representa um tempo de simulação. Podemos observar que, para um tempo de simulação fixo, r_f atinge um valor máximo quando o tempo de coalescência é mínimo, ou seja, o menor tempo de coalescência corresponde à maior diversidade de adaptação relativa. Para uma taxa de mutação U suficientemente grande, r_f torna-se pequeno, indicando um cenário *quase* neutro, e praticamente independente de U . Esse raciocínio também ganha suporte quando comparamos com os gráficos de dispersão r_f contra U das figuras 3.2, 3.3, 3.4, 3.5 e 3.6 do capítulo 4 referentes ao limite determinístico; porém naquele caso é necessário uma taxa de mutação muito grande, quando comparada com as observadas aqui, para que a dispersão comece a diminuir.

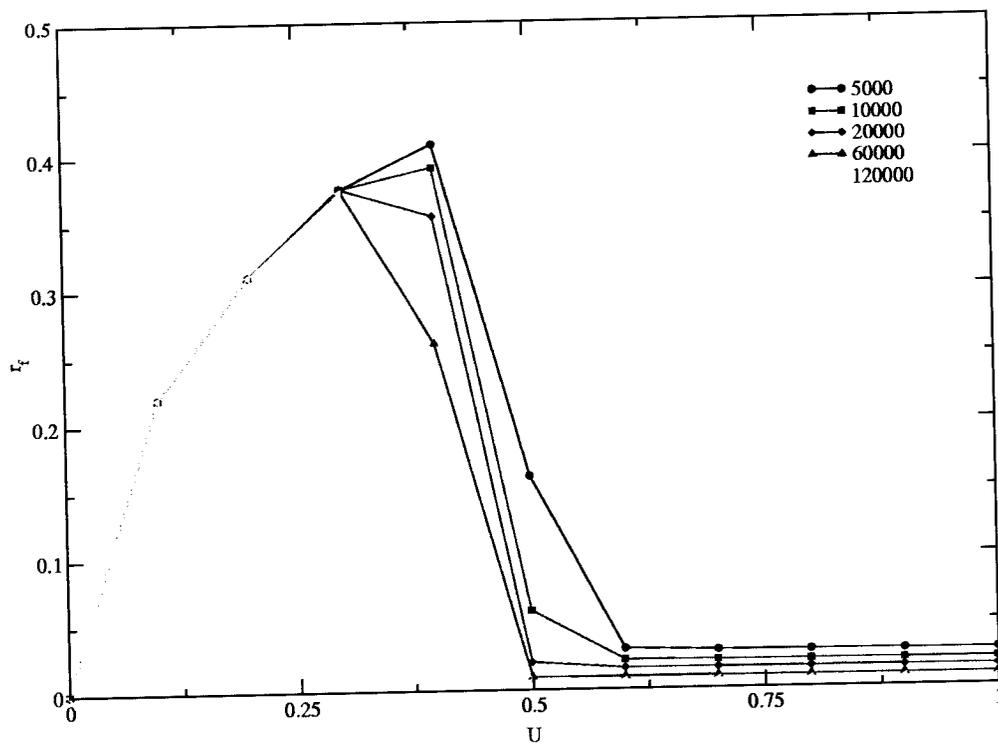


Figura 5.4: Gráfico da dispersão relativa da adaptação r_f contra taxa de mutação U . Cada curva representa um tempo de simulação diferente. Foram realizadas médias em 300 populações de tamanho 50, com coeficiente de epistase $\alpha = 0.5$ e de seleção $s = 0.5$.

Capítulo 6

Conclusão

Para finalizar esta dissertação, nesta seção vamos apresentar os principais resultados e conclusões obtidos nos capítulos anteriores. No capítulo 3, sobre o limite determinístico do modelo, foi apresentado um método de obtenção das frequências das classes de indivíduos no equilíbrio, mais rápido e preciso e, em termos de aplicabilidade, tão geral quanto o usualmente encontrado na literatura. Dos gráficos de número médio $\overline{\langle k \rangle}$ e variância $\overline{\sigma^2(k)}$ no número de mutações nas figuras 3.2, 3.3, 3.4, 3.5 e 3.6 vemos que, como esperado, com o aumento do coeficiente seletivo, a população fica cada vez mais independente do parâmetro de epistase α , porém a mesma afirmação não pode ser feita quando observamos os gráficos de número médio de mutações $\overline{\langle k \rangle}$ contra geração t no caso de população finita (capítulo 4, figuras 4.6, 4.7 e 4.8). Vimos no capítulo 4 que para epistase sinérgica ($\alpha > 1$) a população mantém-se muito tempo com um mesmo número médio de mutações. Isso pode ter importância para a dinâmica de mutações benéficas na população (não considerada nesse trabalho). Mesmo supondo uma taxa de mutações benéficas muito pequena, existe uma probabilidade razoável de que tais mutações ocorram nesse grande período de tempo, aumentando a adaptação média da população. Outro processo que poderia levar a um aumento na adaptação média da população é a recombinação. Por exemplo, o pesquisador Li Chao descreve tal processo em um modelo que leva em conta a troca de material genético entre dois vírus de RNA que infectam uma mesma célula

hospedeira [29]. Esses argumentos envolvendo epistase mostram que a discussão sobre a catraca de Müller como a causa única da vantagem seletiva da reprodução sexuada ainda não está totalmente esclarecida. Do estudo da genealogia de populações com epistase sinérgica (capítulo 5), no gráfico de tempo médio de coalescência $\langle T \rangle$ contra taxa de mutação U na figura 5.2, concluímos que, para valores suficientemente pequenos de U , são os valores intermediários do coeficiente seletivo s que resultam nos menores tempos médios de coalescência.

Ressaltamos que quase não existem investigações teóricas sobre epistase atenuante, $\alpha < 1.0$ (a exceção de [28]), de forma que os resultados apresentados nesta dissertação são inéditos. Com esse tipo de epistase, vimos no capítulo 4 que as mutações acumulam-se linearmente no tempo, porém a uma velocidade menor que a da seleção neutra, e que com a diminuição de α , ou seja, com a epistase cada vez mais atenuante, o comportamento do modelo aproxima-se do observado no limite neutro. Esperava-se que a epistase atenuante, após o acúmulo de muitas mutações, levasse a população ao regime de seleção neutra, identificado pela velocidade de acumulação de mutações igual a do regime neutro. Essa expectativa foi confirmada pelo estudo da genealogia da população, no gráfico de tempo médio de coalescência $\langle T \rangle$ contra taxa de mutação U apresentado na figura 5.3, com $\alpha < 1.0$ fixo e variando o tempo de simulação; porém o mesmo não foi observado nos análise da catraca de Müller, sendo necessário ainda mais investigações para desvendar essa questão. O comportamento do tempo de coalescência médio $\langle T \rangle$ contra taxa de mutação U para um tempo de simulação fixo, mostra aparecimento de um regime semelhante à seleção neutra ($\langle T \rangle$ independente de U) para U suficientemente grande, que é muito intrigante. Embora tenhamos tentado relacionar esse comportamento com a dispersão relativa da adaptação r_f (figura 5.4), achamos que é necessário uma explicação mais satisfatória. É interessante notar que no gráfico da velocidade da catraca R contra taxa de mutação U (figura 4.10) também vemos que, para $\alpha < 1.0$ fixo e U sufi-

cientemente grande, a população comporta-se de forma semelhante ao regime de seleção neutra, com R aumentando linearmente com U . Desses resultados concluímos que com a taxa de mutação suficientemente grande consegue-se compensar a vantagem seletiva dos indivíduos mais aptos.

Para dar continuidade à pesquisa iniciada nesta dissertação, além das linhas indicadas nos parágrafos anteriores, pode-se investigar mais detalhadamente a dependência da velocidade de catraca e da genealogia com o tamanho da população. Também é importante estudar o modelo utilizando seqüências de tamanho finito, possibilitando a ocorrência de mutações reversas. Várias modificações no modelo podem ser realizadas para estudar o efeito de outros processos. Por exemplo, é interessante conhecer a influência da recombinação na acumulação de mutações deletérias. Em particular, nos casos em que as mutações acumulam-se linearmente, saber o quanto de recombinação é o suficiente para parar a catraca. Porém, deve-se mencionar que o estudo da genealogia da população será dificultado, pois o número de ancestrais de uma seqüência não é fixo no caso de recombinação [30, 31]. Outras modificações seriam interessantes para o estudo da genealogia do modelo. Por exemplo, poderia-se levar em conta a distribuição espacial dos indivíduos, com a população dividida em grupos e a seleção atuando dentro destes grupos. Para esse modelo teríamos que supor uma taxa de migração dos indivíduos de um grupo para outro. Podemos ainda fazer com que o tamanho da população varie de forma pré-determinada, por exemplo, exponencialmente, e verificar o efeito na genealogia. Neste sentido, uma genealogia interessante de se investigar seria a do modelo predador-presa estocástico, no qual as flutuações no tamanho da população de cada uma das espécies ocorre de forma natural [32, 33].

Bibliografia

- [1] R. E. Lenski e M. Travisano. *Dynamics of adaptation and diversification: A 10000-generation experiment with bacterial population* Proc. Natl. Acad. Sci. USA, **91**, 6808 (1994)
- [2] A. Moya, S. F. Elena, A. Brach, R. Miralles e E. Barrio. *The evolution of RNA virus: A population genetics view* Proc. Nat. Acad. Sci. U. S. A., **97**, 6967 (2000)
- [3] J. F. C. Kingman. *On the genealogy of large populations* J. Appl. Probab. A, **19**, 27 (1982)
- [4] L. Excoffier. *Ce que nous dit la généalogie des genes* La Recherche. **302**, 82 (1997)
- [5] R. R. Hudson. *Gene genealogies and the coalescent process* Oxford surveys in evolutionary biology V7, (Oxford University Press, 1990)
- [6] M. Pagel. *Inferring the historical patterns of biological evolution* Nature. **401**, 877 (1999)
- [7] H. J. Müller. *The relation of recombination to mutational advance* Mutat. Res., **1**, 2 (1964)
- [8] E. A. Duarte, I. S. Novella, S. C. Weaver, E. Domingo, S. Wainhobson, D. K. Clarke, A. Moya, S. F. Elena, J. C. Delatorre e J. J. Holland. *RNA virus quasi-species - Significance for viral disease and epidemiology* Infect. Agent. Dis., **3**, 201 (1994)
- [9] I. Gordo e B. Charlesworth. *The degeneration of asexual populations and the speed of Müller's ratchet* Genetics, **154**, 1379 (2000)
- [10] H. Bernstein, H. C. Byerly, F. A. Hopf, R. A. Michod and G. K. Vemulapalli. *The darwinian dynamic* Quart. Rev. Biol., **58**, 185 (1983)
- [11] M. Kimura. *Theoretical foundation of population genetics at the molecular level* Theor. Popul. Biol., **2**, 174 (1971)

- [12] M. de La Peña, S. F. Elena e A. Moya. *Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2* Evolution, **54**, 686 (2000)
- [13] P. D. Keightley e A. Eyre-Walker. *Deleterious mutations and the evolution of sex* Science, **290**, 331 (2000)
- [14] J. F. Crow e M. Kimura. *An introduction to population genetics theory*. (Harper and Row, New York, 1970)
- [15] D. L. Hartl e A. G. Clark. *Principles of Population Genetics* (Sinauer Associates, Sunderland, 1989)
- [16] J. F. Crow. *Genetic loads and the cost of natural selection* Mathematical topics in population genetics, (Springer, Heidelberg, 1970)
- [17] A. S. Kondrashov. *Müller's ratchet under epistatic selection* Genetics, **136**, 1469 (1994)
- [18] M. Lynch, R. Bürger, D. Butcher, W. Gabriel. *The mutation meltdown in asexual populations* Jour. of heredity, **84**, 339 (1993)
- [19] G. C. Willians. *Sex and Evolution* (Princeton University Press, Princeton, 1975)
- [20] J. Maynard Smith. *The Evolution of Sex* (Cambridge University Press, Cambridge, 1978)
- [21] S. A. Kauffman e W. G. Macready, *Search strategies for molecular evolution* J. Theor. Biol., **4**, 427 (1995)
- [22] M. Eigen, J. McCaskill e P. Shuster. *The molecular quasi- species* Adv. Chem. Phys., **75**, 149 (1989)
- [23] P. G. Higgs e G. Woodcock. *The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection* J. Math. Biol., **33**, 677 (1995)
- [24] D. Alvez. *Análise estatística da teoria de quase-espécies de evolução molecular* Tese de doutoramento em física básica apresentada no Instituto de Física de São Carlos, USP, (1999)
- [25] W. Li e D. Graur. *Fundamentals of molecular evolution* (Sinauer Associates, Inc. Publisers, Massachusetts, 1991)
- [26] J. Haigh. *The accumulation of deleterious genes in a population - Müller's ratchet* Theor. Popul. Biol., **14**, 251 (1978)

- [27] F. Monteiro e F. Morán. *Biofísica - Procesos de autoorganización en biología* (Eudema Universidad, Madrid, 1992)
- [28] T. Wiehe. *Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models* Genet. Res. Camb., **69**, 137 (1997)
- [29] L. Chao. *Evolution of sex in RNA viruses* Jour. Theor. Biol., **133**, 99 (1988)
- [30] B. Derrida e B. Jung-Muller. *The Genealogical tree of a chromosome* Jour. Stat. Phys., **94**, 277 (1999)
- [31] C. Wiuf e J. Hein. *On the number of ancestor to a DNA sequence* Genetics, **147**, 1459 (1997)
- [32] D. T. Gillespie. *Exact stochastic simulation of coupled chemical-reaction* Jour. Phys. Chem., **81**, 2340 (1977)
- [33] D. T. Gillespie. *General method for numerically simulating stochastic time evolution of coupled chemical-reactions* Jour. Comp. Phys., **22**, 403 (1976)