Acúmulo de mutações em linhagens assexuadas: uma abordagem via experimentos computacionais

Alexandre Colato

USP/IFSC/SBI

8-2-001633

Tese apresentada ao Instituto de Física de São Carlos, da Universidade de São Paulo, para obtenção do título de Doutor em Ciências: Física Básica.

Orientador: Prof. Dr. José Fernando Fontanari

São Carlos – 2004

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

IFSC - SBI
CLASS.....
CUTTER.....
TOMBO.....

Colato, Alexandre

"Acúmulo de mutações em linhagens assexuadas: uma abordagem via experimentos computacionais"
Alexandre Colato – São Carlos, 2004

Tese (Doutorado) – Área de Física da Universidade de São Paulo, 2004 - Páginas: 122

Orientador: Prof. Dr. José Fernando Fontanari

1. Física Biológica.
2. Catraca de Muller.
3. Presa-Predador.

I. Título

UNIVERSIDADE DE SÃO PAULO
Instituto de Física de São Carlos

Av. Trabalhador São-carlense, 400
Caixa Postal 369
13560-970 - São Carlos - SP - Brasil

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE ALEXANDRE COLATO APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, UNIVERSIDADE DE SÃO PAULO, EM 18-11-2004.

COMISSÃO JULGADORA:

Prof. Dr. José Fernando Fontanari (Orientador e Presidente) – IFSC/USP

Prof. Dr. José Antônio Salvador – UFSCar

Profa. Dra. Joyce da Silva Bevilacqua – IME/USP

Dr. Luis Aníbal Diambra – IFSC/USP

Prof. Dr. Otávio Henrique Thiemann – IFSC/USP

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

NÃO ENTRE EM PÂNICO.

"Frase escrita em letras garrafais e amigáveis na capa de um dos livros mais famosos do Universo,

o Guia do Mochileiro da Galáxias."

Do livro o Guia do Mochileiro das Galáxias de Douglas Adams

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

Agradecimentos

Mais uma vez chego ao final de uma etapa em minha vida, e felizmente, com muitas pessoas ao meu redor que direta ou indiretamente me ajudaram nesses quatro anos de doutorado.

Sem dúvida nenhuma a principal pessoa a quem devo agradecer é o Prof. Fontanari, que acreditou em mim quando eu ainda fazia mestrado na Federal e me aceitou como seu aluno de doutorado, por seu suporte essencial para o desenvolvimento desta tese, e pelos conselhos sempre oportunos.

Gostaria de agradecer à colaboração do Prof. Stephen Howard pela ajuda durante os dois últimos anos de trabalho com idéias originais e com uma visão biológica dos processos de evolução. Do mesmo modo agradeço ao Dr. Philip Gerrish (Dr. Fili Gerri) pela orientação, amizade e bons papos no *Brickhouse* e em todo o tempo que estive no México. Além de ser brilhante e estar sempre contente ele me apresentou a pessoas, que sem sombra de dúvida, serão muito importantes para meu futuro acadêmico.

Um parágrafo especial deve ser dedicado a Claudia Pio, que se mostrou uma amiga inestimável ao me apresentar ao Phil e me ajudar em tudo para que minha passagem pelo México fosse a mais perfeita possível, a única coisa que tenho a dizer contra é que ela fez uma enorme propaganda para eu tomar uma *michelada*, mas aquilo é horrível.

Muitos foram os amigos que aturaram as minhas piadinhas e minhas manias, entre eles estão Giovano, Paulo (pac), Paulo (prac), Viviane, Antonios (Seridonio e Francisco), Petrus, Daniela, Danielle, Daniel, Fábio, Mickel, Milton, Louis (Fernando) e em especial ao Léo que me acompanha desde os tempos de Federal.

Muito obrigado a minha família, "Seu" Ary e "Dona" Denise que sempre me incentivaram em tudo e a um excelente amigo, que por coincidência é meu irmão, Tiago. Irmãos eu também encontrei em minha estada em São Carlos, são eles o Taixi (Daniel) e o Flents (José Varalda).

Nestes 11 anos que vivo em São Carlos, posso dizer a grande maioria deles foram especiais, isso porque tive ao meu lado minha namorada, Cristina, que nem sempre estava

fisicamente junto comigo, mas com certeza sempre esteve em meu coração.

Finalmente gostaria de agradecer ao apoio da FAPESP, com o suporte financeiro que possibilitou elaborar esta tese.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

Resumo

Estudos sobre evolução têm sido desenvolvidos desde a publicação dos trabalhos de Charles Darwin sobre a origem das espécies pela seleção natural em 1859. Durante o século XX grandes avanços foram obtidos com a utilização de modelagens matemáticas e computacionais, pois com exceção de algumas espécies que podem ter sua evolução analisada *in vivo*, o tempo necessário para aquisição de dados é enorme e por este motivo o enfoque computacional passou a representar uma ferramenta essencial.

Nesta tese são apresentados os conceitos básicos para se entender o processo evolutivo de populações assexuadas como mutação, seleção e relevos adaptativos, bem como os resultados numéricos sobre sua evolução através do processo conhecido como catraca de Muller, que baseia-se na perda estocástica da classe de indivíduos mais adaptados da população através das mutações adquiridas ao longo de sua linhagem.

Neste trabalho foram estudadas diversas dinâmicas, como a de populações que estão sujeitas à passagens seriais com gargalo, onde observamos que a velocidade da catraca não pára devido aos altos valores de epistase, enquanto que para populações com tamanho variável (crescimento e decrescimento exponencial) a catraca pára durante o período de crescimento até a população atingir o limite permitido pelo meio-ambiente, sendo que a partir deste ponto ela se comporta como no modelo de infinitos sítios tradicional. Por último, são apresentados os resultados de populações que interagem entre si em uma dinâmica presa-predador, onde o comportamento da catraca pode ser entendido com base nas dinâmicas das populações descritas anteriormente.

Um outro problema abordado nesta tese é o da utilização de medidas da topologia de árvores genealógicas para verificar a presença da seleção na evolução de uma população. Apesar dos comprimentos dos ramos das árvores apresentarem alterações quando comparados ao caso neutro, observamos que os testes estatísticos utilizados não são suficientes para inferir o efeito da seleção em populações reais.

Palavras Chaves: Evolução; Catraca de Muller; Populações assexuadas; Passagens seriais; Mutação; Seleção.

Abstract

Studies about evolution have been developed since Charles Darwin's publications about the Origin of species and Natural Selection in 1859. During the XX century major developments were achieved through mathematical and computational modeling, since only few number of species that their evolution can be studied *in vivo*, once that the time scale involed for data acquisition procedure is considerable, and for this reason the computational approach become an important tool in this study.

In this thesis are presented the basic concepts to understand the process of evolution in a population as mutation, selection and adaptive landscapes, in addition some numerical results about the evolution of an asexual population using the process known as Muller's ratchet, that can be characterized by the stochastic loss of the most fitted class of individuals through mutations that are acquired in their lineages.

During this work several dynamics were studied, likewise the populations under serial bottleneck passages, where we observed that the velocity of the ratchet never stops for high epistatic coefficients, while in population whose size can varies (increasing or decreasing exponentially) the ratchet halts during population's increasing until these individuals do not reach the maximum number permitted, and after this point this population behaves like the traditional infinite genome size model. At last, we show the results of populations that can interact between themselves in a predator-prey dynamics, where the behaviour of the ratchet can be understood in the previous dynamics .

Another problem that was studied in this thesis is related with several topology measures of genealogical trees in order to verify the selection in a population evolution. Despite branch's length of the trees changed due to the selection, we could see that the statistical tests used do not be sufficient to infer the effect of selection under real populations.

Conteúdo

1 Introdução **1**

1.1 Um pouco de História 1

1.2 Reprodução sexuada versus reprodução assexuada 5

1.3 Apresentação da tese 7

2 Modelagem de processos evolutivos **10**

2.1 Dinâmica evolutiva 11

2.1.1 Relevos de Replicação 13

2.2 Modelagem computacional 15

3 Evolução de populações assexuadas **19**

3.1 Catraca de Muller em passagens seriais com gargalo 19

3.2 Populações que apresentam tamanho variável 26

4 Evolução de populações assexuadas em um sistema de presa-predador **40**

4.1 O Modelo Presa-Predador 42

4.2 Resultados 48

5 Árvores genealógicas **61**

5.1 Definições básicas 61

5.2 Análise da genealogia utilizando distância de Hamming 65

5.3 Propriedades Estatísticas dos Ramos Internos e Ramos Externos. 72

5.3.1 Medidas da topologia das árvores 79

6 Conclusões **86**

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

A Alguns Resultados Analíticos para o Modelo de Evolução Neutra 91

B Determinação dos Comprimentos de Ramos Externos. 94

C Siglas e símbolos 97

Lista de Figuras

3.1 Passagem serial com gargalo. Neste experimento alguns indivíduos da população que estavam no primeiro tubo são escolhidos aleatoriamente para serem passados para um novo tubo dando origem a uma nova população. . 23

3.2 Comportamento da freqüência mínima C_m para diferentes valores de $\alpha \geq 1$, com $s = 0.5$ e $U = 0.5$. 24

3.3 Comportamento da freqüência mínima C_m para diferentes valores de $\alpha < 1$, com $s = 0.5$ e $U = 0.5$. 25

3.4 Tempo de perda de C_m para diferentes valores de $\alpha \geq 1$, com $s = 0.5$ e $U = 0.5$. 25

3.5 Tempo de perda de C_m para diferentes valores de $\alpha < 1$, com $s = 0.5$ e $U = 0.5$. 26

3.6 Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de mutação U. $s = 0$ e $R = 2$ com médias em 10^6 populações independentes. 30

3.7 Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes. 32

3.8 Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes. 32

3.9 Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes. 33

3.10 Gráfico do número médio de indivíduos na população versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes. 33

3.11 Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2$ com médias em 10^5 populações independentes. 34

3.12 Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2$ com médias em 10^5 populações independentes. Os $\triangle$ mostram o comportamento para $s[0, 0.1]$ enquanto que $\bigcirc$ mostram o comportamento para $s(0.1, 1]$. 34

3.13 Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2$ com médias em 10^5 populações independentes. 35

3.14 Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes. 35

3.15 Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes. 36

3.16 Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes. 36

3.17 Gráfico do número médio de indivíduos na população versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes. 37

3.18 Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes. 38

3.19 Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes. 38

3.20 Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes. 39

3.21 Gráfico do número médio de indivíduos na população versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes. 39

4.1 Comparação entre a evolução temporal para as populações de presas e predadores nos casos em que N_H e N_P assumem valores reais (curva vermelha) e valores inteiros (curva preta), para os parâmetros $N_H(t = 0) = 50$, $N_P(t = 0) = 15$, $r = 2$, $a = 0.05$ e $\kappa = 100$. 45

4.2 Comparação entre a evolução temporal para as populações de presas e predadores nos casos em que N_H e N_P assumem valores reais (curva vermelha) e valores inteiros (curva preta), para os parâmetros $N_H(t = 0) = 50$, $N_P(t = 0) = 15$, $r = 2.5$, $a = 0.05$ e $\kappa = 100$. 46

4.3 Comparação entre a evolução temporal para as populações de presas e predadores nos casos em que N_H e N_P assumem valores reais (curva vermelha) e valores inteiros (curva preta), para os parâmetros $N_H(t = 0) = 50$, $N_P(t = 0) = 15$, $r = 3$, $a = 0.05$ e $\kappa = 100$. 47

4.4 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 0$, $s^N = 0$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 51

4.5 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 0.1$, $s^N = 0.0125$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. . . 52

4.6 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 0.5$, $s^N = 0.0125$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. . . 53

4.7 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 1.0$, $s^N = 0.0125$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. . . 54

4.8 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 0$, $s^N = 0$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 55

4.9 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 0.1$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 56

4.10 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 0.5$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 57

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

4.11 Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 1.0$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 58

4.12 Comparação entre o número de mutações e o tamanho da população. $N(t = 0) = 50$, $U^N = 0.1$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 59

4.13 Comparação entre o número de mutações e o tamanho da população. $N(t = 0) = 50$, $U^N = 0.5$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 59

4.14 Comparação entre o número de mutações e o tamanho da população. $N(t = 0) = 50$, $U^N = 1.0$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes. 60

5.1 Exemplo de uma árvore genealógica e suas partes integrantes. Neste caso a evolução temporal se dá de cima para baixo, sendo os nós mais próximos do topo da árvore mais antigos que os nós que se encontram mais próximos de sua base. 62

5.2 Exemplos de árvores binárias e não-binárias. As figuras a) e b) representam árvores não-binárias, pois possuem um nó com mais de três ramos, enquanto a árvore c) é um exemplo de árvore binária. 63

5.3 Tipos diferentes de árvores. a) Cladograma. b) Aditiva. c) Ultramétrica. . 64

5.4 Árvore ultramétrica para uma amostra com 5 indivíduos. 64

5.5 Distâncias médias em relação à seqüência inicial para seqüências com L sítios. As linhas tracejadas representam seqüências quaternárias, enquanto as linhas cheias representam seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. 68

5.6 Tempo médio de coalescência $T^{\alpha\beta}$ para o modelo de L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$. 69

5.7 Tempo médio de coalescência $T^{\alpha\beta}$ para o modelo de L sítio sem seqüências quaternárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$. 69

5.8 Distância média entre dois indivíduos $d_{\alpha\beta}$ para o modelo L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$. 70

5.9 Distância entre dois indivíduos $d_{\alpha\beta}$ para o modelo de L sítios em seqüências quaternárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$. 70

5.10 Tempo médio de coalescência $T^{\alpha\beta}$ para o modelo de L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0.1$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$. 71

5.11 Distância média entre dois indivíduos $d_{\alpha\beta}$ para o modelo L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0.1$ e taxa de mutação $U = 0.1$. 71

5.12 Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de finitos sítios ($L = 50$) em seqüências binárias com relevo de pico único e tamanho da população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$. 75

5.13 Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo de pico único e tamanho de população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$. 75

5.14 Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de finitos sítios ($L = 50$) em seqüências binárias com relevo multiplicativo e tamanho da população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$. 76

5.15 Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo e tamanho de população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$. 76

5.16 Comprimento dos ramos externos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 e taxa de mutação $U = 1.0$. 77

5.17 Comprimento dos ramos internos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 e taxa de mutação $U = 1.0$. 78

5.18 Comprimento dos ramos externos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 em uma amostra de tamanho 20. 79

5.19 Comprimento dos ramos internos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 em uma amostra de tamanho 20. 80

5.20 *a*) Árvore completamente simétrica: Todos os indivíduos continuam gerando dois novos indivíduos ao longo do tempo. *b*) Árvore completamente assimétrica: Somente um dos dois descendentes de cada geração se reproduz gerando dois novos indivíduos. 80

5.21 Valor esperado de $\overline{N}$ em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. 83

5.22 Valor esperado de σ_N em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. A árvore mais simétrica possui $\sigma_N = 0$. 83

5.23 Valor esperado para a medida Colless em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. A árvore mais simétrica possui $C = 0$, enquanto a mais assimétrica possui $C = 1$. . . 84

5.24 Valor esperado de B_1 em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. 84

5.25 Valor esperado de B_2 em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. 85

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

Capítulo 1

Introdução

1.1 Um pouco de História

A teoria da origem da vida e da evolução das espécies sempre pertenceu ao imaginário do ser humano desde os tempos mais remotos. Durante a antigüidade, Aristóteles, sugeriu uma teoria que seria aceita sem nenhuma contestação por séculos, a teoria da geração espontânea [1]. Segundo ela os seres vivos poderiam ser criados a partir de materiais inanimados quando combinados de maneira adequada. Isto foi absorvido tão fortemente pela população em geral que até o século XVII existiam receitas para a produção de baratas, ratos entre outros animais. Uma das mais famosas receitas foi proposta por van Helmont, um dos precursores da Química e da Biologia científica, na qual camundongos poderiam ser criados da mistura de roupa suja e trigo.

Um dos primeiros pesquisadores a confrontar a teoria da geração espontânea foi Francesco Reidi que em 1688, realizou um experimento muito simples, onde ele cobriu um pedaço de carne com uma tela de malha muito fina, e após algum tempo em um exame minucioso do tecido verificou-se a existência de ovos de moscas, porém nada havia sido encontrado sobre a carne. Apesar deste experimento parecer conclusivo, ele não foi suficiente para derrubar a teoria da geração espontânea.

Após Reidi vários outros cientistas desenvolveram experimentos para corroborar ou não a teoria da geração espontânea, entre eles destacam-se, Anthony van Leeuwenhoek que

descreveu com detalhes os primeiros "animálculos" que mais tarde receberiam o nome de microorganismos. Em 1718, Louis Joblot realizou experimentos fervendo extratos de plantas e mostrando que estes extratos quando não eram expostos ao ar permaneciam estéreis, ou seja, não apresentavam nenhum microorganismo. Tentando derrubar as hipóteses de Joblot, John de Turbeville Needham, demonstrou que mesmo fervendo as soluções, estas apresentariam microorganismos, porém este trabalho foi logo criticado por Spallanzani, que realizou os mesmos experimentos, agora com maiores cuidados, e alegou que a geração de microorganismos só foi possível nos experimentos de Turbeville Needham devido a má esterilização dos recipientes utilizados.

Depois de vários experimentos e discussões ao longo do século XVIII, somente na segunda metade do século XIX três avanços científicos modificaram de maneira contundente essa visão de mundo e na frente desses avanços encontramos dois grandes nomes da ciência, Louis Pasteur e Charles Darwin e um monge, Gregor Mendel.

Louis Pasteur, um dos maiores experimentalistas de todos os tempos, demonstrou que o conceito de geração espontânea de seres vivos através da combinação de elementos inanimados não era plausível para seres em escala macroscópica. Em uma série de experimentos brilhantes e muito bem controlados realizados em 1862, Pasteur mostrou que a teoria da geração espontânea era impossível e que, todos os seres vivos tinham uma origem em gerações anteriores e não em materiais inanimadas.

Em seus experimentos Pasteur desenvolveu um sistema composto por tubos de vidros em um formato bem característico que recebeu o nome de "pescoço´´ de cisne. Inicialmente Pasteur esterilizou um conjunto de frascos e em seguida colocou uma solução que, se deixada exposta ao ar, apresentaria depois de algum tempo os microorganismos de van Leeuwenhoek. No gargalo destes frascos foi fixado o tubo em forma de pescoço de cisne, de maneira que uma das extremidades do tubo ficasse aberta e exposta ao ar. Depois do sistema pronto ele era aquecido lentamente por um longo período até que a solução de seu interior estivesse completamente esterilizada, e então todo esse sistema era colocado para esfriar. Embora o frasco estivesse em constante contato com o ar, a presença de microorganismos só foi observada em parte do tubo, sendo que toda a solução permanecia

estéril.

Após estes experimentos os defensores da teoria da geração espontânea tentaram de todas as formas encontrar meios para derrubar os resultados obtidos por Pasteur, porém sem maiores sucessos. Concomitantemente com a queda da teoria da geração espontânea, perguntas passaram a ser feitas, pois se todos os indivíduos são formados por outros indivíduos da mesma espécie que existiam antes, então como surgiram os primeiros seres vivos? Como pode ser possível a existência de tamanha diversidade de espécies?

Parte dessas perguntas foram respondidas por Charles Darwin [2–4] em sua teoria sobre a evolução das espécies. A base desta teoria está no processo de seleção natural que pode ser entendido da seguinte forma. Imagine uma espécie com um grande número de indivíduos, todos por serem da mesma espécie compartilham características semelhantes, porém não são exatamente iguais entre si. Essa pequena diferença entre os indivíduos fará com que os mais adaptados a um determinado ambiente se reproduzão mais, passando suas características para seus filhos, enquanto que os indivíduos menos adaptados produzirão menos descendentes, desta forma o meio-ambiente *selecionará* naturalmente os mais adaptados para a continuação da espécie.

Com base nesta teoria, podemos responder as perguntas feitas anteriormente, ou melhor, podemos afirmar que não existiu uma primeira geração de cada espécie. O que existiu na verdade foi uma evolução a partir de indivíduos mais simples até indivíduos mais complexos através do processo de seleção natural. Porém Darwin não conseguia explicar o mecanismo necessário para gerar essa diversidade. Na verdade, parte da resposta a esse problema foi obtida através dos trabalhos de Gregor Mendel que em 1866 estudou o cruzamento entre ervilhas e analisando os seus resultados deu os primeiros passos em direção à genética como ela é conhecida hoje. Porém os trabalhos de Mendel ficaram em segundo plano durante toda a segunda metade século XIX, sendo redescoberto no inicio do século XX.

O século XX foi muito importante não só no aspecto teórico mas também no caráter experimental da biologia com avanços proporcionados por pessoas das mais diferentes áreas como engenheiros, matemáticos, estatísticos, etc. Mais especificamente durante as

décadas de 20 e 30, os trabalhos sobre genética de populações de Sewall Wright, Ronald Fisher e John Burdon Sanderson Haldane foram de grande importância ao estabelecer uma base matemática e elaborar conceitos que são utilizados até hoje, como é o caso do modelo de Wright-Fisher, e que será apresentado nos próximos capítulos. Além de importantes contribuições eles formaram um grande número de pessoas que trabalharam ativamente na área da biologia evolutiva durante o século, sendo um principais nomes o de John Maynard Smith.

Durante a década de 60 outras contribuições deram um novo impulso à genética de populações com os trabalhos de Motoo Kimura e Hermann Joseph Muller. Kimura [5–8] apresentou um modelo de evolução onde as variações nos indivíduos de uma espécie ocorrem através de mutações seletivamente neutras ou que possuem um pequeno valor seletivo, sendo a fixação dessas mutações nas populações, e sua conseqüente evolução, dada pelo processo de deriva genética. Em 1969 King e Jukes [9] publicaram um artigo chamando este tipo de evolução de *não-darwiniana*.

Uma das maiores contribuições de H. J. Muller [10] para a genética de populações foi o conceito que é o tema central desta tese e que na década de 70 recebeu o nome de catraca de Muller em sua homenagem. Em artigos publicados na década de 60 Muller propôs um mecanismo no qual se poderia medir a evolução de uma população assexuada através da perda dos seus indivíduos mais adaptados. Essa medida da adaptação recebeu grande atenção da comunidade científica em geral sendo aplicada em estudos teóricos e em medidas experimentais.

Vários nomes foram importantes durante o século XX no avanço deste tema, nomes como John Maynard Smith, Lynn Margulis, Ernst Mayr, Stephen Jay Gold entre outros.

Porém todos esses desenvolvimentos geraram novos questionamentos, como por exemplo, se os seres evoluem de organismos mais simples, então quais foram os primeiros seres? Esta questão ainda encontra-se em aberto, onde vários cenários possíveis já foram construídos, sendo que um dos mais aceitos é o que identifica o RNA como sendo uma das primeiras moléculas na Terra [11, 12] que possuíam a capacidade de se reproduzir e sintetizar alimentos para a sua sobrevivência.

Uma das propostas de que o RNA foi capaz de se reproduzir e evoluir foi apresentada por Orgel [12] e é chamada de Mundo RNA (*RNA World*). Neste cenário o RNA teria todas as condições necessárias para sobreviver e se replicar, porém nos primeiros estágios de sua evolução ele teria tido apenas a capacidade de se duplicar, e somente mais tarde ele desenvolveu a capacidade de organizar aminoácidos para formar as proteínas, sendo que essas moléculas junto com o DNA e o RNA formam o tripé para a manutenção da vida como a conhecemos hoje.

1.2 Reprodução sexuada versus reprodução assexuada

Várias perguntas poderiam ser feitas no que concerne a escolha em estudar a reprodução assexuada ao invés de se estudar a reprodução sexuada, principalmente porque um dos principais argumentos em favor da reprodução sexuada é a grande variabilidade genética que este tipo de reprodução pode produzir em algumas poucas gerações. Na verdade existem inúmeras diferenças entre estes dois tipos de reprodução sendo este tema muito importante na biologia evolucionária como descreveu Graham Bell em seu famoso livro de 1982, *The Masterpiece of evolution* [13]

> "Sex is the queen of problems in evolutionary biology. Perhaps no other natural phenomenon has aroused so much interest; certainly none has sowed as much confusion..."
>
> Graham Bell

Como descrito por Vrijenhoek e colaboradores [14], a grande maioria dos vertebrados se reproduzem sexuadamente, onde de um total de 42300 espécies apenas 22 peixes, 23 anfíbios e 29 répteis se reproduzem assexuadamente. Pensando nesses números, poderíamos imaginar que a reprodução através do sexo é "melhor" quando comparada com a reprodução assexuada, porém isto não pode ser assumindo como completa verdade.

Uma das características mais importantes e citada como principal vantagem da reprodução sexuada, é a variabilidade genética, este ponto é sem dúvida muito importante

para a manutenção de uma espécie, principalmente em um meio que está em constante modificação, pois um determinado genótipo pode se destacar em relação aos demais e por seleção natural se sobrepor aos seus concorrentes, enquanto que uma população que se reproduz assexuadamente levaria mais tempo para produzir um indivíduo melhor adaptado ao meio. Porém os estudos nessa área costumam associar diferentes custos a um determinado tipo de reprodução [13, 15–18]. Podemos tomar como exemplo para a reprodução sexuada o custo associado à procura do parceiro, pois em várias espécies gástasse tempo e energia na busca pelo par, deixando os indivíduos mais suscetíveis a fatores negativos como a própria predação. Outro ponto desfavorável neste tipo de reprodução é que, em geral, nas espécies que apresentam os dois tipos de reprodução, a reprodução sexuada é mais lenta que a assexuada.

Dentre os fatores negativos da reprodução sexuada, o principal destaque é o duplo custo do sexo (*t*wofold cost of sex) [13, 15]*. O duplo custo do sexo pode ser entendido da seguinte forma: Vamos definir como sendo o casal a "unidade´´ de reprodução em uma espécie que se reproduz sexuadamente, pois são necessários dois indivíduos para gerar um descendente, enquanto que essa mesma unidade na reprodução assexuada é um único indivíduo. A menos que a reprodução sexuada produza o dobro de "filhos´´ que a reprodução assexuada, a taxa "per capita´´ de descendentes no primeiro tipo de reprodução será menor quando comparada ao segundo e, se considerarmos o caso em que cada unidade de reprodução produza o mesmo número de descendentes, os indivíduos da reprodução assexuada vão deixar duas vezes mais genes para a geração seguinte que a reprodução sexuada, vindo deste fato o duplo custo do sexo.

Da mesma forma que a reprodução sexuada possui problemas, a reprodução assexuada também possui, sendo o principal deles a degeneração do genoma pela ausência da recombinação. Sabe-se que a maioria das mutações que ocorrem em um genoma são deletérias [20–23] e apenas um pequeno número pode ser considerado como sendo de mutações benéficas [24] de maneira que ao longo do período evolutivo uma população assexuada tem uma redução em sua adaptabilidade devido ao acúmulo dessas mutações

*Também conhecido como duplo custo da meiose [19]

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

deletérias.

Dado a importância deste tema, várias hipóteses surgiram ao longo das últimas décadas para tentar explicar por que algumas espécies passaram a se reproduzir sexuadamente ao longo de sua história evolucionária enquanto outras ainda se reproduzem assexuadamente, entre as principais hipóteses estão a hipótese do melhor homem (*Best-man hypotesis*), a da rainha vermelha (*Red Queen hypotesis*), a da catraca de Muller (*Muller's Ratchet*) entre outras.

O processo da catraca foi descrito durante a década de 60 por Muller [10], onde este propôs que uma população assexuada de tamanho finito tende a adquirir mutações deletérias com uma freqüência muito maior que a de mutações benéficas em seu genoma, devido a erros ocorridos durante o processo de reprodução ao longo de sua linhagem e por isso com o passar do tempo essa população tende à extinção devido a degeneração de seu genoma, sendo a reprodução sexuada uma estratégia inteligente para esta população evitar a sua extinção.

A partir dos trabalhos de Muller durante a década de 60 vários avanços foram conseguidos no problema da catraca, sendo o primeiro deles creditado a John Maynard Smith [25] pela sua formulação matemática do problema. Mais tarde o processo da catraca começou a ser utlizado para quantificar a evolução de populações assexuadas em experimentos de laboratório [26].

1.3 Apresentação da tese

Nesta tese o tema principal não envolve diretamente o problema da origem do sexo, mas sim como utilizar a formulação de Muller para estudar numericamente diversos problemas relacionados à teoria da evolução de populações assexuadas sendo o processo da catraca de Muller apresentado em diferentes dinâmicas.

No capítulo 2 serão descritos os algoritmos básicos para o estudo da evolução de populações assexuadas que são largamente utilizados em experimentos computacionais. Neste capítulo ainda serão desenvolvidos os conceitos necessários para o estudo de uma

dinâmica evolutiva como adaptabilidade, seleção, mutação e relevos adaptativos.

Depois de apresentadas as ferramentas básicas para este estudo, ao longo do capítulo 3 será apresentado o processo evolutivo de uma população que possui um tamanho muito grande e cuja a evolução se dá através de passagens seriais por gargalos (*bottleneck serial passage*) [27]. Em um trabalho de 1994, Kondrashov [21] sugere que, no caso de somente existirem mutações deletérias na população, a catraca de Muler pára, porém nesta tese nós mostramos que isso não é verdade e que apesar da veocidade da catraca diminuir bastante, ela nunca pára efetivamente, entrando em um estado estácionário onde os novos cliques sempre acontecem com um grande intervalo de tempo entre si.

Na seção 3.2 será apresentado a catraca de Muller em uma dinâmica, onde as populações podem variar de tamanho ao longo do tempo [28]. Este estudo apresenta uma importante aplicação experimental onde populações reais em laboratórios podem ser cultivadas em uma dinâmica de crescimento ou decrescimento. Um ponto curioso que obtivemos em nosso trabalho é o comportamento contra-intuitivo da catraca de Muller, quando a população se encontra com baixos valores de seleção.

Todas as dinâmicas estudadas nos capítulos anteriores pressupõem o isolamento das populações, porém na natureza estas não se encontram em perfeito isolamento, tendo um comportamento interativo com outras populações e com o meio-ambiente. No capítulo 4 serão apresentadas as modificações necessárias na dinâmica evolutiva para que o problema da catraca de Muller possa ser estudado em populações que interajam através de uma dinâmica do tipo presa-predador[†]. Para isso nos baseamos em um conjunto de equações desenvolvidas por Beddington [29] e obtivemos um comportamento muito interessante para o número mínimo de mutações na população de presas. Ao invés deste valor crescer monotonicamente, observamos a formação de pequenos platos sempre que a população de presas está crescendo, provocando a parada da catraca, enquanto que este número sobe rapidamente sempre que a população diminui.

Finalmente no capítulo 5 serão estudados padrões evolutivos em árvores genealógicas binárias. Este trabalho foi desenvolvido em conjunto, com o então aluno de doutorado do

[†]Este tipo de dinâmica também recebe o nome de parasita-hospedeiro.

grupo de Evolução Molecular Teórica, Leonardo Paulo Maia. Em sua tese [30], defendida em agosto de 2004, o Dr. Maia apresentou os resultados detalhados deste problema e portanto aqui serão apresentados os principais conceitos de árvores no capítulo 5 bem como os resultados por nós obtidos [31] sobre o efeito da seleção em suas topologias, apesar disto este tipo de estudo é de grande importância por permitir a inferência de informações como taxas de extinção, medidas contínuas de especiação, coevolução entre outras.

Capítulo 2

Modelagem de processos evolutivos

A pesquisa moderna em biologia gera uma enorme quantidade de informações sobre o funcionamento de um organismo, contudo, muito dessa informação permanece puramente descritiva. O maior desafio da biologia funcional consiste em agrupar estes dados moleculares de maneira a encontrar um modelo para o funcionamento global de um organismo [32], tendo como principal objetivo entender e predizer o seu comportamento complexo, sua adaptação ao ambiente, diferenciação e mesmo a sua evolução. Atingir este objetivo requer a união de inúmeros avanços experimentais e conceituais em diferentes áreas de conhecimento como Física, Matemática, Química, Computação e a própria Biologia.

A modelagem matemática vem se tornando uma ferramenta importante na pesquisa biológica, preenchendo a escassez de informações sobre muitos sistemas através da formulações de modelos, que apesar de simplificados, permitem estudar características importantes através de equações e simulações computacionais. Recentemente, este tipo de modelagem tem sido aplicado com sucesso para analisar uma grande variedade de sistemas biológicos, incluindo redes de desenvolvimento de formação de padrões em *Drosophilas* [33], crescimentos de células [34], infecções virais [35–38], dinâmicas difusivas em redes discretas [39–42], e particularmente em meu caso, modelos evolutivos de populações assexuadas [27, 28], entre outros.

Neste capítulo serão apresentados as características dos modelos evolutivos estudados

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

nesta tese como relevo de replicação, mutações, reprodução e como essas características podem ser implementadas na modelagem matemática e computacional [43,44], de maneira a descrever o comportamento de um sistema biológico.

2.1 Dinâmica evolutiva

A dinâmica básica utilizada para descrever um processo evolutivo nesta tese baseia-se no modelo de Wright-Fisher para a deriva genética e no modelo de infinitos sítios [45] com algumas modificações que serão explicadas sempre que necessário. Este tipo de modelagem representa a evolução de uma população de tamanho fixo composta por um total de N indivíduos (seqüências), sendo cada indivíduo constituído por um genoma de infinitos sítios (dígitos ou bases). Inicialmente todos os N indivíduos da população possuem o mesmo genótipo, ou seja, todos apresentam as mesmas características e portanto nenhum destes se encontra mais adaptado ao meio-ambiente que os demais. A cada intervalo de tempo é gerada uma nova linhagem baseada na população atual através da reprodução assexuada. Isto é feito sorteando-se N indivíduos com reposição que irão produzir, em cada sorteio, um descendente. Depois de realizados os sorteios a população parental é descartada, restando somente os seus descendentes, e dessa forma todos os indivíduos presentes na população tem a mesma idade, não existindo sobreposição de gerações.

Esta dinâmica vem sendo utilizada por vários integrantes do grupo de Evolução Molecular Teórica, liderado pelo prof. Fontanari, ao longo dos últimos anos para estudar a construção de genealogias, sempre abordando diferentes características do problema evolutivo, entre estes trabalhos destacam-se as dissertações de Milton T. Sonoda [46] e Daniela F. Botelho [47] e as teses de doutorado de Paulo R. A. Campos [48] e Leonardo Paulo Maia [30].

Outro ponto importante que devemos lembrar ao modelar o comportamento evolutivo de uma população real é que o processo de reprodução não está livre de erros, pois um indivíduo pode produzir descendentes com um genótipo diferente do seu através de mutações ocorridas durante a reprodução, isto já foi verificado experimentalmente em

diversas espécies. Em vírus a taxa de mutação é de cerca de 1 dígito por replicação, em micróbios que possuem DNA essa taxa diminui para 1 em 300, enquanto no ser humano são necessários, em média, 1 milhão de replicações para que ocorra uma simples mutação, o que demonstra que sempre existe uma mutação, com uma menor taxa, mesmo nos melhores mecanismos de replicação [49]. Da mesma forma, mutações ocorrem nos genótipos de cada elemento de nossos experimentos, sendo que estas mutações são pontuais, ou seja, elas ocorrem em cada dígito das moléculas independentemente, com isso podemos gerar indivíduos com um nível de adaptação diferente do seu ancestral apenas com uma simples mutação.

Uma aproximação bastante utilizada na literatura e que consideramos aqui, é que não existem mutações benéficas [24], sendo todas as mutações seletivamente neutras [5] ou deletérias [20–23], de forma que o indivíduo mais adaptado é aquele que possui o menor número de mutações.

A função adaptação (adaptabilidade ou do original em inglês *fitness*) de um indivíduo é um dos pontos mais importantes de toda a modelagem, pois ela definirá qual é a probabilidade de sobrevivência de um dado indivíduo ao meio em que vive*, dessa forma quanto maior a adaptação de um indivíduo, maior é a sua probabilidade de sobrevivência para a próxima geração. Isto ficará mais claro quando definirmos o algoritmo utilizado para o processo evolutivo da população na seção 2.2. Inicialmente vamos definir a adaptabilidade de um indivíduo, identificado pelo subíndice i, como sendo

$$W_i \propto f(s, k_i), \tag{2.1}$$

onde s é a taxa de seleção e k_i é o número de mutações do i-ésimo indivíduo da população. A função $f(s, k_i)$ pode tomar diferentes formas dependendo da dinâmica que estamos querendo reproduzir, por este motivo na seção 2.1.1 serão apresentados os diferentes relevos utilizados nesta tese e suas principais características.

*Também pode ser entendido como o número de descendentes que podem ser produzidos por aquele indivíduo, assim um indivíduo com maior adaptabilidade gera mais descendentes.

O parâmetro s define o quão seletivo será o meio-ambiente em que os indivíduos vivem. Desta maneira quando as mutações não modificam em nada a adaptação do indivíduo consideramos o parâmetro s como sendo 0, que é o caso conhecido como seleção neutra. O caso oposto a este é aquele em que cada nova mutação reduz tanto a sua adaptabilidade em relação ao meio, que este indivíduo não consegue gerar descendentes que sobreviverão, sendo este caso conhecido como seleção extrema e cujo parâmetro s assume o valor 1.

2.1.1 Relevos de Replicação

Os relevos de replicação são expressões matemáticas utilizadas para representar o processo seletivo através das mutações sofridas ao longo das gerações pelos indivíduos e as suas diferentes formas de interações com o meio-ambiente. Podemos utilizar estes relevos para definir o quão benéficas ou prejudiciais são essas mutações e como isto pode afetar a linhagem dos indivíduos. Nesta tese foram utilizados três relevos, o relevo de pico único, o relevo multiplicativo e o relevo epistático.

O relevo de pico único é a mais simples representação da interação entre os indivíduos e o meio, pois ele divide toda a população em dois tipos, ou mais precisamente em duas classes, sendo uma classe formada pelos indivíduos que não possuem nenhuma mutação e a outra formada por todos aqueles que possuem ao menos uma mutação, pois neste relevo consideramos que somente a primeira mutação possui um caráter deletério e todas as demais possuem um caráter neutro, isto é equivalente a dizer que a primeira mutação ocorre em um gene que é importante para o indivíduo e que todos as demais mutações ocorrem em partes do genoma que não são essenciais para o organismo, assim estas mutações tem um caráter neutro. Depois que um indivíduo sofre uma mutação em algum ponto de seu genoma, todas as novas mutações não apresentarão nenhum efeito em sua função adaptação, ou seja, para os indivíduos puros (*wildtype*), que não sofreram nenhuma alteração em seu genoma desde o seu mais antigo ancestral, a sua função adaptação será

$$W_i = 1.$$

Enquanto que a função adaptação para os indivíduos que já sofreram pelo menos uma mutação ao longo de sua genealogia será

$$W_i = 1 - s.$$

Desta forma podemos observar que a probabilidade de sobrevivência para um indivíduo que sofreu mutação é menor do que para o indivíduo que não tem nenhuma alteração em seu genoma.

O segundo relevo utilizado é o multiplicativo, que considera que todas as mutações sofridas por cada indivíduo tem o mesmo caráter deletério; em palavras, cada nova mutação reduz a adaptação de um indivíduo de uma quantidade fixa $(1-s)$, ou seja, cada indivíduo terá uma adaptabilidade dada por

$$W_i = (1 - s)^{K_i}. \tag{2.2}$$

A principal diferença entre o relevo multiplicativo e o relevo de pico único é o caráter prejudicial de cada mutação, o que acarreta uma diferença significativa em relação ao número de classes em que a população pode ser dividida, pois se no relevo de pico único só existem 2 classes na população, neste caso se existirem j seqüências com números de mutações diferentes, existirão j classes, onde cada uma será formada única e exclusivamente pelos indivíduos com o mesmo número de mutações.

O último relevo estudado é o relevo epistático, onde a função adaptação para cada indivíduo é dada por

$$W_i = (1 - s)^{K_i{}^{\alpha}}. \tag{2.3}$$

Este caso é parecido com o relevo anterior pois divide a população no mesmo número de classes quantos forem os números de indivíduos diferentes existentes, porém o coeficiente de epistase α é responsável pelo grau de dano que as mutações provocam nos indivíduos, ou

seja, cada nova mutação possui um caráter mais ou menos deletério segundo o parâmetro α.

Podemos classificar os coeficientes de espistase como sendo sinergística, que ocorre quando $\alpha > 1$, onde cada nova mutação possui um caráter mais prejudicial a adaptabilidade do indivíduo que a mutação anterior; atenuante para α entre 0 e 1, sendo cada nova mutação menos deletéria que a anterior e finalmente o caso em que $\alpha = 1$ e representa o limite não epistático, onde este resultado recupera o relevo multiplicativo.

Estas não são as únicas formas possíveis, existem outros relevos como por exemplo o relevo truncado [30], porém estes não serão discutidos nesta tese.

2.2 Modelagem computacional

Nesta seção será introduzido o algoritmo para a modelagem de uma população assexuada com um tamanho constante ao longo de toda a simulação. Este algoritmo é base para a simulação de processos evolutivos ao longo desta tese e sofrerá algumas pequenas modificações, porém sempre que necessário as modificações efetuadas serão discutidas.

Vamos considerar inicialmente uma população haplóide de tamanho N que se reproduz assexuadamente. A cada unidade de tempo, que aqui é considerada como uma geração, são realizados N sorteios com reposição na população parental para compor a nova população, com isso essa nova população somente será formada pelos descendentes da população parental, e por isso dizemos que não existe sobreposição de gerações, pois nenhum indivíduo da população parental fará parte da nova população . Esta dinâmica foi implementada da seguinte forma:

1. A cada indivíduo i da população parental é associado uma adaptabilidade W_i baseada no número de mutações e do relevo escolhido para a dinâmica.

2. É sorteado um indivíduo da população parental para se reproduzir.

3. É sorteado um número aleatório ξ usando uma distribuição uniforme em $(0, 1)$.

   - Se $W_i \geq \xi$ o descendente passará para a próxima geração;

- Se $W_i < \xi$ o descendente não passará para a próxima geração.

4. A todo o indivíduo que passa pela seleção é acrescido um número de mutações gerado por uma distribuição de Poisson† com valor médio U‡, sendo U a taxa de mutação.

Os passos acima são repetidos até que se obtenha uma população de tamanho igual a população inicial.

Inicialmente este tipo de modelagem na qual a população permanece constante ao longo de sua evolução foi proposto por Wright e Fisher para estudar o problema da deriva genética da seguinte forma: Vamos considerar uma população haplóide, onde cada indivíduo é formado única e exclusivamente por um gene, sendo que este pode assumir dois valores seletivamente neutros§, $\mathcal{A}$ e $\mathcal{B}$, e não existe nenhum tipo de mutação que leve o gene $\mathcal{A}$ ao $\mathcal{B}$ ou vice-versa. Se iniciarmos a simulação com metade dos indivíduos com o gene $\mathcal{A}$ e a outra metade formada pelo gene $\mathcal{B}$ ao longo da evolução desta população um dos genes se fixará na população, ou seja, estará presente em todos os indivíduos, enquanto o outro será extinto. Este processo evolutivo de uma população assexuada com tamanho constante onde um gene se fixa na população através de sorteios aleatórios, sem nenhum tipo de pressão de seleção, define a deriva genética descrita no modelo de Wright-Fisher [46, 50].

Outros fatores além das taxas de mutação, seleção e relevo também são importantes quando tentamos modelar um sistema biológico, fatores esses, como por exemplo, a ploidia dos indivíduos [51, 52], o tipo de reprodução, os números de dígitos que formam o genoma bem como o seu comprimento, entre outros.

†A probabilidade de que um evento de probabilidade p ocorra l vezes num total de L tentativas independentes segue uma distribuição binomial, quando $l << L$ temos que a binomial se aproxima de uma Poisson e por este motivo utilizamos esta distribuição para definir o número de mutações recebidas por um indivíduo com um genoma infinito ($L = \infty$) [47].

‡Existe a possibilidade que o valor calculado na distribuição de Poisson seja 0 de forma que o genoma permanece o mesmo.

§Ser seletivamente neutro representa que um genótipo não possui nenhum tipo de vantagem sobre o outro.

IFSC-USP SERVIÇO DE BIBLIOTECA E INFORMAÇÃO

A ploidia de um indivíduo define quantas cópias de cada cromossomo existem em cada célula deste[¶] [53]. Por exemplo, indivíduos haplóides possuem uma cópia de cada cromossomo, enquanto indivíduos diplóides possuem duas cópias e assim por diante.

A finitude do genoma é um fator essencial na evolução da população em experimentos computacionais, pois existem diferenças importantes quando trabalhamos com um genoma de tamanho finito ou infinito. No caso infinito o número de mutações que uma seqüência recebe é dado por uma distribuição de Poisson, e todas essas novas mutações sempre ocorrem em um dígito que não havia sido mutado anteriormente. Isso pode ser entendido se considerarmos que a probabilidade p de mutar qualquer sítio do genoma como sendo $p = 1/L$, onde L é o tamanho do genoma, quando L é infinito temos que p é 0. Porém quando o genoma é finito o valor de p assume um valor positivo ($p > 0$), assim existe a possibilidade que um dígito mute mais de uma vez podendo, em alguns casos, fazer com que o seu valor volte ao seu estado inicial, quando isso ocorre damos a esta mutação o nome de mutação reversa.

O processo de mutação reversa é muito importante, pois afeta de forma significativa grandezas importantes como as distâncias entre as seqüências[‖]. Isto pode ser exemplificado se assumimos um genoma formado por cadeias binárias. As cadeias binárias são formadas por dígitos (alfabeto) que só possuem dois valores, como por exemplo as bases purinas e pirimidinas, em um caso genérico podemos escrever os dígitos como sendo $\mathcal{A}$ e $\mathcal{B}$, assim temos a seguinte regra de mutação,

$$\mathcal{A} \rightleftharpoons \mathcal{B}$$

onde $\mathcal{A}$ pode mutar para se transformar em $\mathcal{B}$ ou $\mathcal{B}$ pode mutar para se transformar

[¶]Isto só é válido quando não consideramos os gametas da população, pois em indivíduos diplóides que se reproduzem sexuadamente, cada gameta possui uma única cópia de cada cromossomo, porém estas células só formarão um indivíduo quando forem fecundadas por um outro gameta gerando assim um indivíduo diplóide.

[‖]A distância entre seqüências é importante quando analisamos a genealogia dos indivíduos e é baseado na quantidade de dígitos mutados em dois genomas, este conceito será apresentado em detalhes no capítulo 5.

em $\mathcal{A}$. Se inicialmente o dígito do sítio i era $\mathcal{A}$, após sofrer uma mutação ele passa a ser $\mathcal{B}$, e quando uma nova mutação ocorrer no mesmo sítio, dizemos tratar-se de uma mutação reversa, pois este dígito voltará a ser $\mathcal{A}$. Este efeito é menos acentuado quando aumentamos o número de letras do alfabeto que cada sítio pode assumir; por exemplo, para cadeias quaternárias, formadas pelos dígitos $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ e $\mathcal{D}$, as regras de mutação são

$$\begin{array}{lll} \mathcal{A} \rightleftharpoons \mathcal{B} & \mathcal{A} \rightleftharpoons \mathcal{C} & \mathcal{A} \rightleftharpoons \mathcal{D} \\ \mathcal{B} \rightleftharpoons \mathcal{C} & \mathcal{B} \rightleftharpoons \mathcal{D} & \\ \mathcal{C} \rightleftharpoons \mathcal{D} & & \end{array}$$

de forma que a probabilidade de uma mutação reversa ocorrer é bem menor, pois o espaço de possibilidade de mutações é maior. Explicitamente, enquanto que no caso de seqüências binárias uma segunda mutação no mesmo sítio será com certeza reversa, no caso de seqüências quaternárias essa mutação terá apenas uma chance de 33% de ser reversa.

Quando trabalhamos com um genoma infinito não temos o processo da mutação reversa, porém temos um processo muito importante referente a evolução que é conhecido como Catraca de Muller e que será explicado com detalhes no capítulo 3.

Capítulo 3

Evolução de populações assexuadas

Alguns experimentos e modelos teóricos vem sendo propostos para estudar a evolução de diferentes organismos que se reproduzem assexuadamente. Neste capítulo serão abordados o problema da evolução através do processo da catraca de Muller, onde alguns resultados analíticos são apresentados, bem como resultados de experimentos computacionais.

O modelo teórico utilizado na seção 3.1 é baseado em experimentos de passagens seriais com gargalo (*serial bottleneck passage*) cujos resultados são bastante interessantes e apresentam uma possível aplicação experimental [27].

Na seção 3.2 são apresentados resultados onde as populações não apresentam um número fixo de indivíduos, podendo começar com um único exemplar e expandir indefinidamente ou mesmo se extinguir após um período de expansão devido à degeneração de seu genoma pelo acúmulo de mutações deletérias (*mutational meltdown*) [54].

3.1 Catraca de Muller em passagens seriais com gargalo

O estudo do processo evolutivo em populações assexuadas proposto por Muller [10] na década de 60, e que mais tarde recebeu o seu nome [55] tem grande importância teórica [21, 56] e também tem sido observado experimentalmente em algumas populações virais [26]. Este processo pode ser entendido como a perda estocástica da classe de

indivíduos mais adaptados, em uma população de tamanho finito, devido ao acúmulo de mutações deletérias adquiridas ao longo de sua evolução e, a medida que um dígito de uma seqüência sofre uma mutação, esta seqüência jamais poderá voltar ao "estado" anterior com um menor número de mutações, pois consideramos um genoma infinito onde não existem mutações reversas.

Mais precisamente, vamos considerar uma população com N indivíduos cujo genótipo tem um comprimento infinito, onde inicialmente todos os genomas são idênticos. Durante a evolução dessa população, a cada nova geração, os indivíduos podem adquirir novas mutações como descrito no capítulo anterior. Quando todos os indivíduos da população tiverem adquirido pelo menos uma mutação, pode-se dizer que a catraca de Muller atuou, em outras palavras que a catraca deu um"clique". O movimento da catraca pode ser entendido se dividirmos os indivíduos em classes identificadas pelo número de mutações, de maneira que só fará parte da classe j os indivíduos que possuírem j mutações, de forma que a catraca de Muller sempre dará um "clique"quando a classe mais adaptada se extinguir, assim a catraca se movimentará até parar em um novo ponto ("dente da catraca") que é definido pela nova classe que possui o menor número de mutações por indivíduo em toda a população.

Para caracterizar melhor este processo, vamos identificar os indivíduos da população, em um dado tempo t, por um vetor $X(t)$ da forma

$$X(t) = (X_0(t), X_1(t), \ldots, X_k(t), \ldots), \tag{3.1}$$

onde $X_k(t)$ representa a quantidade de indivíduos na classe k, na geração t. Como estamos considerando o tamanho da população igual a N, podemos também identificar um vetor $C(t)$ referente as freqüências das classes, de forma que

$$C_k(t) = \frac{X_k(t)}{N}, \qquad \text{onde} \sum_{i=0}^{\infty} C_i(t) = 1 \tag{3.2}$$

indica a freqüência de indivíduos com k mutações na população.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

Definindo então a probabilidade de termos k novas mutações em uma seqüência como sendo dado pela distribuição de Poisson [6,57] de média U,

$$\mathcal{M}_k = e^{-U}\frac{U^k}{k!}, \tag{3.3}$$

é possível escrever uma relação de recorrência para as freqüências das classes [46,58] como

$$C_m(t+1) = \sum_{k=0}^{m} \mathcal{M}_{m-k}\frac{C_k(t)W_k}{\sum_{i=0}^{\infty} C_i(t)W_i}, \quad \text{onde} \quad W_k = (1-s)^{K^\alpha}. \tag{3.4}$$

Podemos notar nesta expressão que a classe m é formada pelos indivíduos da própria classe que não sofreram mutação, ou por qualquer indivíduo da classe k que tenha sofrido exatamente $m-k$ mutações.

Se considerarmos o tamanho da população muito grande de forma que tanto a mutação que gera novos indivíduos, quanto a seleção, que elimina os menos adaptados se balanceiem, teremos um estado estacionário para a relação de recorrência anterior, assim $C_m(t+1) = C_m(t) = C_m$, $\forall\ m$, de forma que

$$C_m = \frac{1}{T_e}\sum_{k=0}^{m} \mathcal{M}_{m-k}C_kW_k, \tag{3.5}$$

onde $T_e = \sum_{i=0}^{\infty} C_i(t)W_i$ é a adaptação média no regime estacionário.

Fazendo $m = 0$ em (3.5), obtemos

$$\begin{aligned} C_0 &= \frac{1}{T_e}\mathcal{M}_0C_0W_0 \\ C_0 &= \frac{1}{T_e}e^{-U}C_0 \\ T_e &= e^{-U}. \end{aligned} \tag{3.6}$$

Desta maneira notamos pelas eqs. 3.6 que a adaptação média independe do relevo utilizado, já que esta função é independente dos coeficientes de epistase e de seleção,

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

dependendo exclusivamente da taxa de mutação U. Assim podemos reescrever a relação de recorrência 3.5 como

$$\begin{aligned} C_m &= \frac{1}{T_e}\left[\mathcal{M}_0 C_m W_m + \sum_{k=0}^{m-1} \mathcal{M}_{m-k} C_k W_k\right] \\ C_m &= \left(1 - \frac{\mathcal{M}_0 W_m}{T_e}\right)^{-1} \frac{1}{T_e} \sum_{k=0}^{m-1} \mathcal{M}_{m-k} C_k W_k \\ C_m &= \frac{e^U}{1-(1-s)^{m^\alpha}} \sum_{k=0}^{m-1} e^{-U} \frac{U^{m-k}}{(m-k)!} C_k W_k \end{aligned} \tag{3.7}$$

ou simplesmente,

$$C_m = \frac{1}{1-(1-s)^{m^\alpha}} \sum_{k=0}^{m-1} \frac{U^{m-k}}{(m-k)!} C_k W_k. \tag{3.8}$$

Esta expressão se aplica perfeitamente ao caso em que temos a classe inicial na população, ou seja C_0 diferente de zero, porém este resultado pode ser generalizado para o caso em que todas as $j-1$ primeiras classes são extintas como sendo dado por

$$C_m = \frac{1}{(1-s)^{j^\alpha} - (1-s)^{m^\alpha}} \sum_{k=j}^{m-1} \frac{U^{m-k}}{(m-k)!} C_k W_k; \quad m > j \tag{3.9}$$

Dadas as distribuições de equilíbrio das classes dos indivíduos na população, estudamos o processo de evolução através de passagens seriais com gargalo, figura 3.1. Neste tipo de experimento alguns indivíduos são previamente escolhidos para iniciar uma população e para isso são introduzidos em uma placa que permite o seu crescimento livremente durante um período de incubação, no caso do Bacteriófago $\phi 9$ a população pode passar de um único indivíduo para 8×10^9 em 24 horas [58], depois desta incubação, alguns indivíduos são escolhidos aleatoriamente e passados para um outro tubo onde o ciclo de crescimento recomeça.

Com base na passagem serial com gargalo e nas fórmulas para o caso em que as populações se encontram em equilíbrio, desenvolvemos um experimento computacional para

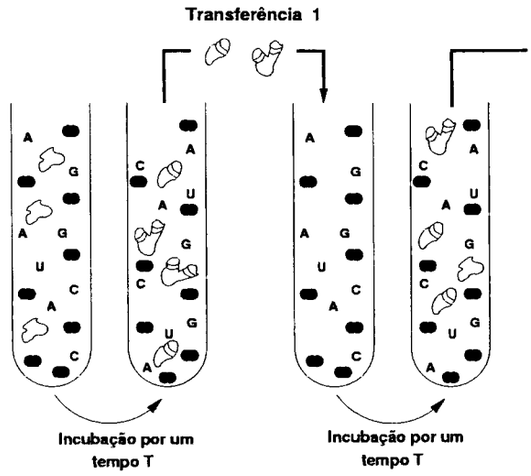

Figura 3.1: Passagem serial com gargalo. Neste experimento alguns indivíduos da população que estavam no primeiro tubo são escolhidos aleatoriamente para serem passados para um novo tubo dando origem a uma nova população.

verificar a evolução de espécies assexuadas, observando o comportamento da catraca de Muller. Para isso inicialmente colocamos indivíduos sem nenhuma mutação na população e deixamos essa população evoluir até atingir o estado de equilíbrio, de maneira que as classes seriam dadas pela expressão 3.9, isto é o equivalente a deixar a população evoluir em um tubo por um período t de incubação. Depois de atingir o equilíbrio é escolhido, aleatoriamente, um indivíduo para ser passado para uma nova cultura, repetindo novamente o processo de incubação. Isso é equivalente ao experimento de passagens seriais descrito na figura 3.1.

O processo da catraca de Muller acontece quando é feita a passagem de um indivíduo que é escolhido aleatoriamente para criar uma nova cultura, isso porque nem sempre o indivíduo que é escolhido é o mais adaptado na população, caracterizando a perda das classes com menor número de mutações. Vale lembrar que essa passagem não implica necessariamente na perda dos indivíduos mais adaptados, dependendo dos parâmetros de seleção, mutação, epistase e do relevo escolhido.

Para tentar entender melhor o processo, analisamos duas grandezas fundamentais, a primeira é o comportamento da classe m com relação ao número de mutações ao longo dos experimentos, e o segundo, é o tempo entre dois cliques da catraca $< \delta\tau >$.

Nas figuras 3.2 e 3.3 o comportamento da freqüência mínima (C_m) em função do número de mutações m na população são apresentados, onde podemos ver nas figuras 3.2 e 3.4, que para $\alpha = 1$, tanto a freqüência de mutações quanto o tempo de perda da classe mais baixa, $< \delta\tau >$ são constantes como esperado para regime não epistático. Neste regime a velocidade da catraca de Muller é constante ao longo da evolução da população.

Quando aumentamos o efeito da epistase, aumentando o parâmetro α para valores maiores que a unidade, era esperado que a da catraca parasse a medida que aumentassem o número de dígitos mutados [21], porém isto não fica evidenciado em nossos resultados que demonstram claramente que a velocidade da catraca entra em um regime estacionário, pois o intervalo de tempo entre dois sucessivos cliques da catraca ($\delta\tau$) permanece constante.

Porém no caso em que $\alpha < 1$, o que acontece é que a velocidade da catraca aumenta, porém não indefinidamente, assim depois do transiente a velocidade da catraca permanece em movimento constante, de maneira que os diversos cliques sempre se dão com o mesmo intervalo de tempo $\delta\tau$, figura 3.4.

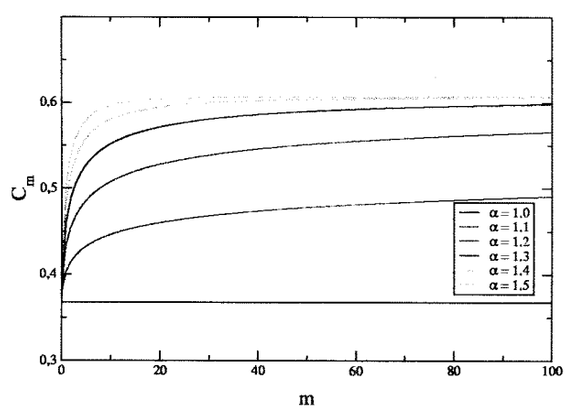

Figura 3.2: Comportamento da freqüência mínima C_m para diferentes valores de $\alpha \geq 1$, com $s = 0.5$ e $U = 0.5$.

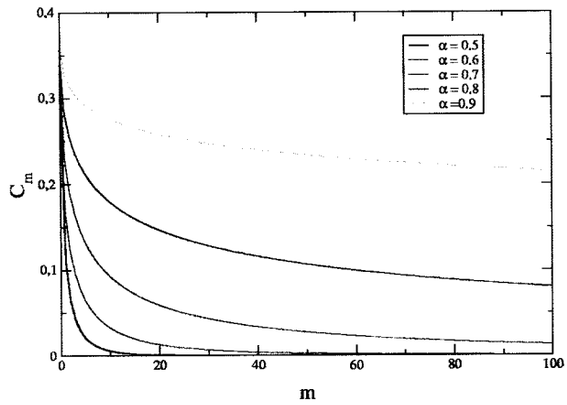

Figura 3.3: Comportamento da freqüência mínima C_m para diferentes valores de $\alpha < 1$, com $s = 0.5$ e $U = 0.5$.

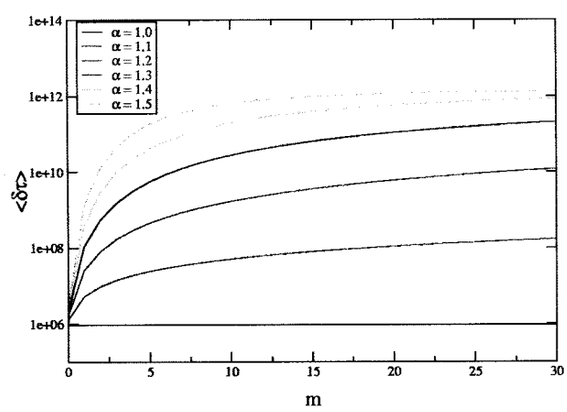

Figura 3.4: Tempo de perda de C_m para diferentes valores de $\alpha \geq 1$, com $s = 0.5$ e $U = 0.5$.

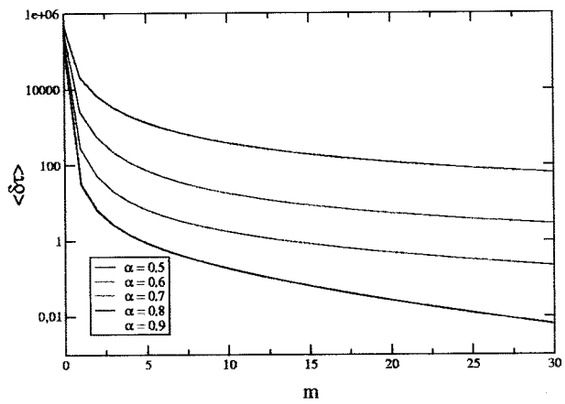

Figura 3.5: Tempo de perda de C_m para diferentes valores de $\alpha < 1$, com $s = 0.5$ e $U = 0.5$.

3.2 Populações que apresentam tamanho variável

Manter a população constante ao longo dos experimentos é algo razoavelmente simples em computação, porém não é tão simples de se conseguir em um laboratório, por este motivo estudar uma população que varia ao longo do tempo é importante quando tentamos modelar a Natureza. Pensando nisso estudamos o comportamento de populações que podem variar de tamanho ao longo tempo. Diferentes cenários são possíveis para este tipo de evolução e baseado na expansão e no decrescimento da população, podemos identificar diversos regimes, como por exemplo:

1. A população inicial, no instante $t = 0$, é formada por um único indivíduo sem nenhuma mutação e cresce exponencialmente com o tempo sem nenhum limitante externo. Esta expansão é conhecida como crescimento Malthusiano.

2. A população inicial, no instante $t = 0$, é formada por um único indivíduo sem mutação e cresce exponencialmente com o tempo até atingir um tamanho N, a partir deste ponto existe um limitante externo, que poder ser entendido como uma restrição do meio-ambiente, como por exemplo, a alimentação disponível ou mesmo o espaço físico para os indivíduos se reproduzirem, de forma que as gerações seguintes serão compostas por N indivíduos.

3. A população inicial, no instante $t = 0$, é composta por N indivíduos distribuídos em classes C_n, onde n identifica o número de mutações de cada indivíduo dentro da classe C . A quantidade desses indivíduos dentro das classes segue uma distribuição de Poisson [27] com média $\frac{U}{s}$, onde U é o taxa de mutação e s a taxa de seleção.

Algumas modificações foram implementadas no algoritmo descrito no capítulo 2 para que o processo da catraca de Muller fosse estudado para o caso em que a população não permanece constante ao longo de sua história.

Cada indivíduo da população é formado por um genoma infinito [45] e tem uma adaptabilidade definida pelo tipo de relevo utilizado, sendo neste caso determinada pelo relevo multiplicativo,

$$W_i = (1-s)^{k_i}, \tag{3.10}$$

onde s é a taxa de seleção e k_i o número de mutações do indivíduo i.

A evolução da população se dá em passos discretos, sem a sobreposição de gerações, onde uma unidade de tempo t é equivalente a uma geração. Nesta seção estudamos os regimes 1 e 3 onde cada indivíduo se reproduz assexuadamente gerando r novos indivíduos, sendo essa a principal diferença introduzida.

O valor de r pode ser entendido como a taxa de reprodução que varia de indivíduo para indivíduo e não como um parâmetro constante. Baseado nesta idéia o valor de r passou a ser uma variável aleatória obtida por uma distribuição de Poisson de média R, onde R representa o número de descendentes gerados por cada indivíduo da população

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

parental. O objetivo principal desta modificação é diminuir a taxa de crescimento da população, pois fixando o valor de $r = R$, a população crescerá exponencialmente com o tempo a uma taxa Rt. Para tornar o crescimento um pouco mais lento modificamos a quantidade de descendentes de um valor constante para um valor gerado pela distribição de Poisson.

Todo novo indivíduo gerado pela reprodução assexuada, recebe as mutações de seu "pai"acrescido de l novas mutações dadas por uma distribuição de Poisson $p(l; U)$, onde U é a taxa de mutação.

Um ponto interessante, é que nem todos os novos descendentes sobrevivem para formar a próxima geração, pois após o nascimento, os indivíduos passam por uma seleção da seguinte forma:

- É calculado a adaptabilidade W dos indivíduos pela equação 3.10
- É sorteado um indivíduo da população parental para se reproduzir.
- $W_{descendente}$ é comparado a um número aleatório ξ gerado por uma distribuição uniforme (0,1), onde
  - Se $W_{descendente} \geq \xi$ o descendente sobreviverá para a próxima geração;
  - Se $W_{descendente} < \xi$ o descendente é retirado da população.

Alguns detalhes devem ser levados em consideração neste caso. O primeiro é que o valor de r deveria depender explicitamente da adaptabilidade do indivíduo, assim um indivíduo mais adaptado deveria produzir um número maior de descendentes do que um menos adaptado, porém o número de descendentes gerados depende somente da taxa de reprodução e não do $f$itness do indivíduo. A correção a este efeito é feita pelo processo seletivo, pois mesmo que um indivíduo com W pequeno produza vários descendentes, apenas um pequeno número destes passarão para a geração seguinte, pois em seu genoma existirão várias mutações que diminuem sua probabilidade de sobrevivência, enquanto que no caso do indivíduo mais adaptado um maior número de descendentes poderá sobreviver devido ao pequeno número de mutações em seu genoma.

Outro ponto extremamente importante é que na seção anterior a população nunca se extinguia, pois esta permanecia no valor de equilíbrio ao longo de toda a simulação. Neste novo caso pode existir a extinção da população (*m*utational meltdown) [54] mesmo para o crescimento exponencial, pois a medida que aumenta o número de mutações no genoma dos indivíduos, menor é o número de descendentes que sobreviverão a seleção. Este efeito é bastante interessante pois podemos observar a catraca de Muller durante todo o processo evolutivo, desde as primeiras gerações até uma possível extinção.

Baseados neste modelo foram realizados diversos experimentos computacionais tentando verificar o efeito da catraca de Muller.

A figura 3.6 mostra o comportamento da média do número mínimo de mutações k_m em função do tempo (gerações). Podemos observar na figura que a catraca, para baixos valores de mutação, pára completamente, o que pode ser verificado pela existência de platôs, enquanto que para altas taxas de mutação a catraca continua em movimento uniforme (detalhe no gráfico 3.6).

Nas figuras de 3.7 à 3.14 estudamos o comportamento de k_m, $< k >$, $< W >$ e $< N >$ para diferentes taxas de reprodução R e uma taxa de mutação fixa U. Aqui podemos observar que a medida que R aumenta, mais rápido se dá o crescimento da população. Na figura 3.10 observamos que a população cresce muito pouco para $R = 1.5$ e logo começa a "sentir"os efeitos da seleção diminuindo rapidamente o seu valor após 8 gerações. Porém para valores de R maiores, como é o caso da figura 3.17 a população cresce exponencialmente quando $s = 0.1$, mas para valores maiores de seleção, podemos dizer que esta taxa equilibra o crescimento, de maneira que a população não expande mais exponencialmente $S \geq 0.2$. O comportamento do número de mutações e da adaptabilidade média são praticamente os mesmos nessas figuras onde estes tendem a uma saturação após algumas gerações.

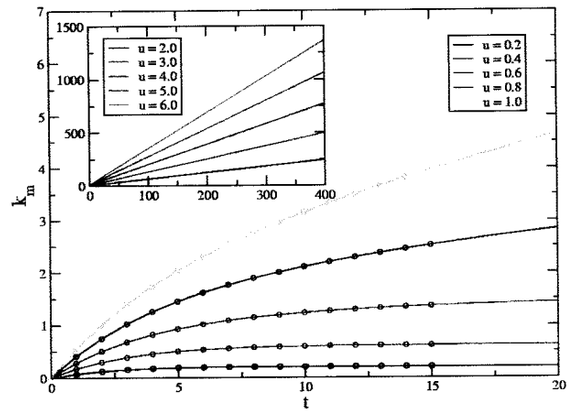

Figura 3.6: Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de mutação U. $s = 0$ e $R = 2$ com médias em 10^6 populações independentes.

Um ponto interessante pode ser observado nas figuras 3.11 e 3.12 mostram o comportamento do número médio de mutações $< k >$ e da média do número mínimo de mutações k_m na população, respectivamente.

O comportamento de $< k >$ para a população constante é conhecido [56] e igual a Ut, como apresentado no apêndice A, sendo que este comportamento é o mesmo para k_m [46]. Assim podemos identificar a velocidade da catraca calculando a inclinação de k_m ou $< k >$, porém nós observamos neste trabalho que isto não ocorre neste modelo. O valor de $< k >$ no regime neutro é exatamente o previsto na literatura, curva preta na figura 3.11, sendo que este comportamento não é acompanhado pelas demais curvas, como esperado.

Como a catraca de Muller é um processo associado à perda estocástica da classe de indivíduos mais adaptada na população, e como os comportamentos de k_m e $< k >$ são diferentes, a velocidade da catraca só pode ser inferida pelo valor de k_m, porém o valor de

k_m possui um comportamento mais complicado, como pode ser observado na figura 3.12, comportamento este que até então nunca tinha sido publicado na literatura. Partindo do caso neutro, a velocidade da catraca aumenta até $s \approx 0.1$, o que é um comportamento totalmente contra-intuitivo, enquanto que para valores de seleção entre 0.1 e 1.0 a catraca diminui a velocidade monotonicamente.

Este comportamento de k_m pode ser explicado pela combinação de dois fatores. O primeiro é que para uma pequena seleção, a diferença na adaptabilidade entre os indivíduos da população (principalmente para poucas gerações) é pequena, figura 3.13, fazendo com que a perda dos indivíduos mais adaptados, pelo fenômeno da deriva genética, seja possível como no caso do regime neutro. O outro fator que atua simultaneamente é o próprio efeito da seleção que é feita após a reprodução, de forma que a combinação destes dois efeitos acelera a catraca para baixos valores seletivos.

Na figura 3.13, observamos o comportamento da adaptabilidade média dos indivíduos na população. Para uma taxa de seleção pequena, $s = \{0.03, 0.05\}$ existe uma diminuição constante na adaptabilidade média da população, enquanto que para valores maiores, existe um rápido decrescimento nas primeiras gerações seguido por uma situação de equilíbrio, onde podemos observar que a adaptabilidade média é maior para maiores valores seletivos. Este comportamento é esperado, pois a seleção tende a eliminar os indivíduos menos adaptados na população, assim quanto maior a seleção, maior é a perda dos indivíduos menos adaptados, e como conseqüência direta maior é o valor da adaptabilidade média na população.

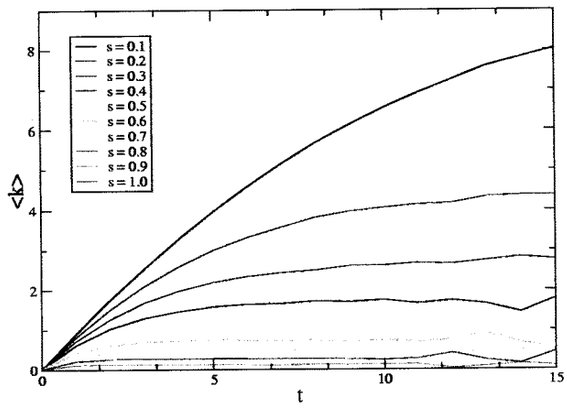

Figura 3.7: Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes.

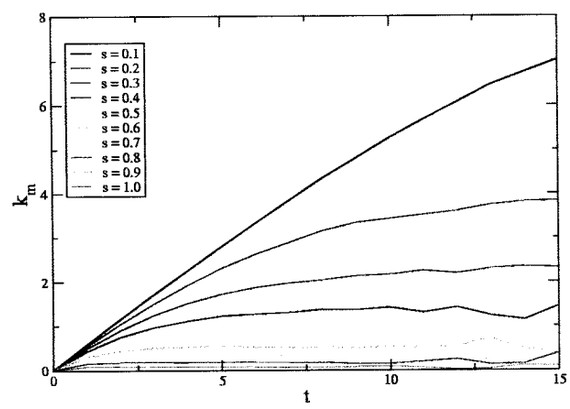

Figura 3.8: Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes.

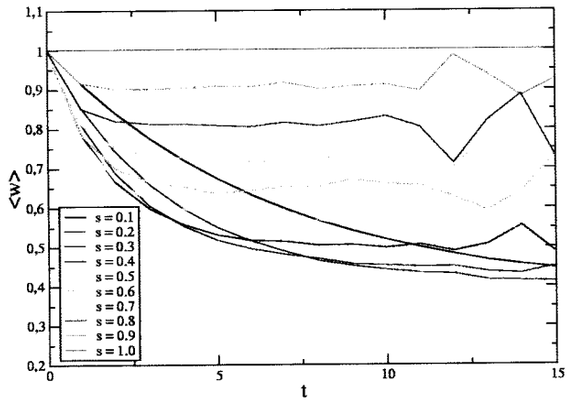

Figura 3.9: Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes.

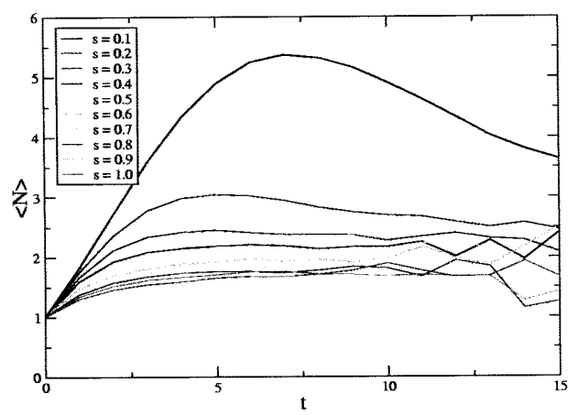

Figura 3.10: Gráfico do número médio de indivíduos na população versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 1.5$ com médias em 10^5 populações independentes.

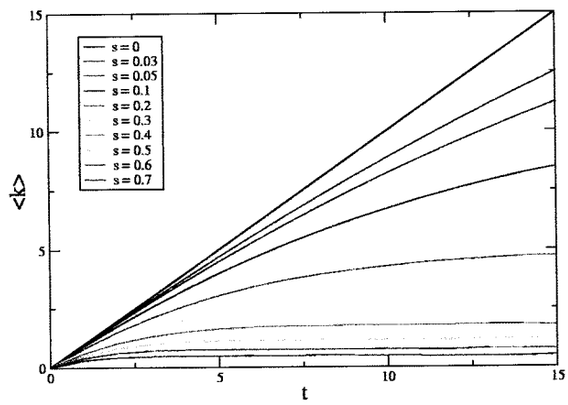

Figura 3.11: Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2$ com médias em 10^5 populações independentes.

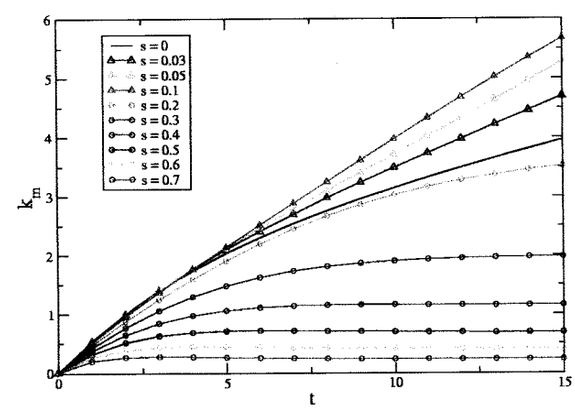

Figura 3.12: Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2$ com médias em 10^5 populações independentes. Os $\triangle$ mostram o comportamento para $s[0, 0.1]$ enquanto que $\bigcirc$ mostram o comportamento para $s(0.1, 1]$.

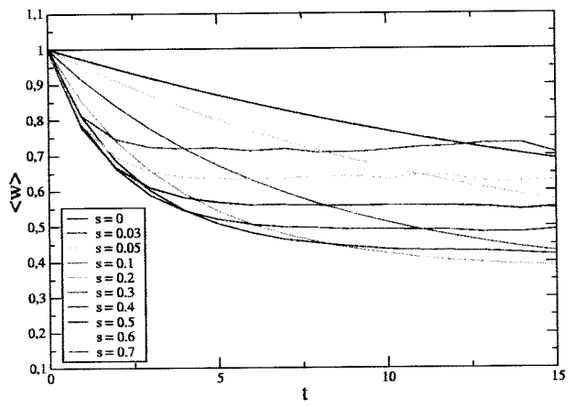

Figura 3.13: Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2$ com médias em 10^5 populações independentes.

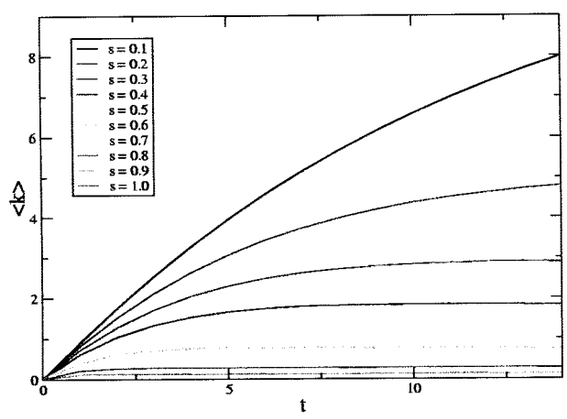

Figura 3.14: Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes.

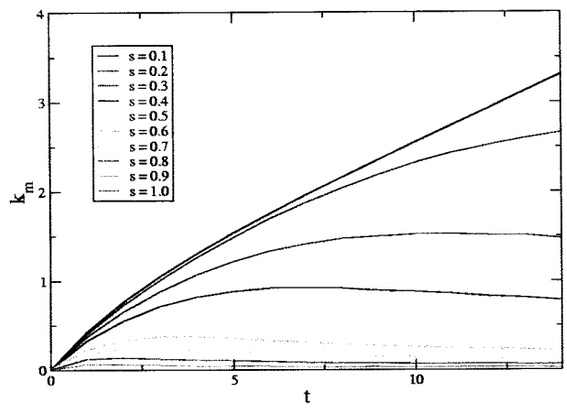

Figura 3.15: Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes.

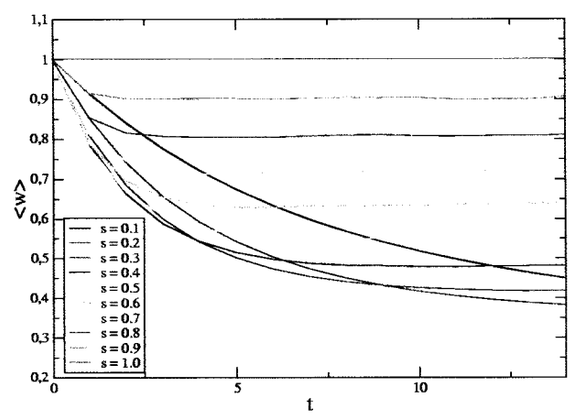

Figura 3.16: Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

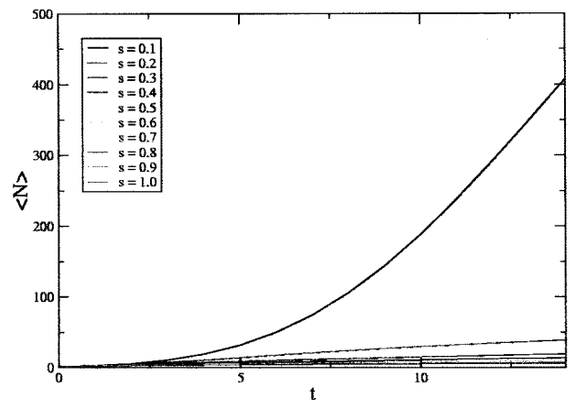

Figura 3.17: Gráfico do número médio de indivíduos na população versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 2.5$ com médias em 10^5 populações independentes.

Para finalizar este estudo consideramos o caso oposto, que representa aquele em que uma população de N indivíduos está no equilíbrio em $t = 0$ e durante a evolução de sua linhagem os indivíduos começam a produzir um número de descendentes, que em média, é menor que a unidade.

Neste caso podemos observar comportamentos distintos para o número mínimo de mutações k_m e para o número médio de mutações na população $< k >$. Enquanto o valor de $< k >$ se mantém praticamente constante ao longo de toda a evolução, o valor de k_m cresce rapidamente a medida que a população decresce, logo nas primeiras gerações. Podemos dizer que mesmo possuindo um número menor de mutações, o processo da deriva genética, não permite a manutenção dos indivíduos mais adaptados, enquanto que o número médio de mutações na população permanece constante.

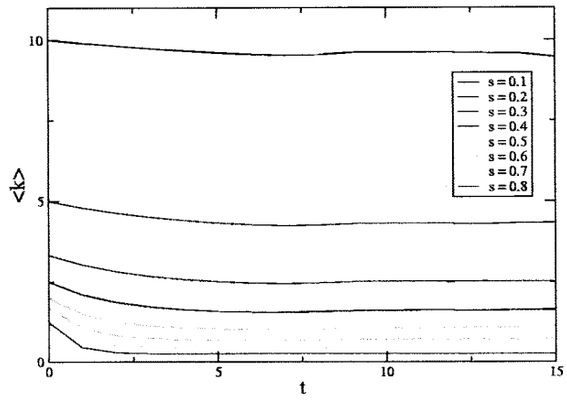

Figura 3.18: Gráfico da média do número mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes.

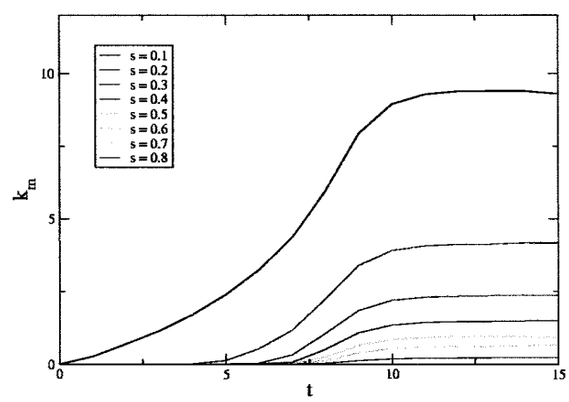

Figura 3.19: Gráfico da média do número mínimo de mutações versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes.

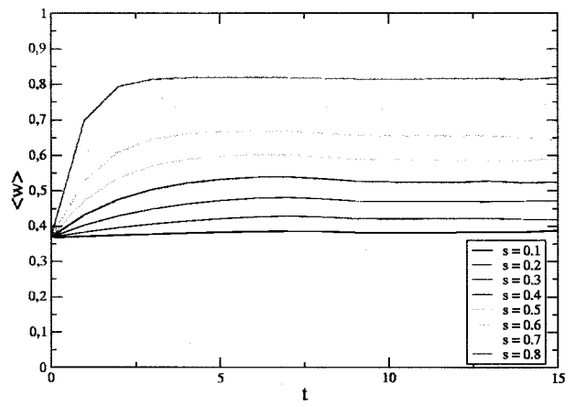

Figura 3.20: Gráfico da adaptabilidade média versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes.

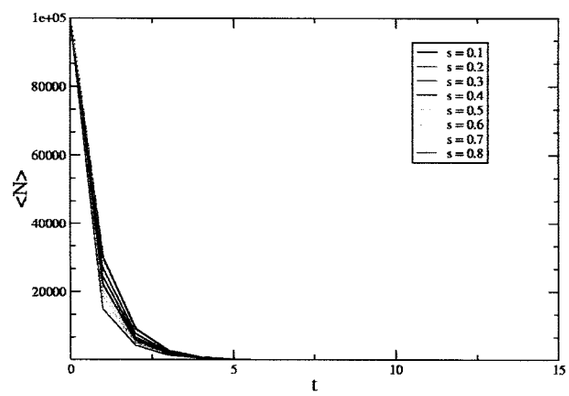

Figura 3.21: Gráfico do número médio de indivíduos na população versus o tempo (gerações) para diferentes valores da taxa de seleção s. $U = 1.0$ e $R = 0.9$ com médias em 10^6 populações independentes.

Capítulo 4

Evolução de populações assexuadas em um sistema de presa-predador

Inúmeras áreas da Biologia tiveram um grande avanço durante o século XX, vários são os exemplos que poderiam ser citados, como a própria teoria da evolução que vem sendo estudada nesta tese, até a modelagem ecológica, onde as interações entre espécies e o meio-ambiente representam um papel importante. O estudo das interações será discutido neste capítulo visando um maior entendimento do processo evolutivo descrito pela catraca de Muller.

Existem diversas formas de classificar as interações entre espécies, sendo os tipos principais [59]:

1. Presa-predador: Quando a taxa de crescimento de uma população decresce a medida que a taxa da outra população cresce.

2. Competição: Quando a taxa de crescimento de cada população diminui com o tempo.

3. Mutualismo ou simbiose: Quando a taxa de crescimento das populações estão relacionadas entre si, de forma que as espécies se beneficiem desta interação.

Neste capítulo serão apresentados estudos numéricos sobre o caso baseado na dinâmica presa-predador.

Um dos primeiros modelos que tentou descrever a interação entre espécies foi proposto por Lotka em 1925 e independentemente por Volterra 1926. O modelo presa-predador de Lotka-Volterra [59], em sua formulação inicial, apresenta uma dinâmica para duas populações, a de presas H e a de predadores P, dadas pelas equações

$$\begin{aligned} \frac{dN_H}{dt} &= N_H(a - bN_P) \\ \frac{dN_P}{dt} &= N_P(cN_H - d), \end{aligned} \tag{4.1}$$

onde a,b,c,d são constantes positivas. Nestas equações podemos observar que as duas populações estão acopladas, de maneira que na ausência de predadores ($N_p = 0$), o número de presas cresce exponencialmente a uma taxa a, sendo este crescimento denominado de crescimento Malthusiano. Porém quando a população de presas se extingue temos a extinção das populações de predadores. Estas equações mostram um possível comportamento entre duas espécies que interagem entre si, porém modelos com n espécies, baseados na eq. 4.1 vem sendo propostos [60], como uma tentativa de melhor entender comportamentos do tipo presa-predador na natureza.

Neste capítulo estudamos o processo da catraca de Muller em populações que interagem em uma dinâmica do tipo presa-predador, porém decidimos estudar um modelo que fosse um pouco mais realista e que de alguma forma, apresentasse o efeito do meio-ambiente nas equações de Lotka-Volterra. Por este motivo decidimos trabalhar com um conjunto de equações que foram inicialmente propostas por Beddington e colaboradores em 1975 [29] e possuem uma dinâmica que se adéqua muito bem ao trabalho originalmente proposto nesta tese, sendo essas equações

$$\begin{aligned} N_H(t+1) &= N_H(t)\exp\left[r(1 - N_H(t)/\kappa) - aN_P(t)\right] \\ N_P(t+1) &= N_H(t)\left[1 - \exp(-aN_P(t))\right], \end{aligned} \tag{4.2}$$

onde r é a taxa de crescimento da população de presas, κ é sua constante de acoplamento com o meio em que vive, sendo esta constante um limitador do seu crescimento e a uma

constante positiva.

Dentre as principais características deste modelo destacamos que no caso da extinção da população de presas $N(t) = 0$, temos também uma extinção para a população de predadores devido ao acoplamento das duas populações na eq. 4.2. Porém quando há a extinção da população de predadores, observamos o crescimento da população de presas seguindo a dinâmica do modelo Malthusiano, ou seja, para uma taxa de crescimento r maior do que 1 a população de presas cresce, porém esse aumento do número de indivíduos é limitado pelo valor definido por κ, enquanto que para r menor que a unidade a população se extingue, como descrito nas equações de Lotka-Volterra.

4.1 O Modelo Presa-Predador

O modelo que estudamos possui uma dinâmica parecida com a apresentada no capítulo 2, porém existem algumas diferenças devido principalmente ao estudo de mais de uma espécie, por isso ela será descrita em maiores detalhes nesta seção.

As duas populações são formadas por um número finito de indivíduos, sendo N_H a quantidade de presas e N_P a de predadores, dados pelas eqs. 4.2. Cada indivíduo da população de presas é formado por um genoma infinito [45] e tem uma adaptabilidade associada ao seu número de mutações dada por

$$W_i^H = (1 - s^H)^{k_i^H}, \tag{4.3}$$

onde s^H é a taxa de seleção e k_i^H o número de mutações do indivíduo i da população de presas. A mesma função adaptação é aplicada para a população de predadores, de forma que sua equação será

$$W_i^P = (1 - s^P)^{k_i^P}. \tag{4.4}$$

Neste caso, como considerado anteriormente, todas as mutações são deletérias, de

forma que com o aumento do número de mutações há uma diminuição da adaptabilidade do indivíduo.

A evolução das populações se dá em passos discretos, sem a sobreposição de gerações, onde uma unidade de tempo t é equivalente a uma geração. Também admitimos que a população de presas e de predadores se reproduzem simultaneamente, ou seja, a cada unidade de tempo obtemos uma nova geração para as populações de presa e de predadores. Apesar desta suposição ser contrária ao que se observa na natureza onde predadores e presas possuem diferentes tempos para a produção de descendentes [61, 62], isto tem se mostrado como uma boa aproximação.

Uma das primeiras alterações apresentadas para esse problema está na forma como a nova população é gerada pela população parental. Até o capítulo anterior que no caso do modelo de Wright-Fisher a população formada pelos descendentes é obtida realizando um sorteio aleatório com reposição entre os indivíduos da geração parental, onde o número de sorteios é, pelo menos, igual ao número de indivíduos na população. No caso de populações que não possuem um tamanho constante ao longo das simulações, cada indivíduo produz r novos descendentes que são passados para a geração seguinte. Nesta nova dinâmica estes métodos apresentam um problema, pois o tamanho da nova população possui um vínculo dado pelas eqs. 4.2. A maneira utilizada para contornar este problema foi definindo no início de cada geração os novos tamanhos das populações dados pelas eqs. 4.2, em seguida são sorteados aleatoriamente e com reposição os indivíduos da população parental para se reproduzirem e gerar seus descendentes até atingir atingir os números determinados pelas equações presa-predador.

Depois de definidos os novos tamanhos de populações, todo novo indivíduo gerado pela reprodução assexuada, recebe as mutações de seu "pai"acrescido de l novas mutações geradas por uma distribuição de poisson $p(l; U^H)$, onde U^H é a taxa de mutação das presas, o mesmo acontecendo para os predadores, porém com uma taxa de mutação dada por U^P.

Como feito anteriormente, nem todos os novos descendentes continuam na população para formar a próxima geração, pois após o nascimento, os indivíduos passam por uma

seleção da seguinte forma:

- É calculado a adaptabilidade W do indivíduo pelas eq. 4.3 e 4.4
- W é comparado a um número aleatório ξ gerado por uma distribuição uniforme (0,1), onde
  - Se $W \geq \xi$ o indivíduo passa para a próxima geração;
  - Se $W < \xi$ o indivíduo é retirado da população.

Este efeito é importante, pois permite a extinção da população devido ao acúmulo de mutações deletérias em seu genoma [54], efeito esse explicado no capítulo anterior. Outro fator que pode provocar a extinção das populações é uma grande variação do número de presas e de predadores. No caso de uma extinção das presas existe uma extinção dos predadores, devido à forma das eqs. 4.2, porém isto não ocorreria no caso inverso, ficando as presas livres para o crescimento ou decrescimento dependendo do valor de r e κ como descrito anteriormente.

Como última modificação em nossa dinâmica, temos que alterar os tamanhos das populações que são gerados pelas eqs. 4.2. Esses valores de N_H e N_P são reais, enquanto que em nossas simulações eles devem ser inteiros, já que cada indivíduo gera um número inteiro de descendentes. Para resolver este problema foi feita uma "discretização" nos valores gerados pelas eqs. 4.2, de forma que cada população a cada nova geração possui um número inteiro de indivíduos, por exemplo, se em um determinado intervalo de tempo a população de presas e de predadores eram 57.39 e 23.76, após a discretização passavam a ser 57 e 23, respectivamente. Apesar desta mudança o comportamento para as populações de presas e predadores ao longo das gerações não foi significativamente afetado, como pode ser observado nas figuras 4.1, 4.2 e 4.3.

A discretização de um dinâmica contínua não é simples e na maioria dos casos produz comportamentos bastante diferentes entre si. Estudos sobre o efeito de discretizar sistemas contínuos são temas correntes na literatura [63–65], porém é importante salientar que não estaremos discretizando um sistema contínuo com o objetivo de reproduzir uma dinâmica

discreta equivalente, mas sim estamos interessados em uma dinâmica discreta de interação entre populações, afinal, na Natureza todos os indivíduos gerados sempre produzem um número inteiro de indivíduos [64].

Para as eqs. (4.2), a medida que a taxa de reprodução r aumenta, aumenta a não-linearidade desta dinâmica, porém podemos observar que desde $r = 2$, figura 4.1 até $r = 3$, figura 4.3, as amplitudes das oscilações apresentam um comportamento semelhante, e de forma que a dinâmica desenvolvida se adequa muito bem ao estudo da catraca de Muller.

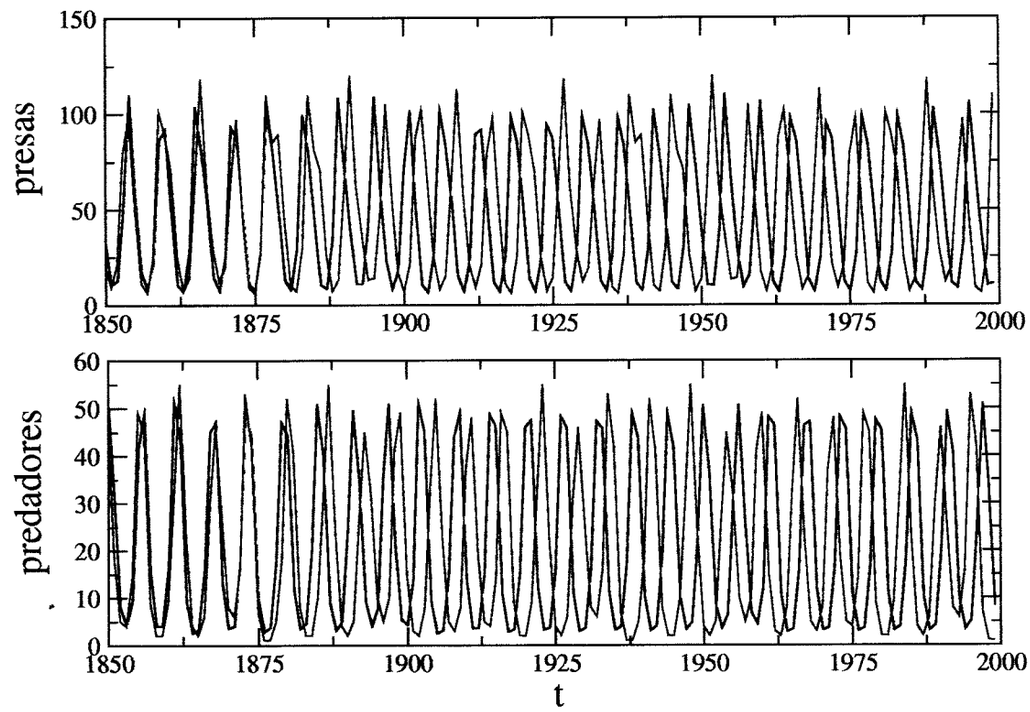

Figura 4.1: Comparação entre a evolução temporal para as populações de presas e predadores nos casos em que N_H e N_P assumem valores reais (curva vermelha) e valores inteiros (curva preta), para os parâmetros $N_H(t = 0) = 50$, $N_P(t = 0) = 15$, $r = 2$, $a = 0.05$ e $\kappa = 100$.

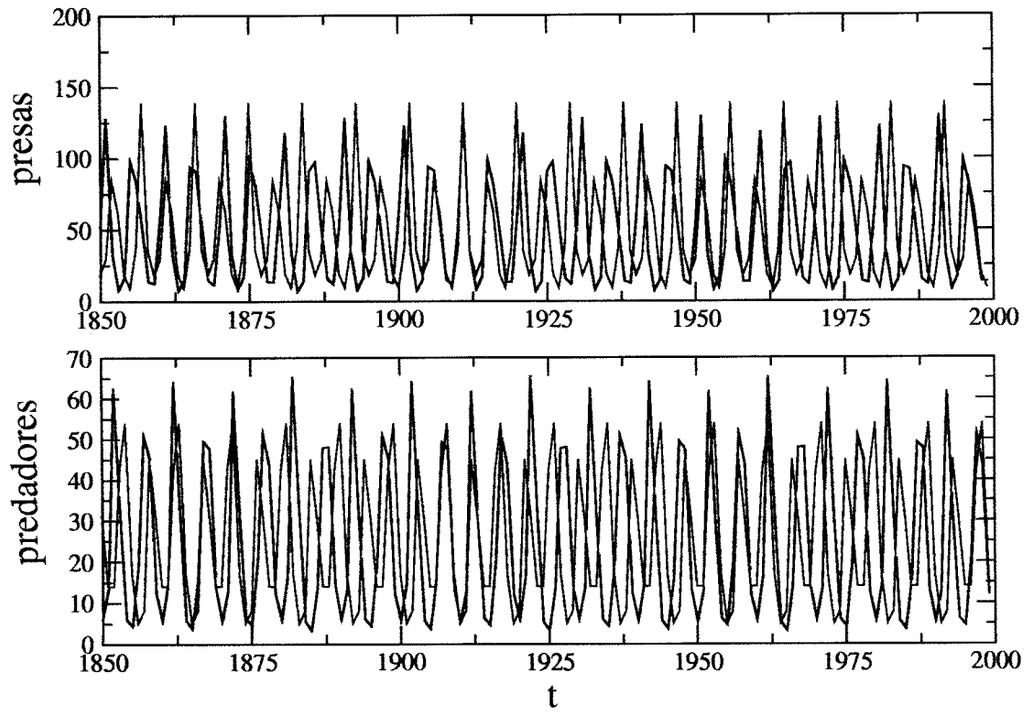

Figura 4.2: Comparação entre a evolução temporal para as populações de presas e predadores nos casos em que N_H e N_P assumem valores reais (curva vermelha) e valores inteiros (curva preta), para os parâmetros $N_H(t=0) = 50$, $N_P(t=0) = 15$, $r = 2.5$, $a = 0.05$ e $\kappa = 100$.

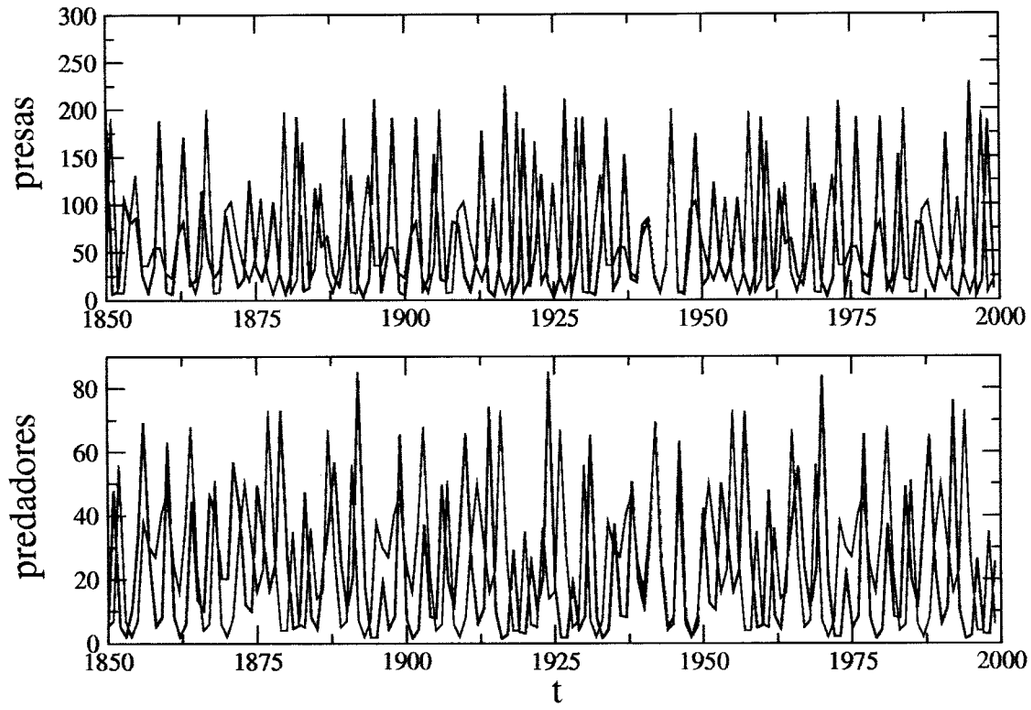

Figura 4.3: Comparação entre a evolução temporal para as populações de presas e predadores nos casos em que N_H e N_P assumem valores reais (curva vermelha) e valores inteiros (curva preta), para os parâmetros $N_H(t=0) = 50$, $N_P(t=0) = 15$, $r = 3$, $a = 0.05$ e $\kappa = 100$.

4.2 Resultados

Para melhor compreender os resultados obtidos por essa nova dinâmica estudamos os comportamentos do número mínimo de mutações (k_m) e o número médio ($< k >$) para diferentes conjuntos de parâmetros, já que este é o procedimento usual neste tipo de análise. Além destas medidas também estudamos o comportamento do tamanho das populações ao longo das gerações. Este estudo foi realizado em duas etapas, sendo o primeiro caso em que a população de presas evolui pela eq. 4.2 sem a presença de predadores ($N_P = 0$) e em seguida para o caso em que a população de predadores é não nula.

Cada uma das figuras de 4.4 à 4.7 apresentam dois gráficos, sendo o gráfico superior os números de mutações mínimo e médio, k_m e $< k >$, respectivamente, enquanto que o gráfico inferior é o tamanho da população ao longo do tempo. Todos os dados são referentes a um valor médio obtido sobre 1000 populações independentes.

Nestas figuras a população de presas evolui sem a presença de predadores e podemos observar que a medida que a taxa de mutação aumenta a população se extingue mais rapidamente, isso é devido ao número de mutações deletérias adquiridas durante a evolução e a conseqüente degeneração de seu genoma. Podemos observar também o aumento linear do número de mutações dos indivíduos em todos os casos, com exceção da figura 4.4, pois esta possui $U^H = 0$.

Na figura 4.4 a população de presas oscila nas primeiras gerações até atingir um estágio de equilíbrio, que é igual a κ, isto acontece devido as ausências de predadores e da pressão da seleção, porém quando consideramos que a população de predadores é não nula, figura 4.8, observamos que o seu comportamento é completamente diferente, pois devido a presença dos predadores a população de presas não apresenta o estado de equilíbrio estacionário, oscilando durante toda a simulação.

As figuras de 4.9 à 4.11, quando comparadas com as figuras 4.5 à 4.7, mostram que as populações de presas se extinguem mais rapidamente quando estão sujeitas a predação, como por exemplo na figura 4.10 onde a população de presas sobreviveu por aproximadamente 350 gerações enquanto que na figura 4.6 a população sobreviveu por 500 gerações.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

Outro ponto interessante é que as populações de presas, na presença de predadores, apresenta uma pequena oscilação ao longo de suas gerações, porém essa amplitude de oscilação é bem menor do que as apresentadas nas figuras 4.1 à 4.3, onde os indivíduos não sofriam mutações ao longo da gerações, enquanto que no caso em que não existe predação, figuras 4.4 à 4.7 a evolução do tamanho das populações praticamente não apresenta oscilações, devido a pressão de seleção.

Quando analisamos o efeito da catraca observamos que tanto para a dinâmica sem predação, figuras 4.5 até 4.7, quanto para a com predação, figuras 4.9 à 4.11, a medida que aumentamos a taxa de mutação a velocidade da catraca fica maior, como esperado, pois existe um acúmulo maior de mutações durante as gerações e como conseqüência a perda da classe dos indivíduos mais adaptados. Porém podemos observar também que a predação aumenta a velocidade da catraca quando comparada com o caso sem predação, um exemplo claro pode ser visto nas figuras 4.5 e 4.9, onde a inclinação das curvas de k_m e de $< k >$ é maior para o caso com predação.

O ponto mais interessante nesta dinâmica pode ser observado nas figuras 4.12, 4.13 e 4.14, onde os gráficos para o tamanho da população de presas e os números de mutações foram sobrepostos. Podemos notar que existe uma relação direta entre o tamanho da população de presas e o número mínimo de mutações, pois enquanto a população está em crescimento a classe de indivíduos mais adpatados não é perdida, produzindo um pequeno patamar na curva k_m, ou seja, parando a catraca de Muller por algumas gerações. Isto pode ser justificado por um comportamento observado na Natureza, mas que não foi utilizado explicitamente em nossa programação.

A medida que a população aumenta fica mais "fácil´´ para os predadores encontrarem as presas e por este motivo estes acabam predando aqueles menos preparados para escapar, ou seja, os menos adaptados, e por este motivo notamos que durante o crescimento da população o número mínimo de mutações permanece constante. Quando a população de presas está decrescendo observamos justamente o efeito contrário, já que a medida que a população começa a diminuir o patamar desaparece e a o número mínimo de mutações cresce rapidamente, isto acontece devido à escassez do número de presas o que obriga uma

predação sobre todos os indivíduos desta população, inclusive sobre os mais adaptados aumentando o número mínimo de mutações k_m até a chegada de um novo patamar, quando este ciclo se repete. Contudo este efeito de patamares logo desaparece quando a amplitude das oscilação diminui, como mostrando na figura 4.14, demonstrando que esta característica da catraca de Muller está ligada a oscilação das populações de presas e predadores e que até então não havia sido apresentada na literatura.

Para finalizar, é importante ressaltar que as oscilações em k_min e $< k >$, alguns momentos antes das populações se extinguirem é simplesmente um problema no cálculo das médias não possuindo nenhum outro significado.

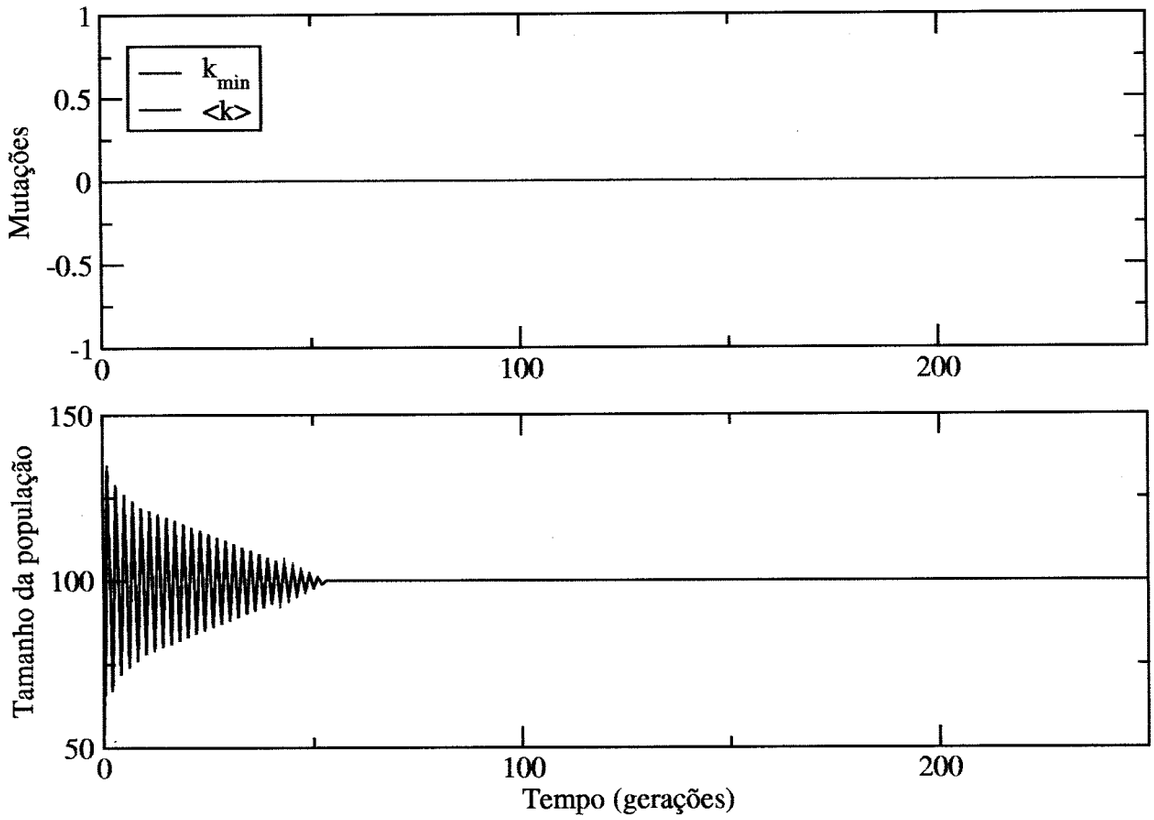

Figura 4.4: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 0$, $s^N = 0$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

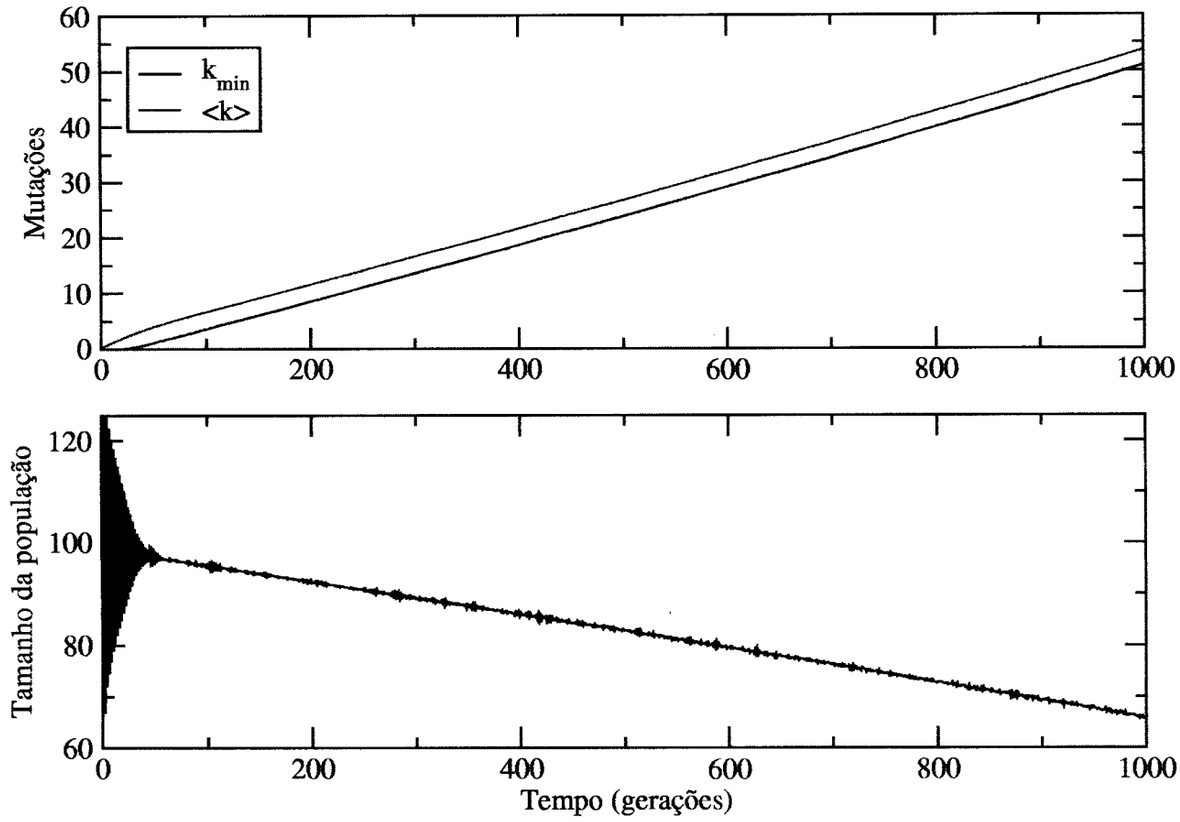

Figura 4.5: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 0.1$, $s^N = 0.0125$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

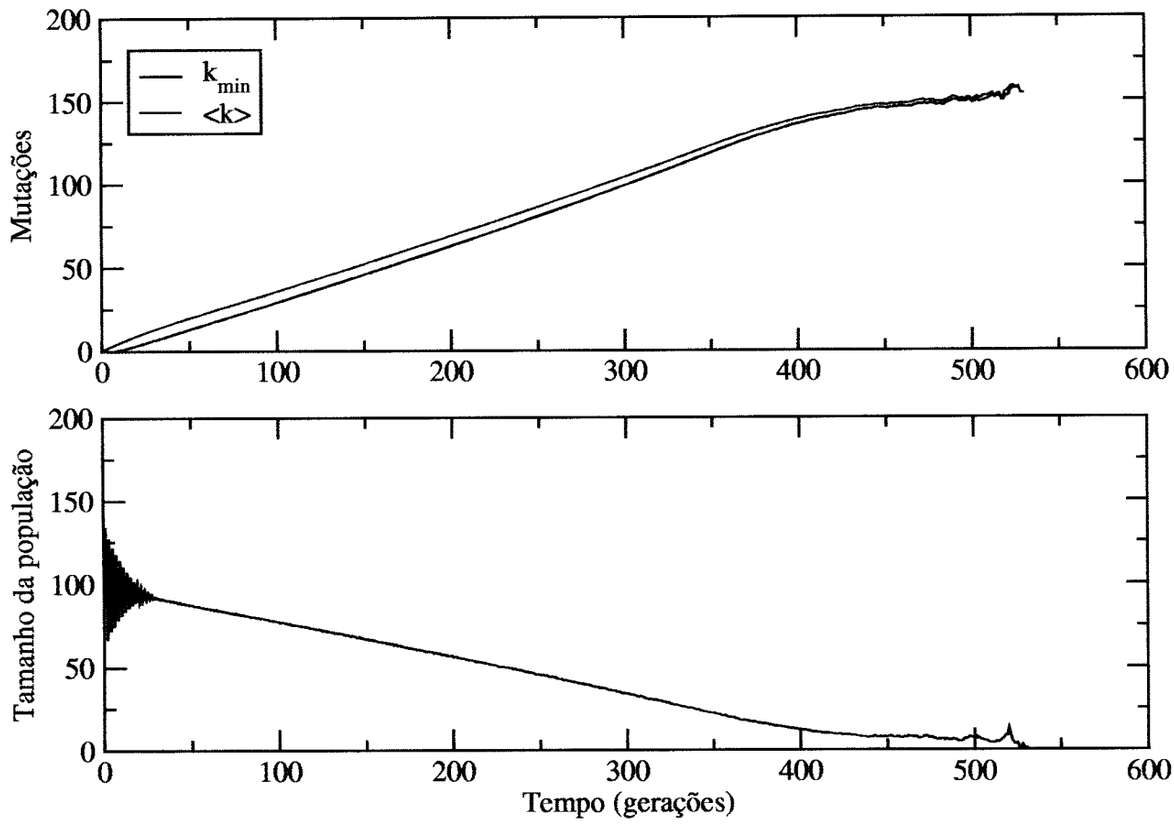

Figura 4.6: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 0.5$, $s^N = 0.0125$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

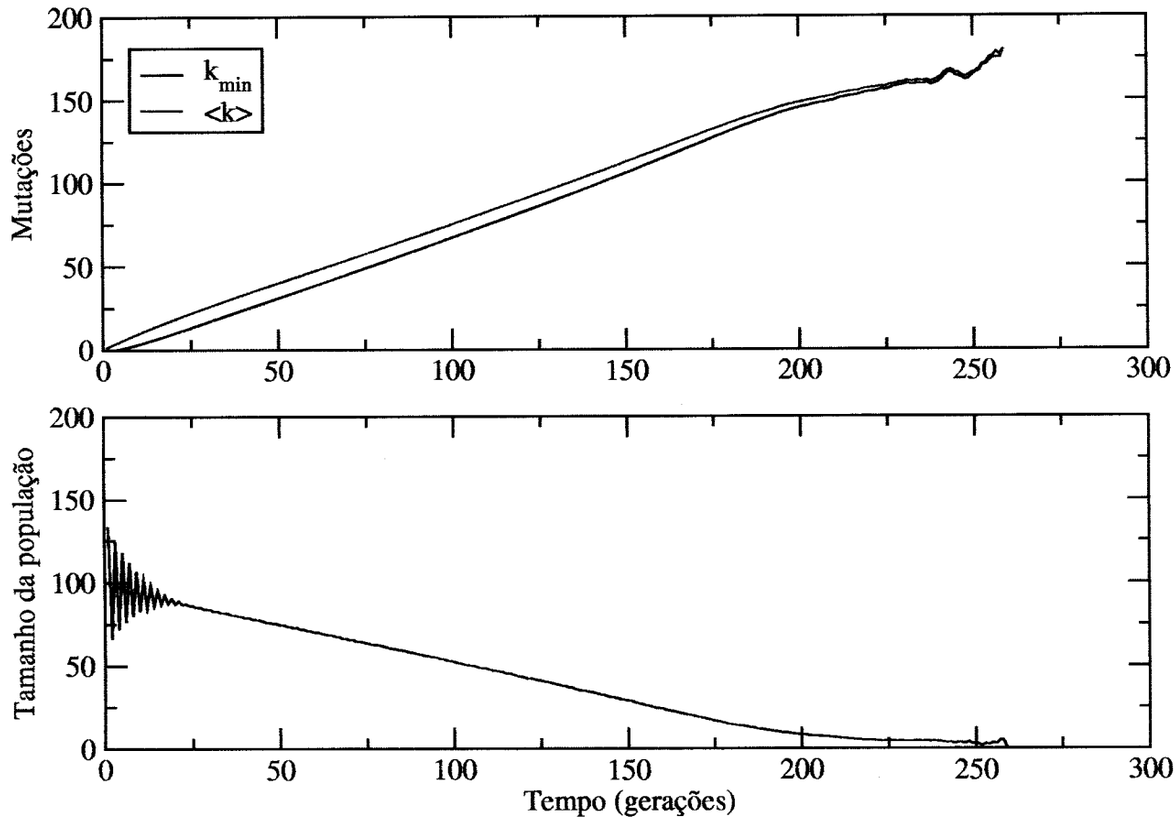

Figura 4.7: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na ausência de predadores. Para $N(t = 0) = 50$, $U^N = 1.0$, $s^N = 0.0125$, $P(t = 0) = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

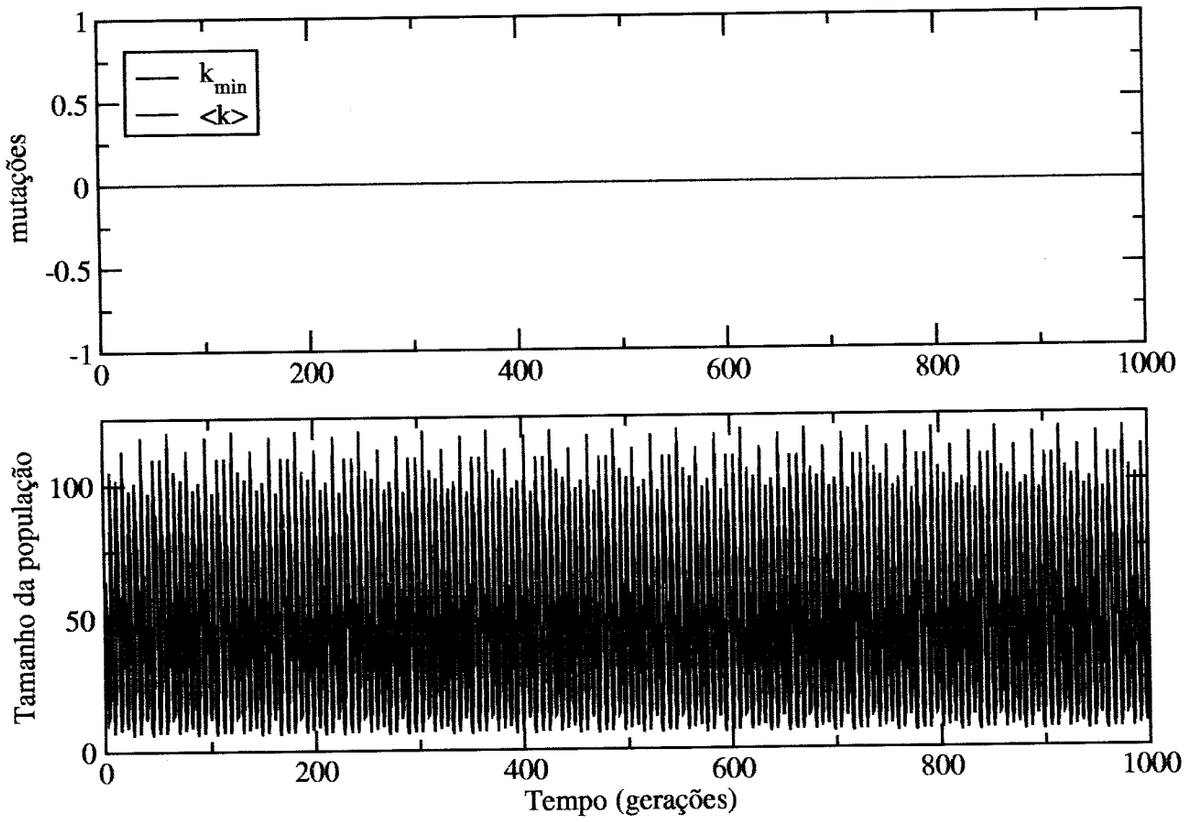

Figura 4.8: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 0$, $s^N = 0$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

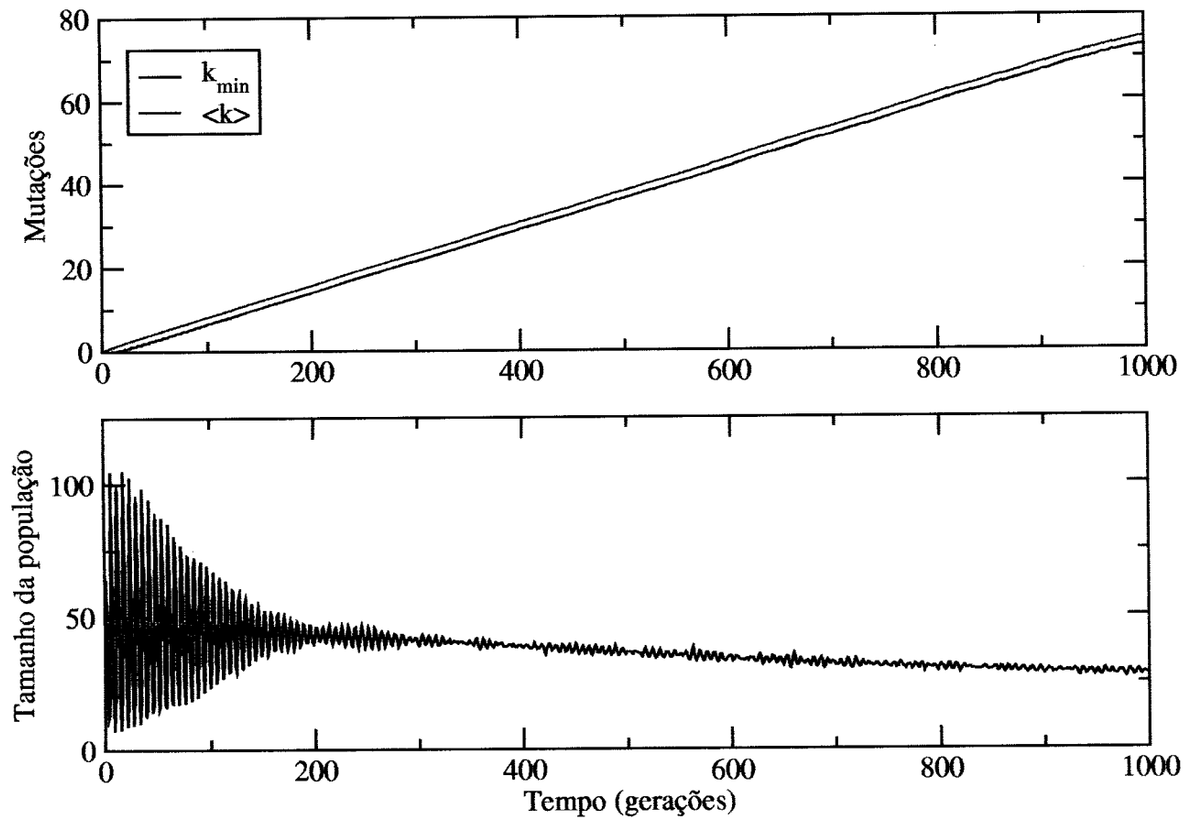

Figura 4.9: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 0.1$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

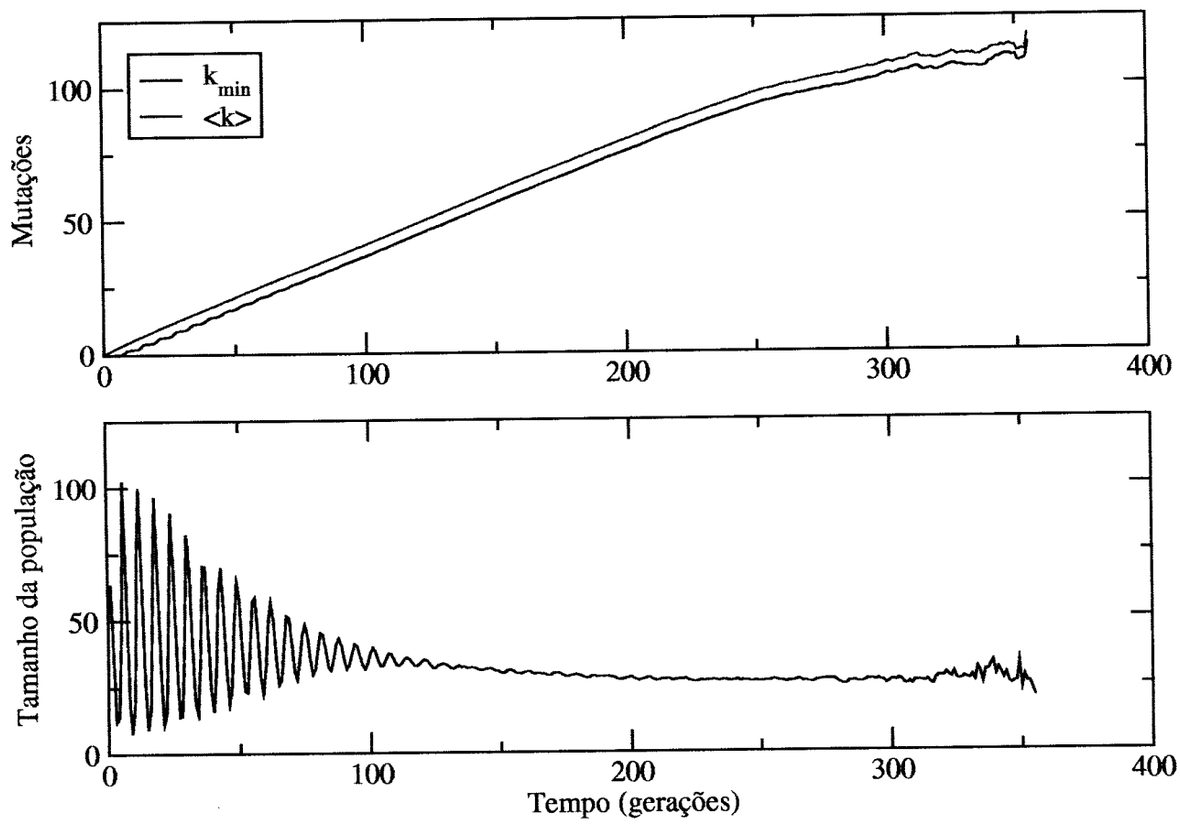

Figura 4.10: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 0.5$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

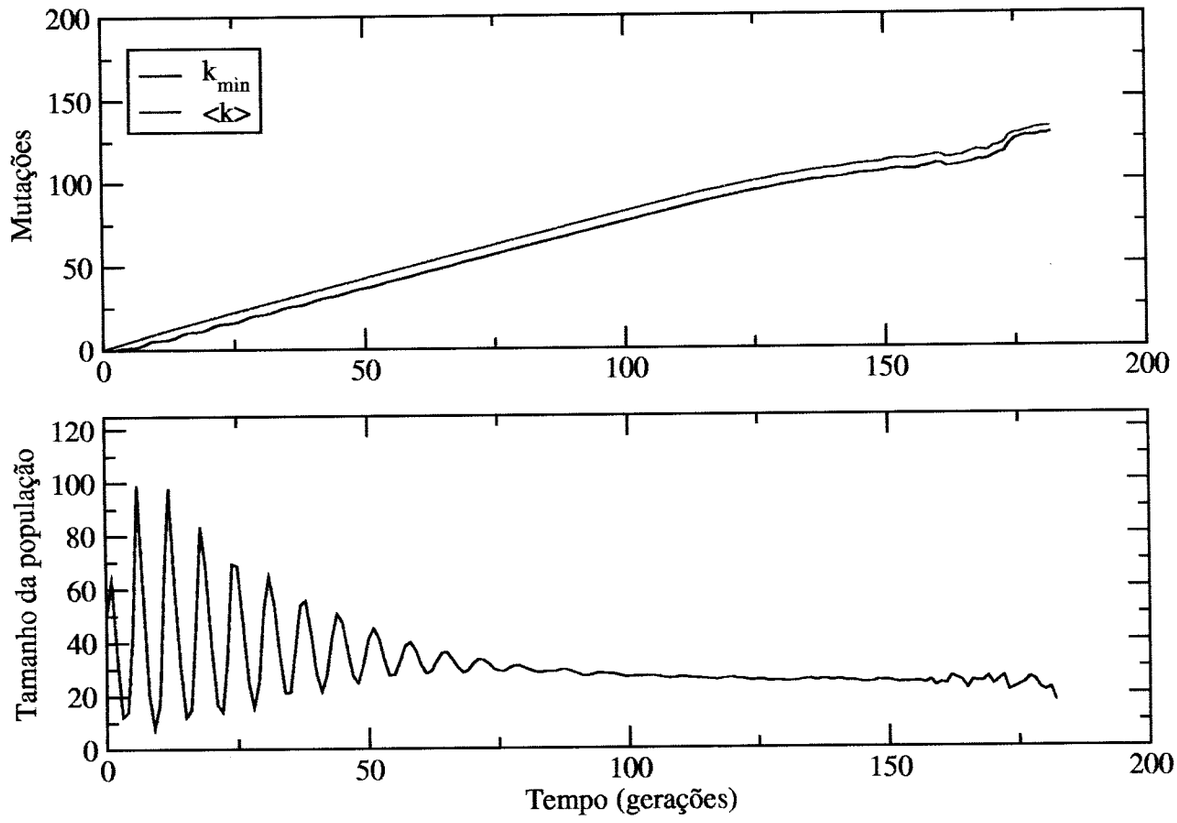

Figura 4.11: Número de mutações mínimo k_m, número médio de mutações $< k >$ e tamanho da população de presas na presença de predadores. Para $N(t = 0) = 50$, $U^N = 1.0$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

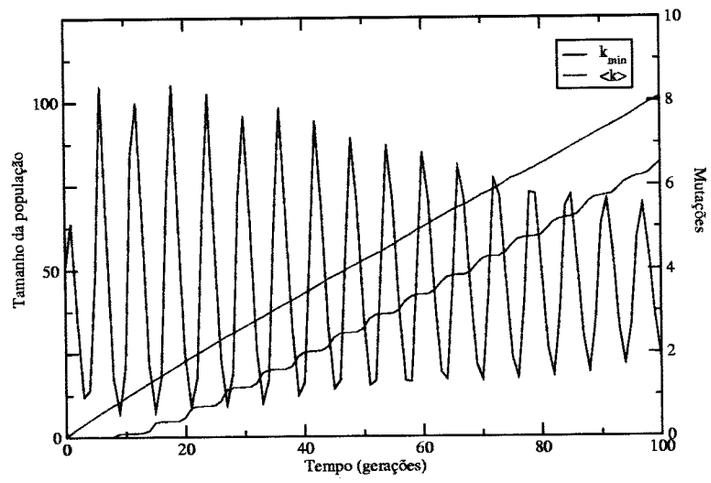

Figura 4.12: Comparação entre o número de mutações e o tamanho da população. $N(t = 0) = 50$, $U^N = 0.1$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

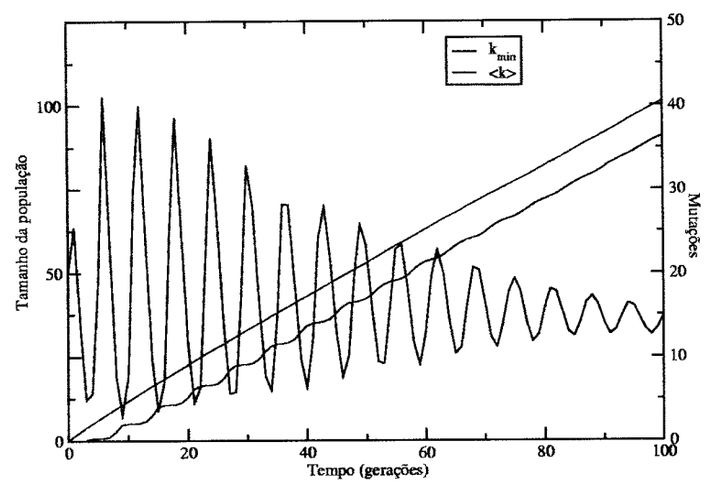

Figura 4.13: Comparação entre o número de mutações e o tamanho da população. $N(t = 0) = 50$, $U^N = 0.5$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

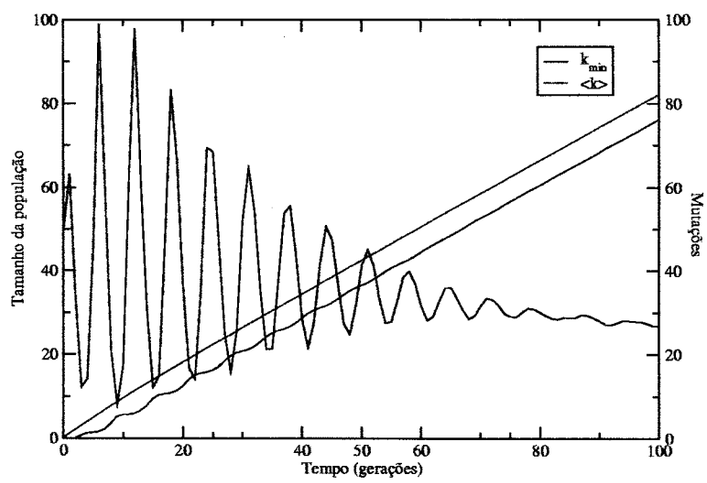

Figura 4.14: Comparação entre o número de mutações e o tamanho da população. $N(t = 0) = 50$, $U^N = 1.0$, $s^N = 0.0125$, $P(t = 0) = 15$, $U^P = 0$, $s^P = 0$, $r = 2$, $a = 0.05$ e $\kappa = 100$. Os valores médios foram obtidos sobre 1000 populações independentes.

Capítulo 5

Árvores genealógicas

Toda a vida pode ser relacionada à ancestrais comuns, dessa forma um importante trabalho da biologia evolucionária é tentar reconhecer padrões evolutivos em árvores genealógicas. Este tipo de análise nos permite inferir informações importantes das populações analisadas, como taxas de extinção, medidas contínuas de especiação, coevolução, entre outras e por poder ser aplicado tanto à evolução pré-biótica quanto à macro-evolução.

Neste capítulo serão discutidas as principais características das árvores genealógicas como sua nomenclatura e classificação na seção5.1, análise da genealogia utilizando distância entre indivíduos de uma árvore, seção 5.2 e as propriedades estatísticas de simetria na seção 5.3.

O problema sobre topologia das árvores será apresentado em populações cujos indivíduos possuem um genoma finito. Esta escolha foi feita para apresentar alguns conceitos como o de mutação reversa e distância de Hamming, sendo que sua generalização para o caso do genoma infinito é direta e não acarreta maiores problemas.

5.1 Definições básicas

Árvores são estruturas matemáticas usadas para representar a história evolucionária*

*Novas propostas para o estudo da história evolucionária entre organismos estão sendo propostos utilizando-se de inúmeras técnicas em conjunto para um melhor entendimento do problema evolutivo.

IFSC-USP SERVIÇO DE BIBLIOTECA E INFORMAÇÃO

de um grupo de seqüências ou indivíduos de maneira que características como simetria e quantidade de nós possam ser usadas para se obter diferentes tipos de medidas referentes a sua evolução. A constituição básica de uma árvore pode ser vista na figura 5.1.

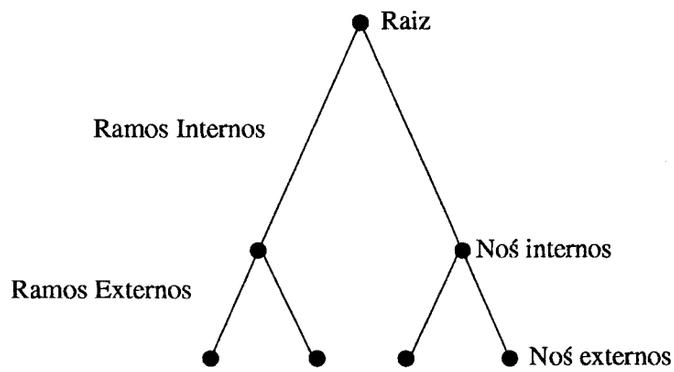

Figura 5.1: Exemplo de uma árvore genealógica e suas partes integrantes. Neste caso a evolução temporal se dá de cima para baixo, sendo os nós mais próximos do topo da árvore mais antigos que os nós que se encontram mais próximos de sua base.

- **Raiz:** Representa o ancestral comum mais recente à todos os indivíduos que se encontram na base da árvore.
- **Nó:** Identificação de um indivíduo em uma árvore genealógica.
- **Ramo:** Parte da árvore que conecta dois nós, sendo que estes ramos podem ser internos ou externos.
  - **Ramos Internos:** Ramos que conectam nós internos.
  - **Ramos Externos:** Ramos que conectam nós internos a nós externos.

Os estudos desenvolvidos neste capítulo referem-se a modelos de árvores binárias, que são árvores formadas por populações onde cada indivíduo gera exatamente dois novos descendentes, ou seja, cada nó interno apresentará exatamente três ramos, sendo que um destes ramos conecta o indivíduo ao seu ancestral mais recente enquanto os outros dois o

[66, 67]

conectam aos seus descendentes. Podemos ter árvores cujos nós apresentem mais de três ramos, como mostrado na figura 5.2, isto significa que cada indivíduo pode dar origem a mais de dois descendentes, porém nesta tese consideramos somente o caso de árvores binárias.

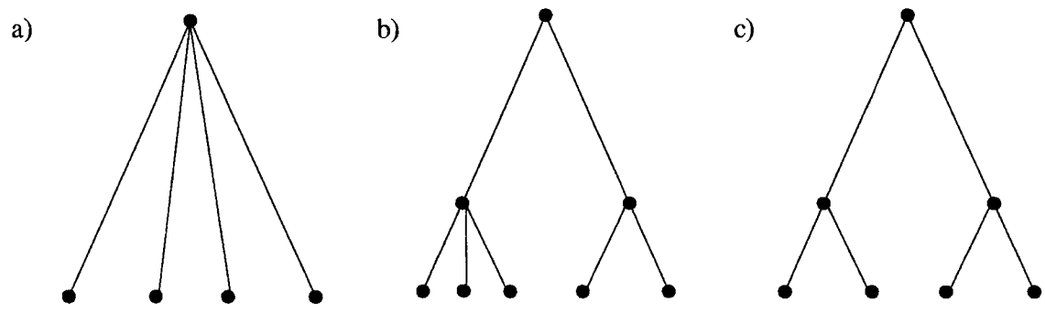

Figura 5.2: Exemplos de árvores binárias e não-binárias. As figuras a) e b) representam árvores não-binárias, pois possuem um nó com mais de três ramos, enquanto a árvore c) é um exemplo de árvore binária.

Distintos tipos de árvores podem ser usados para representar diferentes aspectos evolucionários, assim quando estamos interessados em descrever simplesmente a relação entre os indivíduos de uma determinada população utilizamos os cladogramas, que são aquelas árvores genealógicas que costumam ser feitas para mostrar o grau de parentesco e descendência em uma família. Contudo se quisermos detalhar os números de mutações sofridos durante as gerações ou o tempo de surgimento de cada novo indivíduo, utilizamos as árvores aditivas e ultramétricas, respectivamente. Assim podemos classificar os três tipos básicos de árvores como:

1. **Cladograma:** Utilizado simplesmente por mostrar a relação entre os indivíduos ao seus ancestrais e descendentes. Figura 5.3 a).

2. **Aditiva:** Além de mostrar a relação entre os indivíduos, como na cladograma, este tipo de árvore contém uma informação adicional relacionada aos comprimentos dos ramos, onde o seu tamanho está associado ao número de mutações adquiridas por um indivíduo e que não estão presentes em seu ancestral. Figura 5.3 b).

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

3. **Ultramétrica:** Estas não contém informações sobre o número de mutações, porém mostram o tempo em que os indivíduos surgiram ao longo da evolução da população. Figuras 5.3 c) e 5.4.

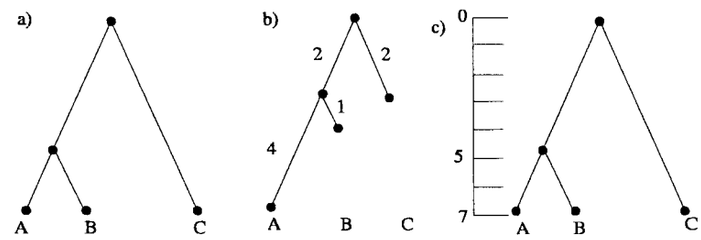

Figura 5.3: Tipos diferentes de árvores. a) Cladograma. b) Aditiva. c) Ultramétrica.

Na figura 5.4 podemos ver uma árvore ultramétrica mais detalhadamente. Os ramos i_n representam os ramos internos, enquanto e_n representam os ramos externos. Desta forma podemos associar um comprimento a cada ramo interno e externo, ou seja, o ramo i_1 tem um comprimento t_1, já o ramo i_2 tem um comprimento que é a soma de t_1, t_2 e t_3. O mesmo raciocínio deve ser aplicado aos ramos externos: e_5 com um comprimento t_4, enquanto e_1 tem comprimento $t_3 + t_4$.

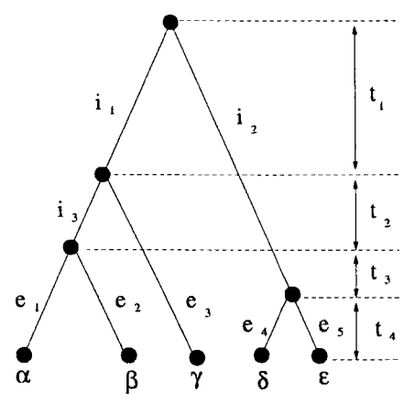

Figura 5.4: Árvore ultramétrica para uma amostra com 5 indivíduos.

Outro aspecto importante que podemos observar na figura 5.4 é o tempo de coa-

lescência $T^{\alpha\beta}$. Este tempo representa o número de gerações entre os indivíduos α e β e seu último ancestral comum. Portanto $T^{\alpha\beta}$ é $t_3 + t_4$, enquanto $T^{\alpha\gamma}$ é $t_2 + t_3 + t_4$.

Existem outros tipos e formas de classificação das árvores genealógicas, bem como denominações diferentes para uma mesma característica, pois ainda não há uma uniformização para esta nomenclatura, porém estas informações não são relevantes para a compreensão do problema aqui apresentado e por isso não são abordadas nesta tese.

Alguns resultados analíticos para o problema da coalescência em evolução neutra foram obtidos e estão descritos no Apêndice A.

5.2 Análise da genealogia utilizando distância de Hamming

Para se trabalhar com a maioria dos modelos evolutivos é necessário introduzir uma medida que diferencie pares de seqüências. Isto pode ser feito de diversas formas, como por exemplo, utilizando-se o conceito de distância de Hamming [68, 69]. Este conceito baseia-se na diferença entre dois dígitos (duas bases) que se localizam na mesma posição (sítios) do genoma em duas seqüências (indivíduos) diferentes, ou seja, considerando que estas seqüências são formadas por um número finito de dígitos, podemos definir a unidade da distância ($d_{\alpha,\beta}$) como sendo a diferença de um dígito, na mesma posição, entre os indivíduos α e β. Dessa forma se duas seqüências possuírem os mesmos dígitos nas mesmas posições ao longo de todo seu genoma, dizemos que a distância de Hamming entre elas é zero, enquanto que, se apenas o i-ésimo sítio for diferente entre os dois genomas, dizemos que a distância entre elas é igual a 1, e assim sucessivamente.

Diferentemente dos outros capítulos em que os estudos foram realizados em uma população onde os indivíduos eram formados por genomas infinitos, neste caso estudaremos o comportamento da população onde os indivíduos são compostos por um número finito de dígitos em seu genoma. Isto foi feito por estarmos interessados em estudar a genealogia entre indivíduos baseados na distância de Hamming, de maneira que optamos por "construir" indivíduos onde poderíamos saber com exatidão qual a posição do dígito alterado.

Este procedimento também poderia ser feito com cadeias infinitas e admitir que toda nova mutação acontece em um sítio que antes não havia mutado, porém isso não se enquadra aos objetivos deste estudo que se utliza do conceito de distância de Hamming em cadeias finitas.

Neste capítulo vamos considerar uma população haplóide formada por N seqüências, onde cada seqüência α pode ser representada como um número finito de L sítios, sendo que cada sítio possui um único dígito, da forma $\{l_1, l_2, \cdots, l_L\}$, onde, por exemplo, para cadeias binárias l_i pode assumir apenas dois valores, enquanto nas cadeias quaternárias cada sítio pode conter um dígito que assume até quatro valores diferentes, como por exemplo A, G, C e T.

Várias são as formas de se analisar uma genealogia, iniciaremos este estudo pelo tempo de coalescência de dois indivíduos. Este tempo representa o número de gerações entre esses indivíduos e seu último ancestral comum, também conhecido MRCA (do inglês *Most Recent Common Ancestor*).

Para iniciar a análise do tempo de coalescência entre dois indivíduos, podemos construir uma matriz baseada nos tempos de coalescência de toda a população e atualizá-la a cada nova geração, para isso devemos definir uma matriz T, com a forma $T_t^{\alpha\beta}$, onde cada elemento $\alpha\beta$ representa o tempo de coalescência entre as seqüências α e β [56], enquanto a sua atualização será da seguinte forma

$$T_{t+1}^{\alpha\beta} = T_t^{G(\alpha)G(\beta)} + 1, \tag{5.1}$$

onde $G(\alpha)$ representa a geração parental do indivíduo α. Assim esta equação relaciona os indivíduos da matriz no tempo t com os indivíduos da matriz no tempo $t+1$.

Este estudo apresenta um problema, pois a medida que a população aumenta de tamanho, também aumenta o tamanho da matriz T, de maneira que para uma população de tamanho N a matriz terá $N \times N$ elementos, porém esta matriz apresenta a característica de ser simétrica já que $T_t^{\alpha\beta} = T_t^{\beta\alpha}$, o que facilita os cálculos dos elementos.

Partindo-se da hipótese que todas as seqüências iniciais são idênticas, podemos definir

a distância de Hamming entre uma seqüência qualquer da população em uma dada geração e uma seqüência da populaçao inicial. Na figura 5.5 é apresentada essa distância para o modelo de L sítios com relevo multiplicativo. Nesta figura é interessante notar que essa distância é maior em cadeias quaternárias do que em cadeias binárias e isto pode ser entendido devido ao fato de em cadeias finitas existe o processo de mutações reversas. Uma mutação reversa é aquela que modifica um dígito já mutado em seu dígito original. Assim em cadeias binárias quando ocorre uma mutação, a distância de Hamming da seqüência inicial aumenta uma unidade, porém quando ocorre uma mutação reversa, esta distância diminui uma unidade, pois essa nova mutação gerou o mesmo dígito da seqüência inicial. No caso das cadeias quaternárias o efeito de mutações reversas é menor, porque um sítio que já tenha sido mutado uma vez, ao sofrer nova mutação, terá probabilidade $1/3$ de voltar ao dígito inicial e $2/3$ de mutar para um outro dígito, por este motivo a distância de Hamming no caso das cadeias quaternárias tende a ser maior [68, 70].

Também podemos observar que existe uma saturação das distâncias médias $d_{\alpha\beta}$ para seqüências de sítios finitos, enquanto que para as seqüências infinitas a distância cresce linearmente com o tempo t, ou seja a distância é dada por Ut, sendo este resultado bem conhecido e pode ser entendido como o efeito da catraca de Muller, como apresentado anteriormente (ver apêndice A para solução analítica) e desta forma podemos dizer que a probabilidade de que uma mutação ocorra em um sítio que já sofreu mutação é nula no caso $L = \infty$.

Nas figuras 5.6 e 5.7 são apresentados os tempos médios de coalescência para o modelo de L sítios em cadeias binárias e quaternárias, respectivamente, em uma dinâmica com relevo multiplicativo. Estes gráficos foram obtidos utilizando-se a eq. 5.1 e mostram que nem o tamanho da seqüência L nem o tamanho do alfabeto (binário ou quaternário) afetam os tempos de coalescência (lembramos, é claro, que esses tempos dependem dos parâmetros N, s e U, que são o tamanho da população, a taxa de seleção e da taxa de mutação, respectivamente). Isto se deve ao fato de que esta grandeza independe do alfabeto usado, já que esta medida faz referência apenas aos ancestrais de cada seqüência.

Também foram realizadas análises das distâncias entre os indivíduos da população

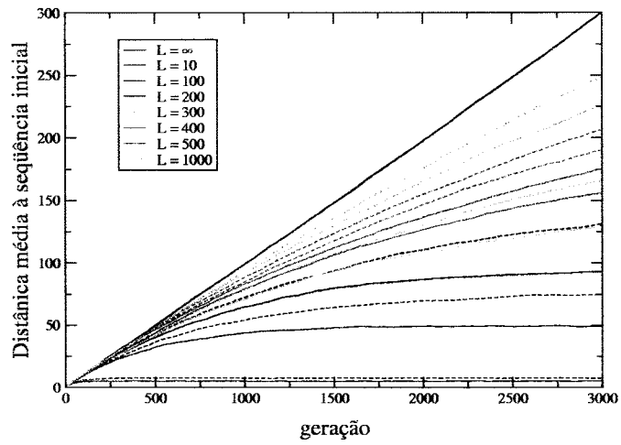

Figura 5.5: Distâncias médias em relação à seqüência inicial para seqüências com L sítios. As linhas tracejadas representam seqüências quaternárias, enquanto as linhas cheias representam seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$.

onde podemos observar, figuras 5.8 e 5.9, apesar das flutuações, que todas as distâncias tendem a um limite, que aumenta com o número de sítios. Em particular, para $L = \infty$ e $s = 0$ (evolução neutra) esse limite é $2NU$.

Quando observamos o comportamento da distância $d_{\alpha\beta}$ e do tempo médio de coalescência $T^{\alpha\beta}$, figuras 5.11 e 5.10 respectivamente, observamos que os resultados analíticos obtidos para o modelo neutro não se aplicam nesse caso. Fica bem claro que o efeito da seleção modifica o comportamento dessas grandezas, cujos valores acabam ficando abaixo do esperado para o caso neutro.

Podemos observar que os resultados de seleção neutra $s = 0$ apresentados nesta seção, figuras 5.6-5.9, estão de acordo com os resultados analíticos obtidos no apêndice A. Também é possível observar que a distância média entre os indivíduos se aproxima do caso dado pela evolução neutra a medida que o número dos sítios aumenta, figuras 5.8 e 5.9, sendo que o caso que $L = 10$ apresenta valores bem menores que o caso neutro.

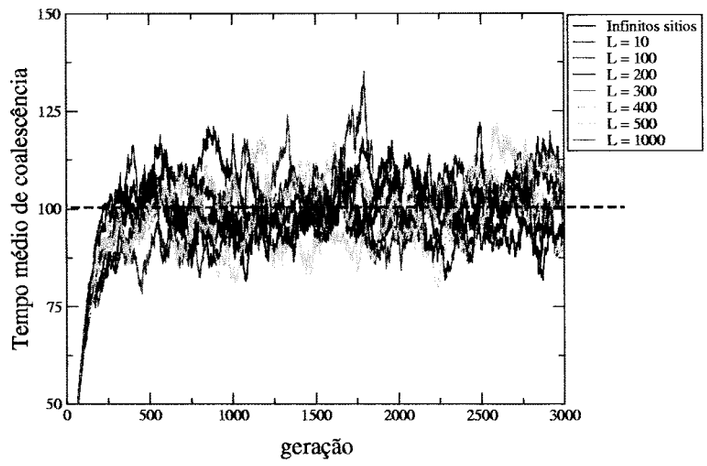

Figura 5.6: Tempo médio de coalescência $T^{\alpha\beta}$ para o modelo de L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$.

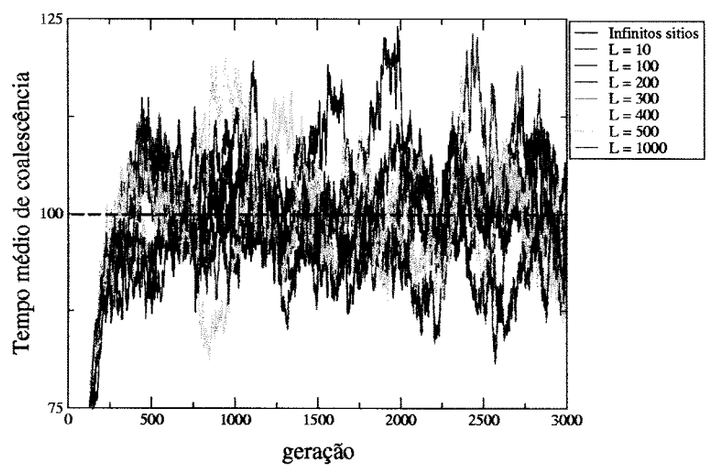

Figura 5.7: Tempo médio de coalescência $T^{\alpha\beta}$ para o modelo de L sítio sem seqüências quaternárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$.

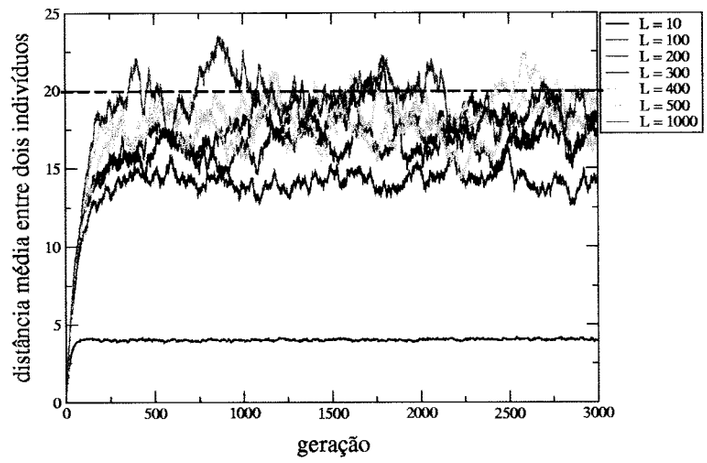

Figura 5.8: Distância média entre dois indivíduos $d_{\alpha\beta}$ para o modelo L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$.

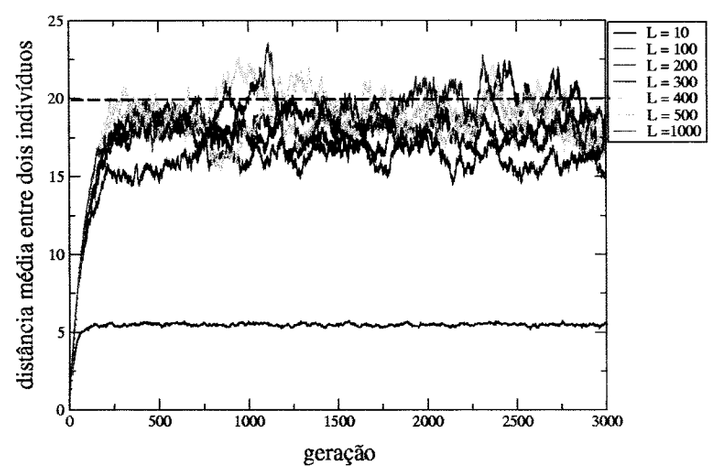

Figura 5.9: Distância entre dois indivíduos $d_{\alpha\beta}$ para o modelo de L sítios em seqüências quaternárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$.

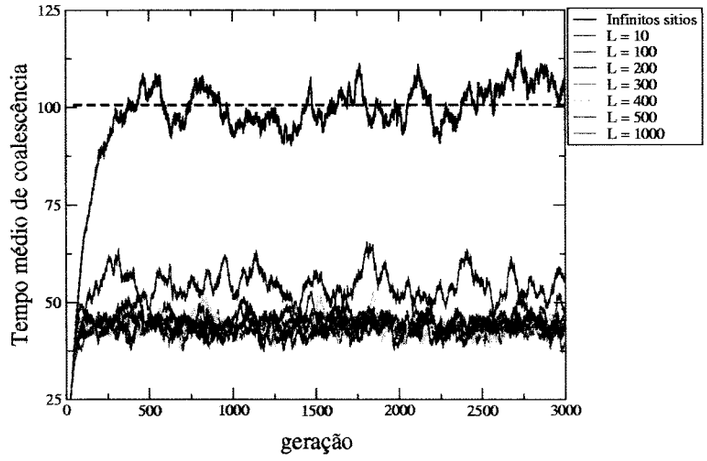

Figura 5.10: Tempo médio de coalescência $T^{\alpha\beta}$ para o modelo de L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0.1$ e taxa de mutação $U = 0.1$. A linha tracejada representa o resultado analítico para $L = \infty$.

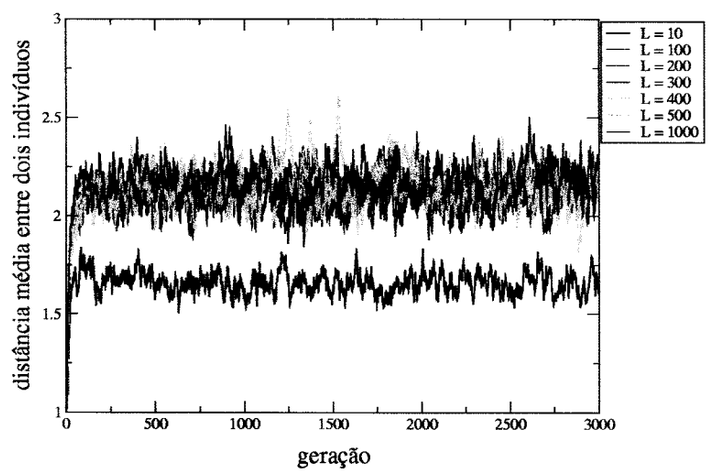

Figura 5.11: Distância média entre dois indivíduos $d_{\alpha\beta}$ para o modelo L sítios em seqüências binárias. Os parâmetros são: tamanho da população $N = 100$, taxa de seleção $s = 0.1$ e taxa de mutação $U = 0.1$.

5.3 Propriedades Estatísticas dos Ramos Internos e Ramos Externos.

Ao observarmos uma amostra aleatória de dígitos (genes) em uma dada população, encontraremos mutações que há muito estão presentes nos indivíduos, tendo sido adquiridas em tempos remotos de sua evolução, e que portanto estarão presentes nos ramos internos da genealogia. Por outro lado as novas mutações estarão presentes nos ramos externos. Esses efeitos modificam os indivíduos da população e, dependendo da taxa com que eles ocorram, podem ajudar-nos a inferir informações úteis quanto a evolução da população.

Podemos então tentar inferir a seleção em uma população analisando os indivíduos antes e após a seleção. Porém este tipo de análise apresenta grandes problemas, pois mesmo que ela seja feita de geração em geração, o grau de correlação das seqüências (indivíduos) entre gerações subseqüentes é grande o que dificulta uma análise coerente. Outro fator é a demanda computacional necessária para analisar uma espécie por várias gerações, já que uma investigação desse porte necessita de alta capacidade de processamento. Como as informações sobre a seleção não são simples de serem obtidas via observação direta, podemos fazer isso de forma indireta, analisando o comprimento dos ramos das árvores genealógicas. Uma maneira de fazer esta inferência foi proposta por Fu e Li [71] em 1992. O método proposto por Fu e Li baseia-se nos comprimentos dos ramos internos e externos das árvores genealógicas e em como os novos ramos e novos genes são adicionados ao longo do tempo (gerações).

Para definir as propriedades dos ramos são necessárias algumas considerações. Primeiro vamos escolher aleatoriamente uma amostra de n seqüências de uma população haplóide com tamanho N. Vamos supor que todas as mutações sejam seletivamente neutras e que não ocorra recombinação entres as seqüências. Desta forma os indivíduos escolhidos na amostra farão parte de uma única árvore genealógica, como mostrado na figura 5.4. Fu e Li encontraram duas relações que descrevem o tamanho médio dos ramos dentro de uma árvore genealógica no regime neutro com $L = \infty$, sendo que para os ramos externos é

$$\overline{E_n} = 2N, \tag{5.2}$$

onde E_n é a soma de todos os ramos externos[†]. O mais importante desta equação é que a soma dos ramos externos independe do tamanho n da amostra escolhida, sendo em média igual tanto para uma amostra de tamanho 4 quanto para uma de tamanho 40. O mesmo não acontece para os ramos internos que possui uma dependência explícita no tamanho da amostra, da forma

$$\overline{I_n} = 2N(a_n - 1), \tag{5.3}$$

onde I_n é a soma dos ramos internos e

$$a_n = \sum_{k=1}^{n-1} \frac{1}{k}. \tag{5.4}$$

Baseados nas medidas analíticas dos comprimentos dos ramos externos e internos elaboradas por Fu e Li, estudamos o efeito da seleção em as árvores genealógicas obtidas através de experimentos computacionais. Nestes experimentos uma amostra de 10 indivíduos era escolhida aleatoriamente em uma população de N indivíduos e sua genealogia reconstruída de forma a obtermos a árvore genealógica correspondente com todos os comprimentos dos ramos internos e externos. Esta reconstrução foi feita com base na matriz dos tempos de coalescência, eq. 5.1. O mesmo procedimento foi repetido até se obter 1000 genealogias, com um intervalo de N gerações entre cada medida[‡], para estudarmos como o efeito da seleção, representado aqui pelos relevos de replicação, e o tamanho finito das seqüências L afetam as previsões das eqs. 5.3 e 5.4, que são válidas para o regime neutro ($s = 0$) e $L = \infty$.

Em cada simulação foram utilizados populações compostas por 50 indivíduos, cujo tamanho permanecia constante ao longo de sua evolução, de maneira que os valores

[†]A demonstração deste resultado é apresentada no Apêndice B

[‡]O intervalo de N gerações foi escolhido para tentar evitar o efeito de correlação entre as amostras

analíticos para o comprimentos dos ramos externos é de $\overline{E_n} = 100$ enquanto para os ramos internos é $\overline{I_n} \approx 183$.

Golding [72] verificou que a seleção não poderia ser inferida pela árvore genealógica para o relevo de pico único, resultado que foi obtido também em nossas simulações, figuras 5.12 e 5.13, onde os gráficos apresentam condições insuficientes para identificar qualquer efeito da seleção nos comprimentos dos ramos. Podemos notar nestas figuras que o valor médio dos ramos internos e externos se aproxima muito dos valores analíticos dados pela eqs. 5.3 e 5.4 para todos os conjuntos os valores utilizados para a taxa de mutação e portanto este teste deveria ser substituído por outros testes estatísticos [73–76] que pudessem obter um melhor resultado.

Porém, quando consideramos o mesmo problema, mas com o relevo multiplicativo, figuras 5.14 e 5.15, o teste proposto por Fu e Li apresenta uma mudança significativa no comprimento dos ramos, mais acentuada nos ramos internos que nos ramos externos, principalmente para altas taxas de mutações. Desta forma este teste poderia ser utilizado para identificar quando uma dada população está sobre o efeito da pressão de seleção, desde que cada nova mutação seja mais prejudicial ao indivíduo que as mutações anteriores como considerado no caso deste tipo de relevo.

Identificado que os comprimentos dos ramos variavam para o caso do relevo multiplicativo, algumas perguntas surgem naturalmente como por exemplo, qual seria ou quais seriam os efeitos nas árvores genealógicas quando elas são construídas a partir de diferentes tamanhos de amostras colhidas na população? E no caso de uma alta taxa de mutação? O mesmo acontece quando variamos a taxa de seleção do regime de seleção neutra ($s = 0$) até o caso de seleção extrema ($s = 1$)?

Para tentar responder a primeira pergunta fizemos simulações onde são apresentadas as curvas para os comprimentos dos ramos externos e internos, como função do tamanho da amostra (n), figuras 5.16 e 5.17. Podemos observar que para o caso de seleção neutra $s = 0$ obtivemos as previsões feitas pelas eqs. 5.3 e 5.4, porém quando há seleção no processo evolutivo, os comprimentos dos ramos ficam menores à medida que a seleção aumenta, o que evidência que os novos indivíduos da população que possuem uma alta

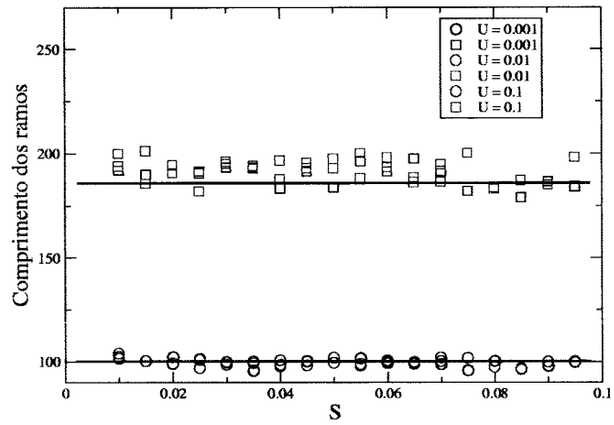

Figura 5.12: Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de finitos sítios ($L = 50$) em seqüências binárias com relevo de pico único e tamanho da população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$.

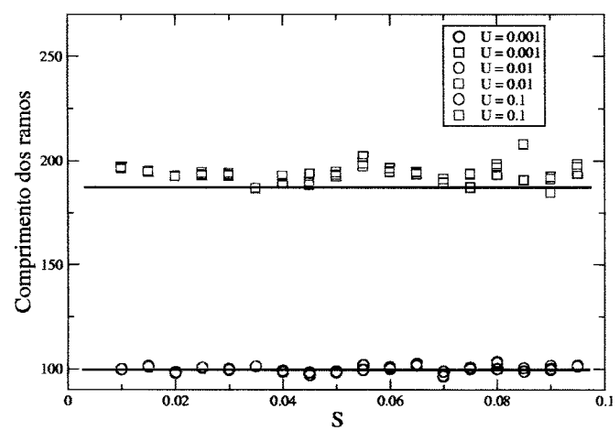

Figura 5.13: Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo de pico único e tamanho de população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$.

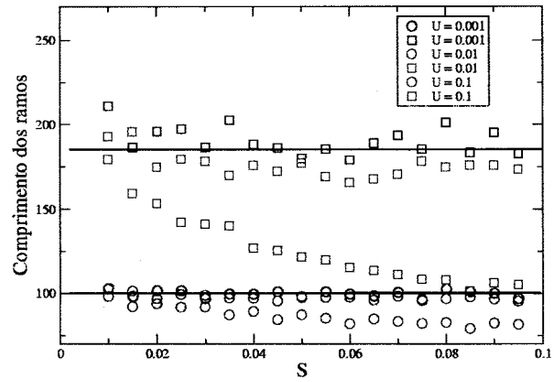

Figura 5.14: Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de finitos sítios ($L = 50$) em seqüências binárias com relevo multiplicativo e tamanho da população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$.

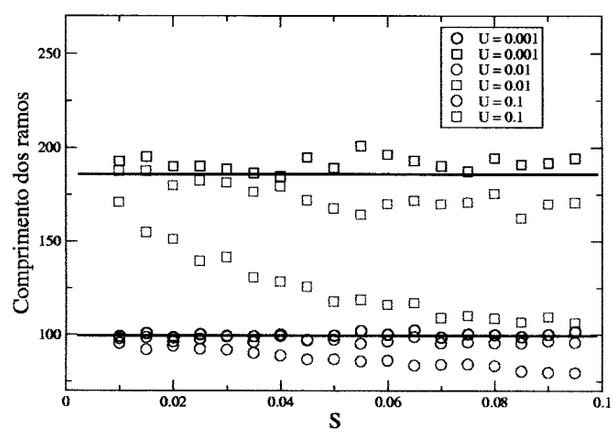

Figura 5.15: Comprimento dos ramos internos(□) e ramos externos(○) para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo e tamanho de população $N = 50$. As linhas cheias representam os resultados analíticos para o regime de seleção neutra $s = 0$.

taxa de mutação logo são eliminados da população.

Outro fator importante é que o comportamento previsto pela eq. 5.3 de que os comprimentos dos ramos externos são independentes do tamanho n da amostra, deixa de ser verdadeiro no caso do relevo multiplicativo quando deixamos de trabalhar no regime neutro e passamos para o regime com seleção, isso porque podemos observar que quanto maior o tamanho da amostra maior é o comprimento dos ramos externos para uma mesma taxa de seleção, figura 5.16.

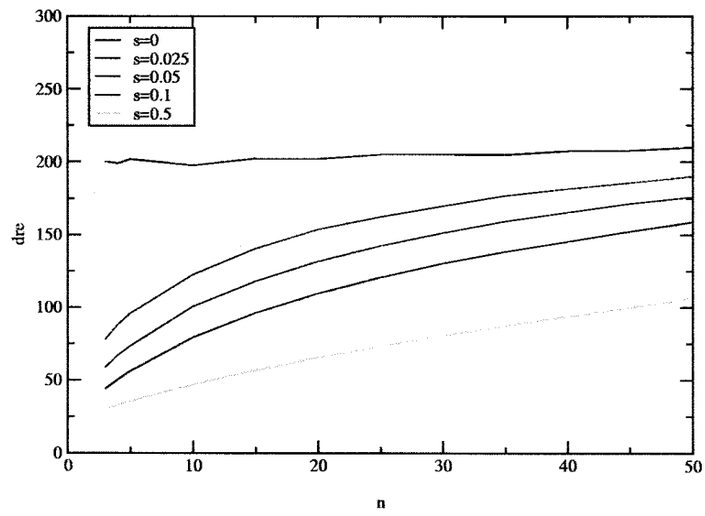

Figura 5.16: Comprimento dos ramos externos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 e taxa de mutação $U = 1.0$.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

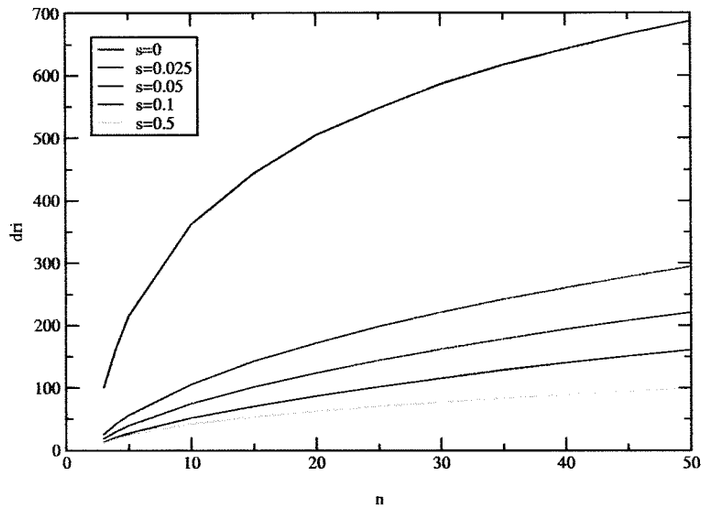

Figura 5.17: Comprimento dos ramos internos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 e taxa de mutação $U = 1.0$.

Para completar o estudo do efeito da seleção nos comprimentos dos ramos, fixamos um tamanho de amostra e variamos o parâmetro seletivo desde 0 até 1 para diferentes taxas de mutação.

Podemos verificar nas figuras 5.18 e 5.19 que o efeito da seleção não é monotônico e que este se torna mais evidente quanto maior a taxa de mutação, de maneira que quanto menor é o número de sítios mutados, menor é o efeito no comprimento dos ramos. Isso pode ser facilmente entendido se considerarmos que para uma pequena taxa de mutação, poucas mutações serão acrescidas a cada indivíduo da população de forma que a seleção não afetará o comprimento dos ramos, porém quando a taxa de mutação aumenta, o processo de seleção eliminará os indivíduos mais novos que possuem maior número de mutações em seu genoma modificando significativamente os comprimentos dos ramos e medida que o valor seletivo se aproxima da unidade, só os elementos com nenhuma mutação permanecem na população recuperando o caso de seleção neutra.

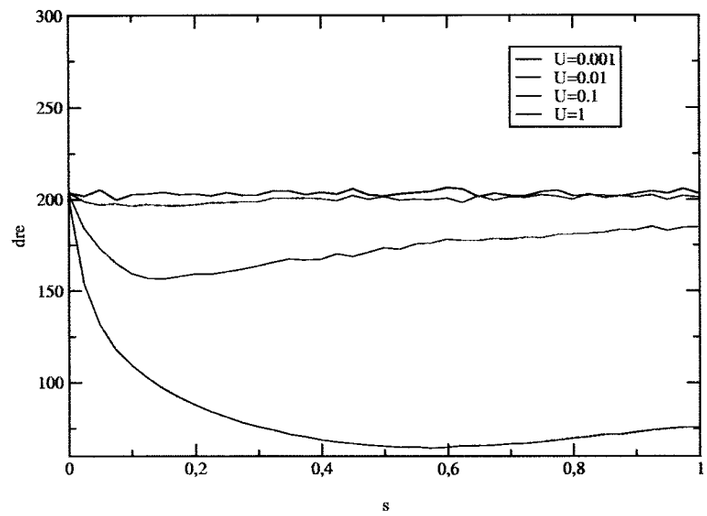

Figura 5.18: Comprimento dos ramos externos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 em uma amostra de tamanho 20.

Depois de observarmos que os comprimentos dos ramos são afetados em um relevo multiplicativo, decidimos tentar quantificar esses efeitos utilizando-se de medidas estatísticas conhecidas [77] e que vem sendo aplicadas a modelos de evolução atualmente, estes resultados serão apresentados na próxima seção.

5.3.1 Medidas da topologia das árvores

Para entender melhor o efeito da seleção na topologia das árvores, aplicamos diferentes testes estatísticos em nossas amostras. Os testes escolhidos [77] servem para definir o grau de simetria de uma árvore binária, figura 5.20, ou seja, como os comprimentos dos ramos são afetados pelo processo seletivo, afetando diretamente a sua topologia.

Como apresentado na seção anterior, as árvores serão formadas por n indivíduos,

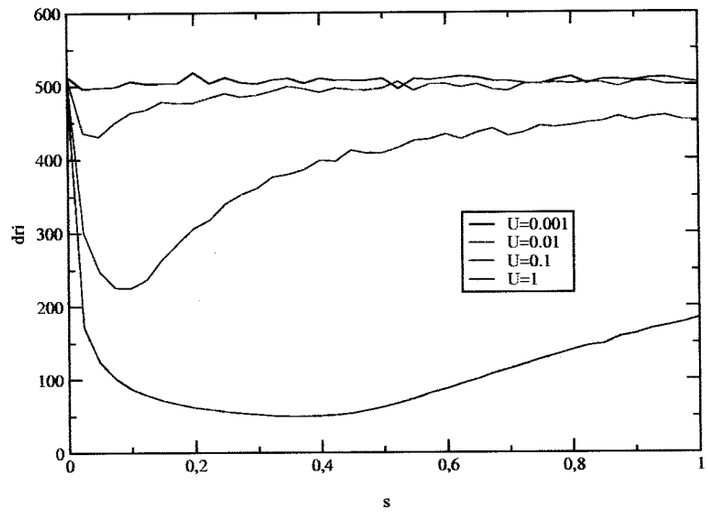

Figura 5.19: Comprimento dos ramos internos para o modelo de infinitos sítios ($L = \infty$) em seqüências binárias com relevo multiplicativo para uma população de tamanho 100 em uma amostra de tamanho 20.

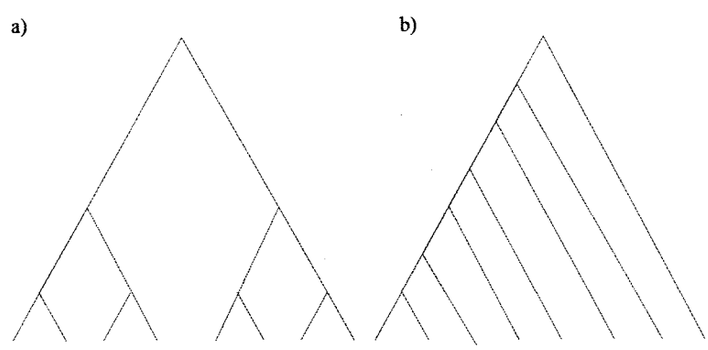

Figura 5.20: *a*) Árvore completamente simétrica: Todos os indivíduos continuam gerando dois novos indivíduos ao longo do tempo. *b*) Árvore completamente assimétrica: Somente um dos dois descendentes de cada geração se reproduz gerando dois novos indivíduos.

escolhidos aleatoriamente na população, na geração atual (nós externos), e portanto terão $n-1$ ancestrais (nós internos), onde Φ é a raiz da árvore e representa o último ancestral comum mais recente (MRCA), da amostra de tamanho n. Uma medida bastante útil é a quantidade de ancestrais (nós internos) entre o i-ésimo indivíduo e o MRCA, inclusive, e será denotada por N_i. Além disso, cada ancestral (nó interno) i possui duas sub-árvores da qual ele é o MRCA. Estas sub-árvores possuem r_i e s_i descendentes na geração atual, onde faremos $r_i \geq s_i$, sem a perda de generalização [77].

Com base nestas medidas podemos definir os testes utilizados para conduzir o estudo sobre a topologia das árvores que são:

$$\overline{N} = \frac{1}{n}\sum_{i=1}^{n} N_i, \tag{5.5}$$

onde $\overline{N}$ é a média sobre todos os nós internos N_i descrita acima.

$$\sigma_N = \left\{\frac{1}{n}\sum_{i=1}^{n}(N_i - \bar{N})^2\right\}^{\frac{1}{2}}, \tag{5.6}$$

sendo σ_N é a variância de N_l.

$$C = \frac{2}{n(n-3)+2}\sum_{i}^{n-1}(r_i - s_i). \tag{5.7}$$

C é conhecida como medida Colless [78] e é baseada nas diferenças entre as sub-árvores r_i e s_i somadas sobre todos os nós internos.

$$B_1 = \sum_{i \neq \Phi} \frac{1}{m_i}, \tag{5.8}$$

B_1 é calculado com base nos nós interiores de uma árvore, excluindo Φ.

$$B_2 = \sum_{l=1}^{n} 2^{-N_l} N_l, \tag{5.9}$$

B_2 é a estatística Shannon-Wiener [79] e é calculada com base na forma da árvore e pode ser utilizado para o estudo de padrões de genealogias.

As figuras 5.21 até 5.25 mostram os valores esperados para as topologias apresentadas nas estatísticas (5.5-5.9), respectivamente, como função do tamanho da amostra n para diversos valores de seleção. Em todas as figuras as barras para os diferentes valores de seleção representam o desvio padrão associado a cada medida realizada.

Nestes gráficos podemos observar que a seleção modifica a topologia das árvores, porém as barras de erro não possuem uma resolução suficientemente pequena para diferenciarmos as curvas através de sua taxa de seleção, ou seja, a seleção é um fator importante para a modificação dos comprimentos dos ramos das árvores, bem como para uma alteração em sua topologia, porém essa modificação é praticamente indistinguível em modelos reais.

Para aumentar a precisão, ou melhor, para diminuir o tamanho das barras de erro, deveríamos dividir o valor de cada barra pela raiz quadrada do número de amostras utilizadas para realizar o experimento. Isso é fácil de ser obtido nas simulações, pois o tempo computacional necessário para produzir gráficos, como os apresentados nas figuras 5.21 até 5.25, com um número de médias grande o suficiente para começarmos a perceber a diferença entre os ramos[§] é de menos de uma semana. Porém existe um problema experimental grande, que é o de conseguir uma população suficientemente grande para que as amostras obtidas não tenham nenhum grau de correlação, ou que este seja o menor possível entre as árvores, pois caso contrário a medida seria afetada por esta correlação.

Portanto podemos dizer que os testes estatísticos aplicados só terão validade no caso em que muitas amostras puderem ser obtidas na população, pois caso contrário o efeito da seleção não poderá ser observado nessas árvores.

[§]Quando executado em um Pentium IV, com precessador de 2 GHz e 1 GB de memória RAM

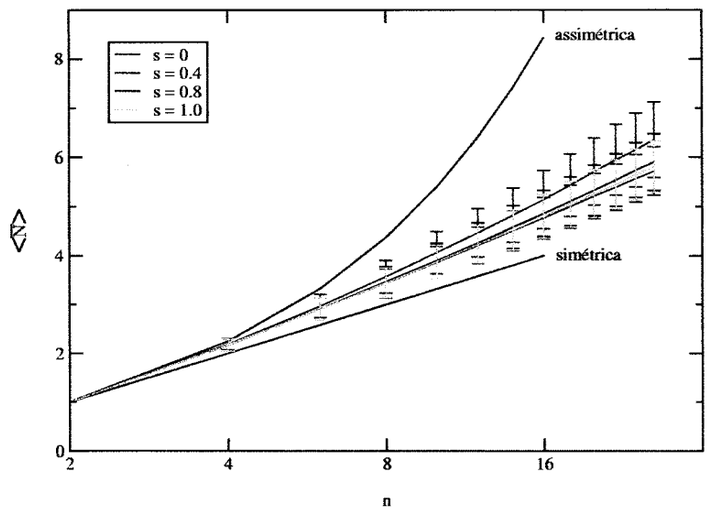

Figura 5.21: Valor esperado de $\overline{N}$ em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva.

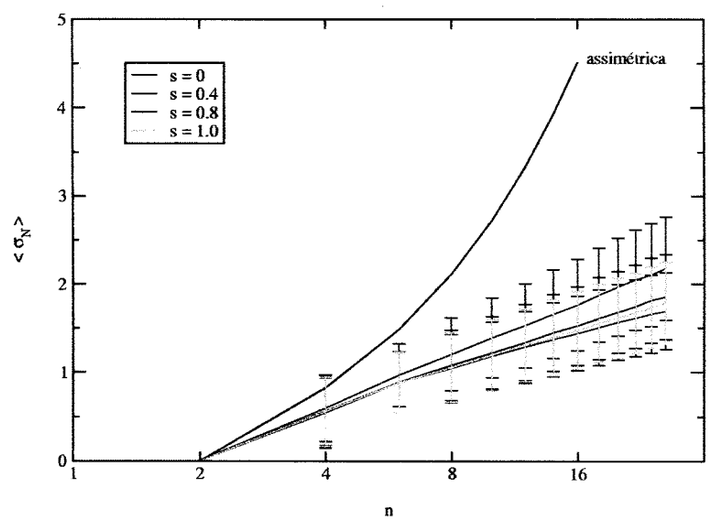

Figura 5.22: Valor esperado de σ_N em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. A árvore mais simétrica possui $\sigma_N = 0$.

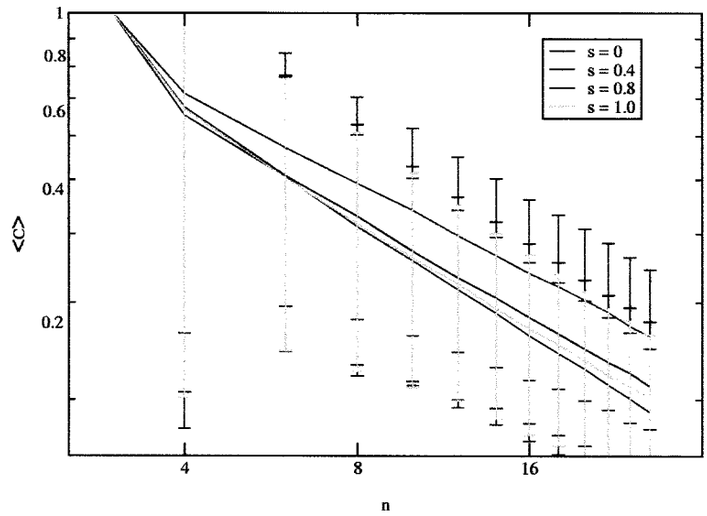

Figura 5.23: Valor esperado para a medida Colless em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva. A árvore mais simétrica possui $C = 0$, enquanto a mais assimétrica possui $C = 1$.

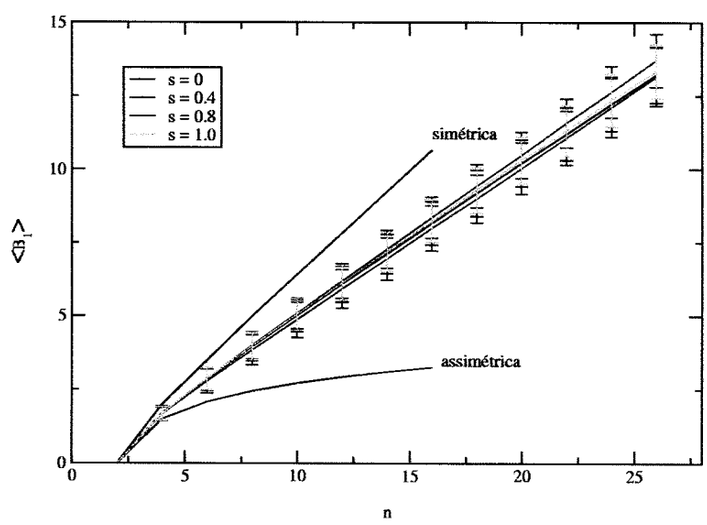

Figura 5.24: Valor esperado de B_1 em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva.

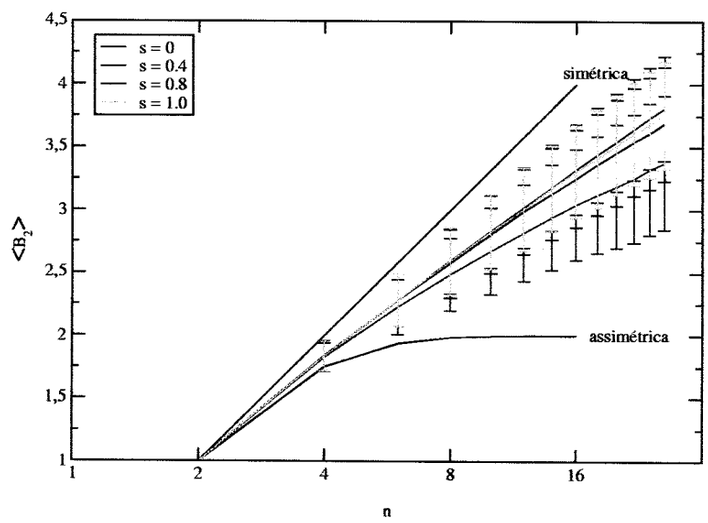

Figura 5.25: Valor esperado de B_2 em função do tamanho da amostra, onde as barras mostram o desvio padrão associado a cada curva.

Capítulo 6

Conclusões

As novas abordagens através de formulações matemáticas deram uma grande contribuição ao estudo do processo evolutivo fundamentado por Darwin em meados do século XIX, principalmente depois dos trabalhos de Haldane, Wright e Fisher ao longo dos anos 20 e 30 e das teorias de evolução, chamadas de não-darwinianas, como a teoria de evolução neutra de Kimura.

A partir da massificação do uso do computador tornou-se possível aliar uma importante ferramenta, como a aplicação de métodos computacionais, para tentar descrever a evolução de uma população ou de um organismo em geral, sendo este o principal ponto abordado nesta tese. Com o estudo numérico experimentos que antes levariam meses ou até mesmo anos puderam ser modelados em detalhes com diversas aplicações de técnicas numéricas.

Com o auxílio da pesquisa genética, novos experimentos com vírus e bactérias estão sendo desenvolvidos, de forma que organismos simples, e que possuem um ciclo de vida relativamente curto, possam ser analisados em tempo real, como é o caso de alguns vírus e bactérias. Unificar estas ferramentas tem sido o objetivo de estudo na últimas décadas, para tentar obter uma completa descrição do problema.

Como apresentado ao longo de toda a tese, o objetivo principal deste trabalho foi de utilizar ferramentas computacionais no estudo de populações assexuadas através do processo conhecido como catraca de Muller e utilizar testes estatísticos na análise topológica de árvores genealógicas.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

O processo evolutivo em populações assexuadas foi estudado em detalhes sobre diferentes dinâmicas, como a de passagens seriais com gargalo e o problema das populações que variam de tamanho com o tempo no capítulo 3 e populações que interagem em uma dinâmica do tipo presa-predador, capítulo 4, onde observamos resultados interessantes nos diversos casos.

Inicialmente estudamos o problema de uma população que pode crescer de tamanho indefinidamente no recipiente em que é feita a sua cultura, sem nenhum limitante de espaço, alimento ou mesmo predação até atingir o ponto de equilíbrio, que é caracterizado pela presença de todas as classes na população, ou seja, existem indivíduos com diferentes quantidades de mutações em seu genoma. Quando a população atinge este estado de equilíbrio, um indivíduo representante de uma classe específica, é escolhido aleatoriamente e é passado para um novo recipiente para começar uma nova cultura. Este tipo de experimento vem sendo desenvolvido em laboratórios para medir exatamente como se dá o processo evolutivo de organismos reais, sendo que este experimento recebe o nome de passagem serial com gargalo.

Este processo pode ser observado na natureza em uma população viral, como o vírus da AIDS, que se desenvolve em um hospedeiro e aumenta rapidamente a sua população nos primeiros momentos depois da infecção. Porém quando existe a transmissão para um novo hospedeiro, apenas alguns poucos indivíduos da população desenvolvida no primeiro hospedeiro passam para o segundo, começando assim uma nova linhagem que pode evoluir de maneira diferente da primeira população.

Em nosso experimento computacional, utilizamos um relevo com epistase* para descrever o tipo de meio-ambiente em que a população evolui e pudemos observar que a catraca de Muller não para completamente para altos valores de epistase, como foi inicialmente sugerido por Kondrashov [21]. Nesta tese foram apresentados resultados que demonstram que a catraca cada continua atuando, porém a taxas cada vez menores à medida que a população evolui, enquanto que a velocidade da catraca permanece constante para baixos

*relevo este que cada nova mutação possui um caráter mais prejudicial (deletério) que as mutações anteriores quando o coeficiente $\alpha > 1$ e menos prejudicial quando $\alpha < 1$

valores, como apresentado na seção 3.1.

Na seção 3.2, estudamos o comportamento das populações assexuadas que podem aumentar e diminuir de tamanho devido a características intrínsecas ao meio em que ela vive. Estudamos diversos casos, como o crescimento e decrescimento exponencial, sendo este estudo feito sobre diversas taxas de crescimento, e ainda estudamos o caso em que a população apresenta um comportamento oscilatório, ou seja, em alguns períodos ela aumenta de valor e em outros ela diminui. Neste trabalho observamos também o efeito da extinção das populações devido ao acúmulo de mutações sofridas ao longo de suas gerações, bem como a parada da catraca devido ao crescimento exponencial da população e ainda a volta de seu movimento quando ocorre a saturação desta devido ao meio-ambiente.

No capítulo 4 estudamos o comportamento de populações assexuadas que interagem em uma dinâmica do tipo presa-predador. A interação entre as populações de presas, predadores e o meio-ambiente foram baseadas nas equações de Beddington [29], por apresentar um comportamento discreto no tempo além da característica de não permitir que as populações de presas aumentem indefinidamente sempre que as populações de predadores são extintas, isso se deve ao fato da introdução do parâmtetro k que não fazia parte das equações originais de Lotka-Volterra.

Para adequar esse tipo de dinâmica ao modelo que trabalhamos modificamos alguns pontos importantes, onde vale ressaltar como principal modificação a discretização no número de indivíduos gerados pelas equações do número de indivíduos $N_H(t)$ e $N_P(t)$ das populações de presas e predadores, respectivamente. Isso foi necessário para podermos identificar quantas mutações eram passadas aos descendentes a cada nova geração, assim deveríamos identificar qual era o "pai" e qual era o "filho". Feito essa modificação trabalhamos com o processo da catraca de Muller como nos casos anteriores.

Observamos que as população de presas se extingue mais rapidamente na presença dos predadores, do que quando aquela evolui sozinha, ou mesmo quando as duas populações interagem, porém sem o efeito da seleção. Na realidade existe um processo que atua negativamente nas duas linhagens que é referente à degeneração das populações devido ao acúmulo de mutações sofridas no decorrer da evolução, onde cada vez mais a seleção

irá diminuir o número de indivíduos sobreviventes.

Um ponto interessante que foi observado é a presença de patamares que representam uma parada na velocidade da catraca sempre que a população está crescendo, enquanto observamos que a mesma velocidade aumenta a medida que a população decresce. Este processo pode ser explicado em duas etapas distintas e com a ajuda dos resultados do capítulo 3. Quando a população está crescendo, a probabilidade desta perder a classe de indivíduos mais adaptada na população é menor, pois podemos considerar que se a população cresce indefinidamente ela vai atingir o caso de equilíbrio descrito na seção 3.1. Porém a medida que a população sofre um limitante externo, que freia o seu crescimento, a catraca volta a atuar e as classes de indivíduos mais adaptadas são perdidas como acontece no modelo de infinitos sítios tradicional. Porém quando a população está diminuindo rapidamente, além da perda dos indivíduos pela predação, existe ainda a perda dos indivíduos pelo processo de seleção o que aumenta a velocidade da catraca. Ao juntarmos estes dois efeitos recuperamos o comportamento dos gráficos descritos no capítulo 4.

Finalmente estudamos no capítulo 5, a possibilidade de inferir o processo de seleção em uma população se estudarmos a forma topológica da árvore genealógica que descreve a sua evolução. Inicialmente estudamos o processo da formação dos ramos internos e externos em uma árvore binária e notamos que o processo seletivo modifica o comprimento dos ramos internos e externos segundo a estatística de Fu e Li. Partindo deste raciocínio, decidimos estudar o comportamento de uma população fazendo a reconstrução de sua árvore genealógica e depois calculando medidas referentes a sua simetria.

Para estes testes utilizamos 5 estatísticas que foram: $\overline{N}$ que representa a média sobre todos os nós internos N_i; σ_N é a variância de N_l; a medida de Colless [78] (C) que baseia-se nas diferenças entre as sub-árvores r_i e s_i; a estatística B_1 que é calculada com base nos nós interiores de uma árvore, excluindo a sua raíz; B_2 que é a estatística Shannon-Wiener [79] e é calculada com base na forma da árvore e pode ser utilizada para o estudo de padrões de genealogias.

Em todos os testes observamos que existe uma modificação nos padrões de simetria da árvores, porém estes desvios são muito pequenos para poderem ser medidos em uma

população real, pois os erros associados a estas medidas são consideravelmente grandes para o número de amostras utilizadas, causando na maioria dos casos uma sobreposição das curvas, anulando completamente a possibilidade de inferir qualquer resultado sobre essas medidas. Estes testes só apresentam uma boa validade se conseguíssemos obter um número muito grande de amostras para que o erro associado a essa medida diminuísse consideravelmente, isto pode ser conseguido em um experimento computacional, onde podemos aumentar o número de amostras até obtermos o resultado que desejamos, porém este mesmo método não pode ser obtido experimentalmente, já que a medida que aumentamos o número de amostras, aumentamos também a correlação entre elas, devido a finitude da população em estudo.

Apêndice A

Alguns Resultados Analíticos para o Modelo de Evolução Neutra

Vamos somente considerar o modelo de evolução neutra na sua forma mais simples, na qual a população é composta por um número fixo $N \gg 1$ de indivíduos haplóides que se reproduzem assexuadamente. A cada indivíduo α associamos um genitor (pai) $G_t(\alpha)$. De acordo com o modelo de evolução neutra, $G_t(\alpha)$ é escolhido aleatoriamente com probabilidade uniforme na população para cada indivíduo α e a cada geração t. A qualificação *neutra* deve-se ao fato de que todos os genótipos têm chances iguais de produzirem descendentes; se houvesse seleção favorecendo alguns genótipos, estes teriam maior probabilidade de contribuírem com descendentes para a geração subseqüente.

Em uma população assexuada, todos os indivíduos estão relacionados entre si, uma vez que todos eles descendem de um único ancestral comum que existiu em algum momento no passado. Seja $T^{\alpha\beta}$ o tempo em gerações transcorrido desde o último ancestral comum dos indivíduos α e β. Uma grandeza de importância fundamental no estudo probabilístico de genealogias é a probabilidade $\bar{P}(T)$ de que dois indivíduos escolhidos aleatoriamente na população tenham $T^{\alpha\beta} = T$. Essa grandeza caracteriza a estrutura da árvore genealógica, e depende das propriedades do processo de ramificação envolvido na dinâmica evolucionária. A notação $\bar{\cdot\cdot}$ indica uma média sobre populações independentes. A probabilidade de que dois indivíduos escolhidos aleatoriamente na população não tenham um

ancestral comum na geração anterior é simplesmente

$$(1 - 1/N)\,, \tag{A.1}$$

de onde podemos concluir que a probabilidade de que eles não tenham um ancestral comum t gerações atrás é

$$(1 - 1/N)^t \approx \mathrm{e}^{-t/N}. \tag{A.2}$$

A escala relevante de tempo é então proporcional ao tamanho da população N. Assim, a probabilidade de que dois indivíduos tenham um ancestral comum entre 0 e T gerações atrás é

$$1 - \mathrm{e}^{-T/N}. \tag{A.3}$$

Uma vez que $\bar{P}(T)$ é a probabilidade de que o ancestral comum ocorra exatamente a T gerações atrás, podemos escrever

$$\bar{P}(T) = \left[1 - \mathrm{e}^{-(T+1)/N}\right] - \left[1 - \mathrm{e}^{-T/N}\right] \approx \frac{\mathrm{e}^{-T/N}}{N}. \tag{A.4}$$

Vejamos agora como introduzir mutações neste modelo básico. Naturalmente, mutações neutras, ou seja, mutações que não afetam a adaptabilidade ("fitness"), não modificam a genealogia dos organismos. Vamos supor (a) que cada dígito (alelo) tenha a probabilidade Udt de sofrer uma mutação durante o intervalo de tempo dt e, (b) que as mutação sempre produzam um novo dígito (alelo), nunca antes presente na população. Um modelo 'microscópico' que reproduz essa situação é o modelo de infinitos sítios descrito a seguir. Considere uma população de genes de tamanho $L \to \infty$ cujos sítios podem tomar os valores $\mathcal{A}$ ou $\mathcal{B}$. Se o valor de um dado sítio for $\mathcal{A}$, então esse sítio ainda não sofreu mutação, ou seja, o nucleotídeo (A, C, G ou T) presente naquele sítio é idêntico ao do sítio correspondente do ancestral que originou a população. Por outro lado, se o valor do sítio for $\mathcal{B}$ então o seu nucleotídeo é diferente daquele do sítio homólogo do ancestral. Daí é imediato verificar que a probabilidade $\phi_n(t)$ de que um indivíduo difira de n mutações do ancestral obedece a equação de recorrência

$$\phi_n(t + dt) = Udt\phi_{n-1}(t) + (1 - Udt)\phi_n(t) \tag{A.5}$$

que no limite de dt infinitesimal reduz-se a seguinte equação diferencial de diferenças

$$\frac{d\phi_n(t)}{dt} = U\left[\phi_{n-1}(t) - \phi_n(t)\right] \tag{A.6}$$

cuja solução é a distribuição de Poisson

$$\phi_n(t) = \mathrm{e}^{-Ut}\frac{(Ut)^n}{n!}. \tag{A.7}$$

Em particular, a distância média (em termos do número de mutações) entre o indivíduo e seu ancestral cresce linearmente com o tempo, $\bar{n} = Ut$, figura 5.5. Podemos agora calcular a probabilidade P_n de que a distância entre dois indivíduos quaisquer seja igual a n. Lembrando que os dois indivíduos devem ter necessariamente se originado de um mesmo ancestral em algum ponto do passado podemos escrever

$$P_n = \int_0^\infty dT\bar{P}(T)\phi_n(2T) \tag{A.8}$$

onde $\bar{P}(T)$ dá a probabilidade de que o último ancestral comum dos dois indivíduos tenha existido a T gerações atrás. Utilizando a equação (A.4) podemos efetuar a integral e obter

$$P_n = \frac{\lambda}{(\lambda+1)^{n+1}} \tag{A.9}$$

onde $\lambda = 1/2UN$. Daí, a distância média $d_{\alpha\beta}$ entre dois indivíduos arbitrários α e β é simplesmente $d_{\alpha\beta} = 1/\lambda = 2UN$, figuras 5.8 e 5.9.

Apêndice B

Determinação dos Comprimentos de Ramos Externos.

Com base na figura 5.4, vamos considerar t_m como sendo o tempo necessário para que m seqüências coalesçam em $m-1$. Tajima [80] demonstrou que a distribuição destes tempos em uma árvore binária segue uma distribuição exponencial da forma

$$g(t_m) = \frac{m(m-1)}{2N} exp\left[-\frac{m(m-1)}{2N} t_m\right], \tag{B.1}$$

onde o primeiro momento é

$$\overline{t_m} = \frac{2N}{m(m-1)}, \tag{B.2}$$

Para o caso em que a genealogia a ser investigada tenha n seqüências, e considerando uma árvore binária, o número total de ramos (J_n) presentes será de $2(n-1)$, sendo que o número total de ramos externos (E_n) é n e $n-2$ o de ramos internos(I_n), desta forma podemos escrever

$$J_n = I_n + E_n. \tag{B.3}$$

sendo

$$E_n = e_1 + e_2 + \ldots + e_n, \tag{B.4}$$

onde e_j é o comprimento do j-ésimo ramo externo. Como os ramos são numerados ao acaso, ou seja, nenhuma informação referente ao seu comprimento é levada em consideração,

podemos separar o seu valor médio na forma

$$\overline{E_n} = \overline{e_1} + \overline{e_2} + \ldots + \overline{e_n} = n\overline{e_n}, \tag{B.5}$$

A genealogia de uma amostra de n seqüências será gerada adicionando-se dois ramos externos de comprimento t_n a um dos $n-1$ ramos existentes na população. A partir deste momento nossa genealogia terá 2 ramos externos com tamanho t_n e $n-2$ ramos externos com tamanho $e_{n-1} + t_n$, assim podemos dizer que a probabilidade $P(e_n)$ de escolhermos um ramo de de tamanho e_n na população será

$$e_n = \begin{cases} e_{n-1} + t_n & ; \quad P(e_n) = \frac{n-2}{n} \\ t_n & ; \quad P(e_n) = \frac{2}{n} \end{cases}$$

Portanto podemos escrever

$$\begin{aligned} \overline{e_n} &= \left(\frac{n-2}{n}\right)\overline{(e_{n-1} + t_n)} + \frac{2}{n}\overline{t_n}. && \text{(B.6)} \\ &= \left(\frac{n-2}{n}\right)\overline{e_{n-1}} + \overline{t_n} && \text{(B.7)} \end{aligned}$$

Reescrevendo e_n como $g_n = n(n-1)e_n$, temos

$$\begin{aligned} \overline{g_n} &= n(n-1)\overline{e_n} && \text{(B.8)} \\ &= n(n-1)\left[\left(\frac{n-2}{n}\right)\overline{e_n - 1} + \overline{t_n}\right] && \text{(B.9)} \\ &= (n-1)(n-2)\overline{e_{n-1}} + n(n-1)\overline{t_n}. && \text{(B.10)} \end{aligned}$$

Da definição de g_n escrita anteriormente, resulta

$$\overline{g_n} = n(n-1)\overline{t_n} + \overline{g_{n-1}} \tag{B.11}$$

substituindo o valor da eq. B.2 na equação anterior,

$$\begin{aligned} \overline{g_n} &= n(n-1)\left[\frac{2N}{n(n-1)}\right] + \overline{g_{n-1}} && \text{(B.12)} \\ &= 2N + \overline{g_{n-1}} && \text{(B.13)} \\ &= 2N(n-1). && \text{(B.14)} \end{aligned}$$

Assim da eq. B.8 temos que

$$\overline{e_n} = \frac{\overline{g_n}}{n(n-1)} = \frac{2N(n-1)}{n(n-1)} = \frac{2N}{n} \tag{B.15}$$

Portanto das eqs. B.5 e B.15 temos que o comprimento médio dos ramos externos é

$$\overline{E_n} = 2N. \tag{B.16}$$

Apêndice C

Siglas e símbolos

Neste apêndice serão apresentados uma lista dos símbolos e siglas escritos nesta Tese.

- t: tempo
- k_i: Número de mutações do i-ésimo indivíduo
- W_i: Adaptabilidade do i-ésimo indivíduo
- U: Taxa de mutação
- s: Taxa de seleção
- N: Número de seqüências (indivíduos) de uma população
- α: Coeficiente de epistase
- $X_k(t)$: Quantidade de indivíduos em uma população que possuem k mutações no tempo t
- $C_k(t)$: Freqüência de indivíduos em uma população que possuem k mutações no tempo t
- R: Quantidade de descendentes produzidos por indivíduo
- r: Quantidade de descendentes produzidos por indivíduo, cujo valor é dado por uma Poisson de média R

- H: Identifica a população de presas, por exemplo, o número de presas em uma população é dado por N_H
- P: Identifica a população de predadores, por exemplo, a taxa de mutação dos predadores é dada por U^P
- n: Tamanho de uma amostra colhida em um população
- E_n: Comprimento dos ramos externos de uma amostra de tamanho n
- e_n: Identificam cada um dos ramos externos de uma árvore genealógica
- I_n: Comprimento dos ramos internos de uma amostra de tamanho n
- i_n: Identificam cada um dos ramos internos de uma árvore genealógica
- L: Número de sítios (dígitos) que forma o genoma de um indivíduo
- $T^{\alpha\beta}$: Matriz para o tempo de coalescência entre dois indivíduos α e β
- MRCA: Do inglês *most recent common ancestor* e refere-se ao indivíduo mais "antigo" da árvore genealógica em estudo
- Φ: Raiz ou MRCA de uma árvore
- $\overline{N}$: É a média topológica de uma árvore referente aos nós internos de uma árvore
- σ_N: É a média topológica de uma árvore referente a variância de $\overline{N}$
- C: Medida Colless [78] e é baseada nas diferenças entre as sub-árvores r_i e s_i somadas sobre todos os nós internos
- B_1: É a média topológica de uma árvore calculada com base nos nós interiores de uma árvore, excluindo Φ
- B_2: É a estatística Shannon-Wiener [79] e é calculada com base na forma da árvore]

Bibliografia

[1] L. E. Orgel, *As Origens da vida: Moléculas e Seleção Natural.* Universidade de Brasília, 1985.

[2] C. Darwin, *The origin of species.* John Murray, sixth ed., 1873.

[3] editor:R. C. Stauffer, *Charles Darwin's Natural Selection.* Cambridge, 1975.

[4] S. J. Gould, *The structure of evolutionary theory.* Harvard University Press, 2002.

[5] M. Kimura, *On the probability of fixation of mutant genes in a population*, *Genetics* **47** (1962) 713–719.

[6] M. Kimura and T. Maruyama, *The mutational load with epistatic gene interactions in fitness*, *Genetics* **54** (1966) 1337–1351.

[7] M. Kimura, *Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles*, *Genetical Research* **11** (1968) 247.

[8] M. Kimura, *Evolutionary rate at molecular level*, *Nature* **217** (1968) 624.

[9] J. L. King and T. H. Jukes, *Non-darwiniam evolution*, *Science* **164** (1969) 788.

[10] H. J. Muller, *The relation of recombination to mutational advance*, *Mutation Research* **1** (1964) 2–9.

[11] M. Eigen and R. Winkler-Oswatitsch, *Transfer-rna: The early adaptor*, *Die Naturwissenschaften* **68** (1981) 217–228.

[12] L. E. Orgel, *The origin of life on the earth, Scientific American* **October** (1994) 77–83.

[13] G. A. C. Bell, *The Masterpiece of Evolution: the Evolution and Genetics of Sexuality.* University of California Press, 1982.

[14] R. Vrijenhoek, R. Dawley, C. Cole, and J. Bogart in *Evolution and Cytology of Unisexual Vertebrates* (R. Dawley and J. Bogart, eds.), pp. 19–23. University of State New York, 1990.

[15] S. P. Otto and T. Lenormand, *Resolving the paradox of sex and recombination, Nature Review Genetics* **3** (2002) 252–261.

[16] S. P. Otto, *The advantages of segregation and the evolution of sex, Genetics* **164** (2003) 1099–1118.

[17] E. S. Dolgin and S. P. Otto, *Segregation and the evolution of sex under overdominant selection, Genetics* **164** (2003) 1119–1128.

[18] S. P. Otto and S. L. Nuismer, *Species interactions and the evolution of sex, Science* **304** (2004) 1018–1020.

[19] G. C. Williams, *Sex and Evolution.* Princeton University Press, 1975.

[20] A. S. Kondrashov, *Deleterious mutations and the evolution of sexual reproduction, Nature* **336** (1988) 435–440.

[21] A. S. Kondrashov, *Muller's ratchet under epistatic selection, Genetics* **136** (1994) 1469–1473.

[22] A. S. Kondrashov, *Measuring spontaneous deleterious mutation process, Genetica* **102/103** (1998) 183–197.

[23] B. Charlesworth and D. Charlesworth, *Some evolutionary consequences of deleterious mutations, Genetica* **102/103** (1998) 3–19.

[24] P. J. Gerrish and R. E. Lenski, *The fate of competiting beneficial mutations in an asexual population*, *Genetica* **102/103** (1998) 127–144.

[25] J. M. Smith, *The Evolution of Sex.* Cambridge University Press, 1978.

[26] S. F. Elena and R. E. Lenski, *Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation*, *Nature Review Genetics* **4** (2003) 457–469.

[27] A. Colato and J. F. Fontanari, *Soluble model for the accumulation of mutations in asexual population*, *Physical Review Letters* **87** (2001) 238102.

[28] J. F. Fontanari, A. Colato, and R. S. Howard, *Mutation accumulation in growing asexual lineages*, *Physical Review Letters* **91(21)** (2003) 218101.

[29] J. R. Beddington, C. A. Free, and J. H. Lawton, *Dynamic complexity in prey-predator models framed in difference equations*, *Nature* **255** (1975) 58–60.

[30] L. P. Maia, *Dinâmica adaptativa, genealogias e testes estatísticos de neutralidade em evolução molecular*, *Universidade de São Paulo* (2004).

[31] L. P. Maia, A. Colato, and J. F. Fontanari, *Effect of selection on the topology of genealogical trees*, *Journal of Theoretical Biology* **226** (2004) 315–320.

[32] H. Kitaro, *Computational systems biology.*, *Nature* **420** (2002) 206–210.

[33] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, *The segment polarity network is a robust developmental module.*, *Nature* **406** (2000) 188–192.

[34] M. L. Shuler, S. Leung, and C. C. Dick, *A mathematical model for the growth of a single bacterial cell.*, *Ann. NY Acad. Sci.* **326** (1979) 35–55.

[35] M. Eigen, C. K. Biebricher, M. Gebinoga, and W. C. Gardiner, *The hypercycle. coupling of rna and protein biosynthesis in the infection cycle of an rna bacteriophage.*, *Biochemistry* **30** (1991) 11005–11018.

[36] D. Endy, L. You, J. Yin, and I. J. Molineux, *Computation, prediction, and experimental tests of fitness for bacteriophage t7 mutants with permuted genomes.*, *Proc. Natl. Acad. Sci. USA* **97** (2000) 5375–5380.

[37] B. Reddy and J. Yin, *Quantitative intracellular kinetics of hiv type 1.*, *AIDS Res. Hum. Retrovir.* **15** (1999) 273–283.

[38] L. You, P. F. Suthers, and J. Yin, *Effects of escherichia coli physiology on growth of phage t7 in vivo and in silico.*, *J. Bacteriol.* **184** (2002) 1888–1894.

[39] A. Colato and S. S. Mizrahi, *Effects of random migration in population dynamics*, *Physical Review E* **64** (2001) 011901.

[40] H. Sayama, M. A. M. de Aguiar, Y. Bar-Yam, and M. Baranger, *Spontaneous pattern formation and genetic invasion in locally mating and competing populations*, *Physical Review E* **65 (5)** (2002) 051919.

[41] D. Hadjiavgousti and S. Ichtiaroglou, *Existence of stable localized structures in population dynamics through the allee effect*, *Chaos Solitons & Fractals* **21 (1)** (2004) 199–131.

[42] M. Droz and A. Pekalski, *Role of evolution by natural selection in population dynamics*, *Physical Review E* **69 (5)** (2004) 051912.

[43] E. Baake and W. Gabriel, *Biological evolution through mutation, selection, and drift: An introductory review*, *Los alamos arXiv* **9907372** (1999) 1–62.

[44] B. Drossel, *Biological evolution and statistical physics*, *Los alamos arXiv* **0101409** (2001) 1–106.

[45] J. F. Crow and M. Kimura, *An introduction to population genetics theory.* Harper & Row, New York, 1970.

[46] M. T. Sonoda, *Dinâmica e genealogia de modelos de evolução*, *Universidade de São Paulo* (2001).

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

[47] D. F. Botelho, *Dinâmica de modelos de genética de populações com recombinação, Universidade de São Paulo* (2003).

[48] P. R. de Araújo Campos, *Modelos de informação em evolução pré-biótica, Universidade de São Paulo* (2001).

[49] S. Bonhoeffer and P. Sniegowski, *The importance of being erroneousl.*, *Nature* **420** (2002) 367–368.

[50] D. L. Hartl and A. G. Clark, *Priciples of Population Genetics.* Sinauer Associates, second ed., 1989.

[51] D. Greig and M. Travisano, *Haploid superiority*, *Science* **299** (2003) 524–525.

[52] S. L. Nuismer and S. P. Otto, *Host-parasite interactions and the evolution of ploidy, Proccedings of the National Academy of Sciences of the USA* **101** (2004) 11036–11039.

[53] G. M. Cooper, *The Cell: A molecular approach.* ASM Press, second ed., 2000.

[54] M. Lynch, R. Bürger, D. Butcher, and W. Gabriel, *The mutational meltdown in asexual populations*, *Journal of Heredity* **84** (1993) 339–344.

[55] J. Haigh, *The accumulation of deleterious genes in a population - muller's ratchet, Journal of Theoretical Biology* **14** (1978) 251–267.

[56] P. G. Higgs and G. Woodcock, *The accumulation of mutations in populations and the structure of the genealogical trees in the presence of selection*, *Journal of Mathematical Biology* **33** (1995) 677–702.

[57] A. W. Bowman and D. R. Robinson, *Introduction to statistics: A computer illustrated text.* IOP, revised ed., 1996.

[58] L. Chao, *Evolution of sex in rna viruses*, *Journal of Theoretical Biology* **133** (1988) 99 – 112.

[59] J. D. Murray, *Mathematical Biology.* Springer, second ed., 1993.

[60] T. Cheon, *Evolutionary stability of ecological hierarchy, Physical Review Letters* **90** (2003) 258105.

[61] R. S. Howard and C. M. Lively, *Parasitism, mutation accumulation and the maintenance of sex, Nature* **367** (1994) 554–557.

[62] P. R. de Aráujo Campos, V. M. de Oliveira, R. Giro, and D. S. Galvão, *Emergence of prime number as the result of evolutionary strategy, Physical Review Letters* **93** (2004) 09107.

[63] H. Allen and S. P. Otto, *Does diploidy increase the rate of adaptation?, Genetics* **136** (1994) 1475–1480.

[64] S. M. Henson, R. F. Costantino, J. M. Cushing, R. A. Desharnais, B. Dennis, and A. A. King, *Lattice effects observed in chaotic dynamics of experimental populations, Science* **204** (2001) 602–605.

[65] G. Domokos and I. Scheuring, *Random perturbations and lattice effects in chaotic population dynamics, Science* **297** (2002) 2163a.

[66] M. C. Rivera and J. A. Lake, *The ring of life provides a evidence of genome fusion origin of eukaryotes, Nature* **431** (2004) 152–155.

[67] W. Martin and T. M. Embley, *Early evolution comes full circle, Nature* **431** (2004) 134–137.

[68] M. Eigen and R. Winkler-Oswatitsch, *Statistical geometry on sequence space, Methods in Enzimology* **183** (1987) 505–531.

[69] T. Gojobori, E. N. Moriyama, and M. Kimura, *Statistical methods for estimating sequence divergence, Methods in Enzimology* **183** (1987) 531–551.

[70] M. Eigen, R. Winkler-Oswatitsch, and A. Dress, *Statistical geometry in sequence space: A method of quantitative comparative sequence analysis*, *Proc. Natl. Acad. Sci.* **85** (1988) 5913–5917.

[71] Y.-X. Fu and W.-H. Li, *Statistical tests of neutrality of mutations*, *Genetics* **133** (1992) 693–709.

[72] G. B. Golding, *The effect of purifying selection on genealogy*, in *Progress in population genetics and human Evolution* (P. Donnelly and S. Tavaré, eds.), pp. 271–285. Springer - New York, 1997.

[73] A. O. Mooers and S. B. Heard, *Inferring evolution process form phylogenetic tree shape*, *The Quarterly Review of Biology* **72** (1997) 31–54.

[74] J. D. Wall and R. R. Hudson, *Coalescent simulations and statistical tests of neutrality*, *Molecular Biology and Evolution* **18** (2001) 1134–1135.

[75] F. Depaulis and M. Veuille, *Neutrality tests based on the distribution of haplotypes under an infinite-site model*, *Molecular Biology and Evolution* **15** (1998) 1788–1790.

[76] F. Depaulis, S. Mousset, and M. Veuille, *Haplotypes tests using coalescent simulations on the number of segregating sites*, *Molecular Biology and Evolution* **18** (2001) 1136–1138.

[77] M. Kirkpatrick and M. Slatkin, *Searching for evolutionary patterns in the shape of phylogenetic tree*, *Evolution* **47** (1993), no. 4 1171–1181.

[78] D. H. Colless, *Review of phylogenteics: The theory and practice of phylogenetic systematics*, *Systematic Zoology* **31** (1982) 100–104.

[79] K. Shao and R. R. Sokal, *Tree balance*, *Systematic Zoology* **39** (1990), no. 3 266–276.

[80] F. Tajima, *Statistical method for testing the neutral mutation hypothesis by dna polymorphism*, *Genetics* **123** (1989) 585–595.