

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA

DESEMPENHO DO ALGORITMO GENÉTICO COM
ITERAÇÃO RETROVIRAL PARA OTIMIZAÇÃO DE
FUNÇÕES COM REPRESENTAÇÃO REAL

DIELLE DA SILVA CORREA FRANCO

DM: 21/2015

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA

DIELLE DA SILVA CORREA FRANCO

DESEMPENHO DO ALGORITMO GENÉTICO COM
ITERAÇÃO RETROVIRAL PARA OTIMIZAÇÃO DE
FUNÇÕES COM REPRESENTAÇÃO REAL

DM: 21/2015

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015

**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA**

DIELLE DA SILVA CORREA FRANCO

**DESEMPENHO DO ALGORITMO GENÉTICO COM
ITERAÇÃO RETROVIRAL PARA OTIMIZAÇÃO DE
FUNÇÕES COM REPRESENTAÇÃO REAL**

Dissertação submetida à banca examinadora do Programa de Pós-Graduação em Engenharia Elétrica da UFPA para obtenção do Grau de Mestre em Engenharia Elétrica na área de Computação Aplicada.

**UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015**

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Franco, Dielle da Silva Corrêa, 1984-
Desempenho do algoritmo genético com iteração
retroviral para otimização de funções com representação
real / Dielle da Silva Corrêa Franco. - 2015.

Orientador: Roberto Célio Limão de
Oliveira;

Coorientador: Ádamo Lima de Santana.
Dissertação (Mestrado) - Universidade
Federal do Pará, Instituto de Tecnologia,
Programa de Pós-Graduação em Engenharia
Elétrica, Belém, 2015.

1. Algoritmos genéticos. 2. Infecção -
modelos matemáticos. 3. Otimização combinatória.
4. Computação evolucionária. I. Título.

CDD 22. ed. 006.31

**“DESEMPENHO DO ALGORITMO GENETICO COM ITERACAO RETROVIRAL
PARA OTIMIZACAO DE FUNCOES COM REPRESENTACAO REAL”**

AUTOR: DIELE DA SILVA CORREA FRANCO

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRA EM ENGENHARIA
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 30/06/2015

BANCA EXAMINADORA:



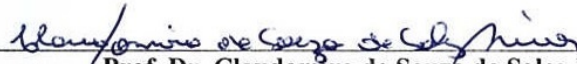
Prof. Dr. Roberto Célio Limão de Oliveira

(Orientador – PPGEE/UFPA)



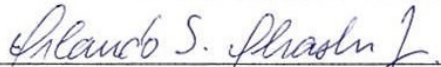
Prof. Dr. Adamo Lima de Santana

(Co-Orientador – PPGEE/UFPA)



Prof. Dr. Claudomiro de Souza de Sales Junior

(Avaliador Externo ao Programa – ICEN/UFPA)



Prof. Dr. Orlando Shiguelo Ohashi Junior

(Avaliador Externo – UFRA)

VISTO:

Prof. Dr. Evaldo Gonçalves Pelaes

(Coordenador do PPGEE/ITEC/UFPA)

Um discípulo de quem jamais se pede nada que não possa fazer, nunca faz tudo o que pode.

John Stuart Mill

Agradecimentos

Dedico meus sinceros agradecimentos para:

– o estimado professor Doutor Roberto Célio Limão, pela orientação, incentivo e muita paciência, pois quando eu nem mais acreditava, ele como todo bom mestre continuou me ajudando, vendo o que nem eu mais via: meu potencial, por isso continuou a prover os meios para que este trabalho fosse concluído.

– o prezado professor Doutor Ádamo Santana que também me deu muitas oportunidades para terminar meu trabalho e representou um valioso auxílio em minha orientação.

– os meus amigos de trabalho que me apoiaram e ajudaram várias vezes durante o desenvolvimento de meu trabalho. Em especial ao meu amigo e co-orientando Sebastião Fonseca, que me ajudou grandemente com incentivo, sugestões e ajuda prática em vários problemas de programação que surgiram.

– minha família. Meus pais Manoel e Sonia, meus irmãos Diego, Danielle e Thiago, minha segunda família Maria e Dulce. Todos foram de fundamental importância por me motivarem neste trabalho e fornecerem grande ajuda emocional em momentos pessoais extremamente difíceis que ocorreram durante a produção deste trabalho.

– acima de tudo, agradeço a Jeová Deus por permitir que eu realizasse e terminasse este trabalho mantendo meus valores e princípios.

Resumo

Desempenho do Algoritmo Genético com Iteração Retroviral para Otimização de Funções com Representação Real

Palavras-chave: Evolução Molecular; Infecção por Retrovírus; Algoritmos Genéticos; Otimização de Funções

A ideia da infecção viral vem sendo utilizada nos Algoritmos Genéticos (AG) para melhorar a taxa de busca dos algoritmos, superando a convergência prematura por meio do controle da diversidade da população, visto que na natureza a recombinação genética por vírus apresenta alta velocidade de replicação e frequente mutação. [Romano 2009].

A meta-heurística denominada AGRI adotou como fonte de inspiração biológica a família *retroviridae*, cujos vírus são baseados em RNA e atendem à necessidade de maior variação alélica do AG, visto que o RNA não possui os mecanismos de correção em seu genoma para eliminar o material viral recombinado.

Neste algoritmo, os vírus são tratados como uma população separada da população de indivíduos. A cada infecção, o material genético viral é transmitido verticalmente entre os hospedeiros difundindo trechos de soluções dos vírus com melhor desempenho por toda população.

A diversidade viral é mantida por meio de um mecanismo que substitui todos o vírus fora da taxa de elitismo viral. Nessa técnica, a população viral evolui junto com a população cromossômica, pois os vírus ineficientes são criados a partir do material genético dos indivíduos mais adaptados e de outros genes novos.

O AGRI segue os princípios biológicos em vários aspectos da infecção e multiplicação viral. Por exemplo: cria a primeira população viral sem o material genético da população somática; escolhe aleatoriamente os vírus que irão infectar um indivíduo, possibilitando que parte da população nunca seja infectada por alguns vírus e que certos vírus infectem mais indivíduos. Além disso, a partir da segunda geração do AG, os

vírus substituídos são criados com material genético de dois indivíduos, e tem diferentes quantidades de genes.

Nesta abordagem, a maximização do espaço de busca é realizado utilizando três mecanismos: alta variabilidade genética da população viral com tamanhos diferentes dos trechos de soluções; efetivação da infecção apenas quando há um aumento no *fitness* do indivíduo; e possibilidade de um indivíduo ser infectado por quaisquer dos vírus da população viral.

Para analisar o efeito dos parâmetros da infecção viral do AGRI e seu desempenho em comparação com outras meta-heurísticas bem conceituadas, as seguintes funções de benchmarking relacionadas com problemas de minimização foram selecionadas: F1 (Shifted Sphere Function), F2 (Shifted Schwefel's Problem), F3 (Shifted Rotated High Conditioned Elliptic Function) e F5 (Schwefel's Problem 2.6 with Global Optimum on Bounds).

Os resultados mostraram que para funções unimodais propostas, o AGRI tem boa performance em comparação com as outras meta-heurísticas selecionadas podendo alcançar o ótimo global ou boas soluções com poucas iterações.

Abstract

Performance Evaluation of Genetic Algorithm with Retroviral Iteration for Real-Valued Functions Optimization

Key words: Genetic Algorithm; Retroviruses Infectio; Real Function; Maximization Problems.

Viral Infection is used to improve the performance in Genetic Algorithms (GA) by reducing premature convergence through the population diversity control, since viruses presents high replication and mutation rates in the nature.

The metaheuristic called AGRI is inspired biologicaly in a viruses family based on RNA, which provide a high allelic variation to GA, since RNA doesn't have genoma correction mechanisms to remove re-combined viral genetic material .

In this algorithm, the viruses are a separate population. To each infection, the better performance viruses genomes are transmitted vertically spreading parts of solutions to GA population.

The diversity viral is maintained through a mechanism that substitutes all viruses out of elitism viral rate. In this method, the virus population evolves along with GA population, so the inefficient viruses are created from genetic material of the better adapted individuals and other new genes.

The algorithm AGRI follows biological principles in several viral infection and multiplication aspects. For example: it creates the first viral population without GA population genetic material; it sorts the viral population before infect an individual, making possible some viruses doesn't infected a part of the population and other viruses infect more individuals. Since GA second-generation, the replaced viruses are created by both individuals genetic material and have different genes quantities.

In this approach, the search space maximization is increased by three mechanisms:

high viral population genetic variability by variety of sizes to solutions pieces; infection validation process that confirms the fitness increases in each individual and infection possibility by any viruses in the viral population.

To analyse the AGRI's viral infection parameters effects and compare his performance with others high-performing metaheuristics, the following minimization benchmarking are selected: F1 (Shifted Sphere Function), F2 (Shifted Schwefel's Problem), F3 (Shifted Rotated High Conditioned Elliptic Function) e F5 (Schwefel's Problem 2.6 with Global Optimum on Bounds).

The results to the functions unimodais proposed showed that AGRI has a good performance in comparison with others metaheuristics reaching in few iterations the global best or good results.

Sumário

1	Introdução	p. 2
1.1	Visão geral	p. 2
1.2	Revisão do Estado da Arte.....	p. 3
1.3	Motivação e desafios	p. 5
1.4	Objetivos.....	p. 7
1.5	Organização do texto	p. 8
2	Referencial Teórico	p. 9
2.1	Sistemas Naturais	p. 9
2.1.1	Biologia Molecular	p. 10
2.1.1.1	Ácido Desoxirribonucléico - DNA	p. 12
2.1.1.2	Ácido Ribonucléico - RNA	p. 14
2.1.2	Vírus	p. 15
2.1.2.1	Ciclo Viral	p. 16
2.1.3	Retrovírus	p. 17
2.2	Considerações Finais	p. 19
3	Computação Evolucionária	p. 21
3.1	Introdução a Algoritmos Genéticos.....	p. 21
3.2	Componentes do AG	p. 24

3.2.1	Indivíduos	p. 24
3.2.2	População Inicial	p. 25
3.2.3	Avaliação	p. 26
3.2.4	Seleção	p. 26
3.2.5	Reprodução	p. 27
3.2.6	Mutação	p. 29
3.3	Comparação com Outras Meta-heurísticas	p. 29
3.4	Considerações Finais	p. 30
4	Algoritmo Genético Retroviral Iterativo	p. 31
4.0.1	O Mecanismo Viral do AGRI	p. 32
4.0.2	Os Parâmetros do AGRI	p. 36
4.1	Considerações Finais	p. 38
5	Resultados	p. 39
5.1	Estudo de Caso	p. 39
5.1.1	Estudo dos Parâmetros da Infecção Viral	p. 42
5.1.1.1	<i>Quantidade de Vírus e Infecção Populacional</i>	p. 43
5.1.1.2	<i>Infecção dos Mais Fracos</i>	p. 46
5.1.1.3	<i>Infecção Única</i>	p. 49
5.1.1.4	<i>Elitismo Viral</i>	p. 52
5.1.1.5	<i>Infecção Interna</i>	p. 54
5.1.2	Desempenho do AGRI - Referências Iniciais	p. 56
5.1.2.1	O AGRI e outras Técnicas de Otimizações Globais	p. 62
5.2	Conclusões do Capítulo	p. 67
6	Conclusões	p. 68
	Referências	p. 70

Lista de Abreviaturas

VEGA	Vírus-Evolutionary Genetic Algorithm
AGIV	Algoritmo Genético com Infecção Viral
AGIVB	Algoritmo Genético com Infecção Viral Básico
AGIVBL	Algoritmo Genético com Infecção Viral e Busca Local
AG	Algoritmos Genéticos
AGRI	Algoritmo Genético Retroviral Iterativo
DNA	Ácido Desoxirribonucléico
RNA	Ácido Ribonucléico
TE	Transposable Elements
ERVs	Endogenous Retroviruses (Retrovírus Endógenos)
LTR	Long Terminal Repeat (Sequência Repetitiva Terminadora)
DE	Differential Evolution
jDE	Adaptive Conventional Differential Evolution Algorithm
PA-jDE	Population Adaptation Differential Evolution Algorithm
PSO	Particle Swarm Optimization Algorithm
BCPSO	Blended Crossover Based Particle Swarm Optimization Algorithm
DVPSO	Distribution Vector Particle Swarm Optimization Algorithm

Lista de Figuras

Figura 1	Substâncias específicas passam pela membrana celular	10
Figura 2	Tipos de células no corpo humano	11
Figura 3	Armazenamento do DNA no núcleo da célula	12
Figura 4	Máquina de Enzimas que divide o DNA em dois filamentos	13
Figura 5	Um filamento exposto de DNA transmite informações para o RNA ...	14
Figura 6	Alguns tipos de vírus	16
Figura 7	Replicação viral com a etapa de transcrição reversa.	17
Figura 8	Fluxograma de um AG Clássico.	23
Figura 9	Cruzamento com um ponto de corte	28
Figura 10	Cruzamento multiponto com dois pontos de corte	28
Figura 11	Cruzamento uniforme	28
Figura 12	Representação de alguns vírus possíveis para um cromossomo de 20 posi-	

ções.	32
Figura 13	Processo de criação de um novo vírus, a partir da 2 ^a Geração: (A) Indivíduo gerado aleatoriamente (B) Indivíduo infectado selecionado da população cromossômica (C) Materiais genéticos cromossômicos mesclados (D) Novo vírus com material genético de um hospedeiro bem adaptado 33
Figura 14	Fluxograma do AGRI. 35
Figura 15	Shifted Sphere Function. Gráfico 3-D para 2 dimensões 40
Figura 16	Shifted Schwefel's Problem 1.2. Gráfico 3-D para 2 dimensões. 40
Figura 17	Shifted Rotated High Conditioned Elliptic Function. Gráfico 3-D para 2 dimensões. 41
Figura 18	Schwefel's Problem 2.6 with Global Optimum on Bounds. Gráfico 3-D para 2 dimensões 41
Figura 19	Função F1: Comparação do Fitness Médio ao variar os parâmetros <i>Quantidade de Vírus e População Infectada</i> 44
Figura 20	Função F1: Avaliação dos Parâmetros <i>Quantidade de Vírus e Infecção Populacional</i> - Comparação da Medida Diversidade Genética 46
Figura 21	Função F1: Avaliação do Parâmetro <i>Infecção dos Mais Fracos</i> - Comparação do Fitness Médio 47
Figura 22	Função F1: Avaliação do Parâmetro <i>Infecção dos Mais Fracos</i> - Comparação da Medida de Diversidade Genética 48
Figura 23	Função F1: Avaliação do Parâmetro <i>Infecção Única</i> - Comparação da Evolução do Fitness Médio 50
Figura 24	Função F1: Avaliação do Parâmetro <i>Infecção Única</i> - Comparação da Medida de Diversidade Genética 51
Figura 25	Função F1: Avaliação do Parâmetro <i>Infecção Única</i> - Comparação do	

	Desvio-Padrão	51
Figura 26	Função F1: Avaliação do Parâmetro <i>Elitismo Viral</i> - Comparação da Evolução do Fitness Médio	52
Figura 27	Função F1: Avaliação do Parâmetro <i>Elitismo Viral</i> - Confrontação da Medida de Diversidade Genética	54
Figura 28	Função F1: Avaliação do Parâmetro <i>Infecção Interna</i> - Comparação do Fitness Médio	55
Figura 29	Função F1: Avaliação do Parâmetro <i>Infecção Interna</i> - Comparação da Medida de Diversidade Genética	56
Figura 30	Função F1: População 1 - Comparação da Evolução do Fitness - Infecção viral de qualidade, AG e AG com Taxa de Mutação Elevada	58
Figura 31	Função F1: População 1 - Comparação da Medida de Diversidade - Infecção viral de qualidade, AG e AG com Taxa de Mutação Elevada	59
Figura 32	Função F1: População 1 - Comparação da Evolução do Fitness - Infecção viral mediana, AG e AG com Taxa de Mutação Elevada	60
Figura 33	Função F1: População 1 - Comparação da Medida de Diversidade - Infecção viral mediana, AG e AG com Taxa de Mutação Elevada	60
Figura 34	Função F1: População 1 - Comparação da Evolução do Fitness - Infecção viral de baixa qualidade, AG e AG com Taxa de Mutação Elevada	61
Figura 35	Função F1: População 1 - Comparação da Medida de Diversidade - Infecção viral de baixa qualidade, AG e AG com Taxa de Mutação Elevada	62
Figura 36	Função F1: Evolução do Fitness Médio do AGRI para 30 dimensões ..	64
Figura 37	Função F2: Evolução do Fitness Médio e Medida de Diversidade Genética do AGRI para 30 dimensões	65
Figura 38	Função F3: Evolução do Fitness Médio e Medida de Diversidade Genética	

do AGRI para 30 dimensões 66

Figura 39 Função F_5 : Evolução do Fitness Médio e Medida de Diversidade Genética
do AGRI para 30 dimensões 66

Lista de Tabelas

Tabela 1	Principais diferenças entre o DNA e o RNA	11
Tabela 2	Analogia entre sistemas naturais e artificiais	22
Tabela 3	Exemplos da decodificação de algumas representações de genótipos em fenótipos	25
Tabela 4	Notações definidas para os testes das funções do <i>benchmark</i>	42
Tabela 5	Configurações do <i>benchmark</i> para os testes das funções	42
Tabela 6	Configurações do <i>benchmark</i> para os testes de avaliação dos parâmetros da infecção viral	43
Tabela 7	Configurações para avaliar o comportamento dos Parâmetros da infecção viral <i>Quantidade de Vírus e Infecção Populacional</i>	43
Tabela 8	Comparação do resultado dos Parâmetros da infecção viral <i>Quantidade de Vírus e Infecção Populacional</i>	44
Tabela 9	Configurações para avaliar o comportamento do Parâmetro da infecção viral <i>Infecção dos Mais Fracos</i>	47
Tabela 10	Resultados para o Parâmetro da infecção viral <i>Infecção dos Mais Fracos</i>	48

Tabela 11	Configurações para avaliar o comportamento do Parâmetro da infecção viral <i>Infeção Única</i>	49
Tabela 12	Resultados para o Parâmetro da infecção viral <i>Infeção Única</i>	50
Tabela 13	Configurações para avaliar o comportamento do Parâmetro da infecção viral <i>Elitismo</i>	52
Tabela 14	Resultados para o Parâmetro da infecção viral <i>Elitismo Viral</i>	53
Tabela 15	Configurações para avaliar o comportamento do Parâmetro da infecção viral <i>Infeção Interna</i>	54
Tabela 16	Comparação do resultado para o Parâmetro da infecção viral <i>Infeção Interna</i>	55
Tabela 17	Parâmetros dos Testes Artificiais	57
Tabela 18	Configuração dos parâmetros do AG clássico e AG com Taxa de Mutação Elevada	57
Tabela 19	Ensaio 1 - Melhor configuração dos parâmetros do AGRI	58
Tabela 20	Ensaio 2 - Configuração viral com qualidade mediana	59
Tabela 21	Ensaio 3 - Pior configuração dos parâmetros do AGRI	61
Tabela 22	Configuração dos Parâmetros do AGRI	63
Tabela 23	Comparação do resultado da melhor configuração do AGRI com outros algoritmos	63

CAPÍTULO 1

Introdução

1.1 Visão geral

A computação natural constitui-se de novas abordagens computacionais caracterizadas pela aproximação com os mecanismos naturais, ou seja, utiliza a natureza como fonte de inspiração para o desenvolvimento de técnicas, ferramentas e dispositivos para solucionar problemas. As metáforas criadas sofrem simplificações em relação aos detalhes do sistema natural, e de maneira geral, os modelos computacionais que são desenvolvidos tem características estocásticas, garantem a solução quase-ótima e apresentam uma maneira declarativa de resolver um problema, isto é, descrevem *o que fazer*, diferentemente dos métodos procedurais que focam em *como fazer*.

A computação natural se dedica à diversas linhas de pesquisa para uma melhor compreensão da natureza e pode ser dividida em três grandes áreas [Castro 2012]:

Computação inspirada na natureza : busca na natureza a inspiração para o desenvolvimento de novas técnicas computacionais para resolver problemas complexos. Como exemplo temos as áreas: Computação Evolucionária, Inteligência Coletiva, Redes Neurais Artificiais, Sistemas Imunológicos Artificiais, Sistemas Endócrinos Artificiais.

Simulação e emulação da natureza utilizando a computação : sintetiza formas, padrões e comportamentos similares aos da natureza, além de desenvolver organismos artificiais. As seguintes áreas estão inclusas: Vida Artificial e Autômatos Celulares, Geometrias Computacional da Natureza ou Fractais; e

Computação com materiais naturais : novo paradigma de computação que tem como objetivo principal substituir as arquiteturas computacionais conhecidas atualmente,

utilizando mecanismos naturais para computar, além de usar novos tipos de matéria-prima para substituir a tecnologia de silício empregada atualmente. São exemplos, as áreas: Computação de DNA e Computação Quântica.

Assim, Castro em [Castro 2012] salienta que os padrões complexos encontrados na natureza são estudados, compreendidos e aplicados para:

- desenvolver novas ferramentas de computação (em software e/ou hardware) para solucionar problemas;
- projetar dispositivos computacionais que simulam, emulam, modelam e descrevem sistemas naturais;
- criar modelos computacionais que sintetizam novas formas de vida por imitar padrões, formas, comportamentos e organismos; e
- desenvolver novos paradigmas de computação, isto é, tipos de computadores em substituição aos atuais baseados em silício.

Este trabalho aplica-se à sub-área da Computação Inspirada na Natureza denominada Computação Evolucionária, mais precisamente utiliza Algoritmos Genéticos.

A fonte de inspiração da natureza é o ciclo de vida dos retrovírus como mecanismo que incorpora novo material genético na malha gênica dos hospedeiros para aumentar a variabilidade genética desses. Meta-heurísticas que utilizam abordagens de renovação da diversidade genética da população de indivíduos do Algoritmo Genético utilizando infecção viral ainda não são largamente aplicadas para a resolução de problemas, por isso, há necessidade de mais estudos nesse campo.

Ademais, este estudo analisa e faz contribuições à meta-heurística *Algoritmo Genético com Iteração Retroviral* por meio da introdução de novos mecanismos que emulam a cadeia de infecção e reprodução viral nos organismos.

O histórico científico a seguir apresenta algumas abordagens desenvolvidas que utilizam a infecção viral para recombinação gênica e aumento da diversidade genética no AG.

1.2 Revisão do Estado da Arte

Kubota et al. propôs o algoritmo evolucionário denominado VEGA (Virus-Evolutionary Genetic Algorithm) em que a evolução ocorre por meio da propagação horizontal e vertical da informação genética. Nessa abordagem a população de hospedeiros é infectada por uma população de vírus. Os vírus são formados por sub-cadeias de DNA, isto é, são vetores com trechos selecionados de DNA da população de hospedeiros, os quais são transcritos nos cromossomos infectados. O algoritmo VEGA utiliza o processo

de evolução do AG, com seus operadores de cruzamento, mutação e seleção, acrescentando uma etapa de transdução viral. [Kubota et al. 1996] [Kubota and Fukuda 1998].

A utilização da metáfora da infecção viral como mecanismo de renovação de diversidade na população do Algoritmo Genético já foi explorada no estudo de Guedes et al. em [Guedes et al. 2005], no denominado AGIV (Algoritmo Genético com Infecção Viral). Nessa variante do AG, a infecção viral é utilizada como um operador que substitui o operador de mutação.

O método consiste em manter uma população de Vírus, cujo genoma é formado por trechos de soluções. No problema do caixeiro viajante, uma rota seria, por exemplo a solução A,F,G,E,G,T,A e um trecho viral seria F,G,E. Para o processo de infecção, um Vírus é selecionado utilizando o processo de *ranking*. Após isso, é realizada a transcrição do genoma viral para um determinado indivíduo (representado apenas por um cromossomo), que passa a ter em seu material gênico um trecho idêntico ao representado pelo agente infectante. [Guedes et al. 2005].

Os vírus possuem um valor chamado de *Nível de Infectibilidade* que indica quão bem o Vírus tem atuado. Quando um Vírus é criado ou modificado, sua infectibilidade é configurada em um valor inicial fixo. Para o Vírus que infecte um indivíduo e aumente a aptidão desse, o nível de infectibilidade desse vírus é incrementado de 1 (um), caso contrário (se diminuir o *fitness* do indivíduo) seu valor é diminuído de 1 (um). O Vírus que tem valor de infectibilidade 0 (zero), descarta suas partes e copia uma parte do cromossomo para si (num processo chamado de transdução). Assim, enquanto população de indivíduos melhora, a população de Vírus tende a ficar com sub-rotas melhores, ou seja, trechos de soluções que são mais desejáveis em uma rota completa. [Guedes et al. 2005].

Goldbarg et al. redefiniram o algoritmo viral VEGA utilizando como base a teoria evolucionária denominada Simbiogênese, a qual enfatiza mais os efeitos das inter-relações entre indivíduos de natureza distinta (incluindo diferentes espécies) do que a sobrevivência do mais apto. Essa teoria diz que a evolução do genoma de um indivíduo a longo prazo é mais influenciada por interações extra-celulares do que pela mutação e seleção natural. Assim, o transporte vertical de material genético via recombinação e mutação foram removidos do processo evolucionário no algoritmo. [Goldbarg et al. 2006].

Nesse algoritmo, considerado como um caso especial e simplificado de uma classe maior de algoritmos evolucionários denominados de Transgenéticos, o processo evolutivo é realizado da seguinte forma: após a criação da população somática, a mesma é avaliada para que sejam selecionados os melhores hospedeiros. A partir desses hospedeiros de elite, são criados vírus onde cada vírus irá infectar toda a população de cromossomos. Após a infecção, esta é avaliada e novos hospedeiros de elite são selecionados para gerarem os novos vírus que procederão com o ataque infeccioso sobre toda a população. Este processo continua até atingir um critério de parada.

Araújo e Moreira em [Araujo and Moreira 2009] fizeram adaptações da metaheurística AGIV e propuseram duas variantes: AGIVB (Algoritmo Genético com Infecção Viral Básico) e AGIVBL (Algoritmo Genético com Infecção Viral e Busca Local). Essas

variantes mantêm idênticas ao AGIV às etapas de geração da População somática inicial, Seleção das soluções pais e Cruzamento (*crossover*). A geração da População Inicial de Vírus dos algoritmos AGIVB e AGIVBL é feita de forma aleatória. Na etapa de Seleção das soluções pais, é utilizada a técnica de Seleção por Torneio. O operador de cruzamento cíclico é utilizado nesses algoritmos e a fase de Infecção Viral substitui o operador Mutação. Após a etapa de infecção, novos vírus podem ser gerados pelo processo de transdução, no qual o vírus se modifica (mutação viral) para melhor se adaptar ao ambiente; ou pelo processo de simbiose virótica, no qual o vírus que infectar um agente e atingir sua taxa de infectividade máxima, irá incorporar um dos memes vizinhos pertencentes ao agente aumentando seu próprio tamanho, transformando-se em um novo vírus com sua taxa de infectividade igual à inicial.

1.3 Motivação e desafios

Problemas com múltiplas dimensões são complexos e exigem alto custo computacional. [Michalewicz 1996]. Utilizar o AG (Algoritmos Genéticos) clássico na solução desse tipo de problema leva a valores ótimos locais, devido a rápida perda de diversidade genética na população. [Falcone 2004].

Uma solução para aumentar a diversidade é incorporar material genético, de forma planejada, de organismos externos à população de indivíduos. Essa recombinação de genes pode ser realizada computacionalmente por meio de uma da metáfora da natureza, utilizando como agente transgenético o vírus, mais especificamente, vírus da família *Retroviridae* que se integram definitivamente ao genoma infectado fornecendo sequencias completas ou parciais de genes.

Moreira em [Moreira 2010] apresentou o modelo computacional, baseado em estruturas virais da família *Retroviridae* para recombinação de material genético e aumento da diversidade populacional, denominado AGRI (Algoritmo Genético Retroviral Iterativo). Nesse estudo, não foi abordado a configuração dos parâmetros da etapa viral, tampouco os aspectos comportamentais do AG com vírus. Ademais, a composição genética dos indivíduos limitava-se a dois cromossomos, o que não possibilitava a análise multi-dimensional dos problemas explorados.

A fim de avaliar o comportamento dos parâmetros da infecção viral do AGRI e comparar seu desempenho frente a outras meta-heurísticas avançadas, foram necessárias algumas modificações no modelo proposto como, por exemplo, habilitar o algoritmo para processar funções multi-dimensionais.

Durante o estudo do algoritmo proposto por Moreira, observou-se que mais modificações poderiam ser introduzidas para otimizar o tempo de processamento, preservar o material genético original da população somática e direcionar a qualidade da população viral, com o objetivo de melhorar os resultados.

Assim, o presente estudo ofereceu como contribuição ao algoritmo de Moreira,

mecanismos de aproximação com os sistemas naturais para solucionar falhas e/ou melhorar a performance do AGRI. As seguintes contribuições foram realizadas:

- o algoritmo adquiriu a capacidade de processamento multi-dimensional, devido o genoma dos indivíduos ser configurável para refletir o número de dimensões desejável;
- o algoritmo adquiriu novos parâmetros configuráveis, tais como: Número de dimensões, Bias da Função e Número de execuções. Além disso, cada cromossomo do indivíduo agora pode ser configurado individualmente (limite inferior, limite superior e precisão);
- os vírus gerados tornaram-se potencialmente mais poderosos, isto é, em uma população viral são criados vírus com genomas que possuem diferentes quantidades de genes, diversificando o tamanho dos trechos de soluções. A adição desta característica ao AGRI emula os vírus na natureza, pois os retrovírus têm tamanho variados podendo ser mais ou menos agressivos. Na versão de Moreira, os vírus sempre tinham quantidade de material genético fixo;
- o material genético viral é gerado individualmente. Para cada vírus a ser criado, um valor aleatório dentro de um intervalo é selecionado. Este valor representa a quantidade total de material genético, que será distribuída de forma aleatória para formar um DNA viral único. Com este método, aumentou-se a capacidade de variabilidade do genoma viral;
- a primeira população viral é criada aleatoriamente, isto é, sem usar material genético da população somática. Este comportamento emula a realidade da natureza, pois quando um indivíduo se expõe a um ambiente, os vírus que estão ali não tem seu material genético. São entidades totalmente independentes. No trabalho de Moreira, em [Moreira 2010], a primeira população viral era criada utilizando o material genético da população de indivíduos;
- nesta versão, os vírus substituídos têm genes novos e genes dos indivíduos mais adaptados, os quais são copiados mantendo suas posições originais. Na versão de Moreira, os vírus substituídos eram gerados somente com o material genético do indivíduo selecionado. Os genes tinham suas posições alteradas (embaralhamento) e alguns eram eliminados, de acordo com a taxa de material genético definida. Essa solução impedia a criação de vírus com bons trechos de soluções derivadas do genoma dos melhores indivíduos;
- outra modificação realizada foi o tratamento aplicado para a evolução da população viral, que deve caminhar junto com a população somática. A contar da segunda geração, uma porção da população viral com pior desempenho será substituída por novos vírus produzidos com material genético apenas dos melhores indivíduos;
- o número de tentativas de infecção para cada indivíduo selecionado foi alterado para preservar o material genético do hospedeiro. Agora, um indivíduo poderá ser

infectado até três vezes numa mesma geração. Anteriormente, na versão de Moreira, cada indivíduo da população somática era submetido ao ataque de toda a população viral, o que mudava totalmente seu material genético logo nas primeiras gerações, além de aumentar bastante o tempo de processamento devido o grande número de tentativas de infecção por indivíduo a cada geração;

- o processo de infecção foi remodelado da seguinte forma: para cada indivíduo selecionado, qualquer vírus da população viral pode ser escolhido para o processo de infecção. Desta maneira, a probabilidade de um vírus infectar mais de um indivíduo se mantém equilibrada, ao mesmo tempo que diferentes vírus podem infectar a mesma população. Este mecanismo controlado mimetiza as seguintes situações da natureza: 1- os indivíduos de uma população dificilmente são infectados por todos os vírus do ambiente em que estão; 2- um vírus ocasionalmente pode infectar mais de um indivíduo no mesmo ambiente; 3- um vírus não infectou nenhum indivíduo de um ambiente;
- o operador mutação do AG clássico também foi ajustado. Nesta versão, a probabilidade de mutação é lançada sobre cada gene do indivíduo, ou seja, um valor aleatório é gerado e caso esse valor seja menor que a taxa de mutação, então o gene é alterado. Caso o número seja maior que essa taxa, então o gene permanece igual. Na versão em [Moreira 2010], a taxa de mutação representava o percentual de genes do indivíduo que seria alterado;

1.4 Objetivos

Os objetivos principais deste trabalho incluem a análise do comportamento do AGRI para problemas com múltiplas dimensões em funções unimodais e, a escolha adequada dos parâmetros da infecção viral para otimizar o desempenho do algoritmo. Como objetivos específicos têm-se:

- Comparar os resultados do AGRI para as funções analisadas, com os resultados do AG clássico e do AG com alta taxa de mutação, a fim de provar que a infecção viral produz material genético de qualidade, a partir do direcionamento de um mecanismo inteligente; e que a infecção viral difere de uma mutação agressiva;
- Comparar o desempenho do AGRI nos problemas analisados, com o resultado de outros algoritmos avançados da Computação Natural, focalizando o alcance do melhor resultado segundo critérios conhecidos de *benchmarking* do *CEC2005* em [Suganthan et al. 2005];
- Conhecer o efeito dos parâmetros da infecção viral no desempenho do AGRI, focalizando a obtenção dos melhores resultados para os problemas analisados.

1.5 Organização do texto

O restante do documento está dividido seguindo o ordenamento descrito abaixo:

- Capítulo 2: Neste capítulo são apresentadas as fontes de inspiração da natureza para a meta-heurística AGRI. Para tanto, os seguintes temas da biologia molecular são explanados: genética, mutação e recombinação de genes na natureza e, o ciclo de vida dos vírus. Além disso, é explicado o processo de infecção da família dos vírus *Retroviridae*, que inclui a etapa de transcrição reversa e a integração definitiva no genoma do hospedeiro.
- Capítulo 3 : Este capítulo explica detalhadamente o mecanismo metafórico da classe particular da Computação Evolutiva denominada Algoritmos Genéticos.
- Capítulo 4: Este capítulo apresenta a meta-heurística AGRI explicando o funcionamento do seu mecanismo viral; que parâmetros compõem sua infecção viral; como funciona a recombinação genética por meio da infecção viral para aumentar a diversidade populacional; qual o ciclo de vida dos vírus durante o processamento e, apresenta a estrutura do Algoritmo Genético com a adição do módulo viral.
- Capítulo 5: Este capítulo apresenta a descrição do problema e o desenrolar dos métodos estatísticos utilizados para gerar os resultados comparativos em três frentes de estudo: 1) a escolha mais adequada da configuração dos parâmetros do AG e da infecção viral para o problema de otimização das funções unimodais multi-dimensionais analisadas; 2) comparação dos resultados do AGRI com os resultados do AG clássico e do AG com alta taxa de mutação; 3) comparação do desempenho do AGRI com o de outras meta-heurísticas avançadas.
- Capítulo 6: São apresentadas as conclusões sobre o desempenho do AGRI e as propostas para continuidade da pesquisa.

CAPÍTULO 2

Referencial Teórico

Os processos naturais simulados ou utilizados como inspiração na solução de problemas têm se mostrado muito eficientes em ajudar a desenvolver e expandir o conhecimento e entendimento para criação de novos mecanismos computacionais. A sequência de explanações sobre os sistemas naturais subsidiará a inspiração da meta-heurística apresentada neste trabalho.

2.1 Sistemas Naturais

Na natureza, os indivíduos de uma população são afetados pelo ambiente em que vivem, a exemplo da escassez de comida e falta de abrigo, e se afetam uns aos outros nas interações de competição por alimento, caça (teia alimentar), acasalamento e outras. Assim, quanto mais adaptado um indivíduo está a essas condições, suas chances de sobreviver por mais tempo e de reproduzir-se aumentam. A prole desses indivíduos herda a informação genética dos pais e, talvez ocorra algum erro de cópia do material genético ao formá-la. Assim, na maioria das vezes, a população adquire as características genéticas dos indivíduos com adaptação acima da média. No entanto, os indivíduos mais suscetíveis também se reproduzem em menor escala, introduzindo material genético novo e possibilitando a criação de descendentes diferenciados.

Nos sistemas naturais não existe uma relação um-para-um entre gene e característica do indivíduo, mas sim a realidade da *pleiotropia*, em que um gene pode afetar vários traços de fenótipo ao mesmo tempo; e da *poligenia* onde uma característica fenotípica é determinada pela interação de vários genes. [Falcone 2004]. Esse extraordinário grau de complexidade e organização nas criaturas viventes gerou a necessidade de estudar-se as coisas vivas no nível dos genes, moléculas e átomos, assim criou-se o campo da ciência denominado biologia molecular. [Tower 2012].

2.1.1 **Biologia Molecular**

A biologia molecular tem mostrado que todos os organismos vivos apresentam um plano único principal de organização, isto é, apresentam como unidades estruturais as células e moléculas.

A célula é a unidade estrutural e funcional básica dos organismos vivos. Com efeito, uma única célula pode constituir um organismo inteiro, como é o caso dos protozoários. Agrupadas e diferenciadas constituem organismos multicelulares como tecidos e órgãos. [Lemos and Basílio 2000].

O projeto básico do sistema celular é essencialmente o mesmo em todos os sistemas vivos na Terra, desde as bactérias até os mamíferos. Em todos os organismos, os papéis do DNA, do RNA mensageiro e das proteínas são exatamente iguais. Além disso, o significado do código genético também é virtualmente idêntico em todas as células, assim como o tamanho, a estrutura e o projeto bioquímico básico do mecanismo sintético e protéico é praticamente o mesmo em todas elas. Com efeito, em termos de seu projeto bioquímico básico não há nenhum sistema vivo primitivo ou ancestral em relação a qualquer outro sistema. [Tower 1987].

Não obstante a complexidade das milhões de diferentes espécies - cada uma com sua própria morfologia e informação genética específica - é possível a simplificação através da classificação das células. Assim, todas as células vivas podem ser classificadas basicamente em duas categorias principais:

Procarióticas: não possui envoltório nuclear. O cromossomo desta célula está dentro de um espaço denominado nucleóide, que está em contato direto com o citoplasma. As células bacterianas são procarióticas, ou seja, não possuem núcleo. A membrana externa que protege as células procarióticas é resistente e flexível, para proteger seu interior de um ambiente perigoso e ao mesmo tempo permitir que moléculas pequenas de oxigênio entrem e saiam dela e outras moléculas úteis não saiam. Essa seleção é feita por moléculas proteicas especiais.

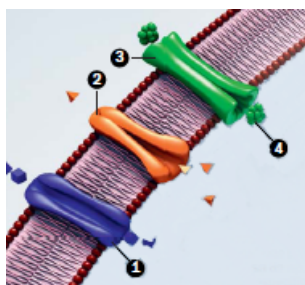
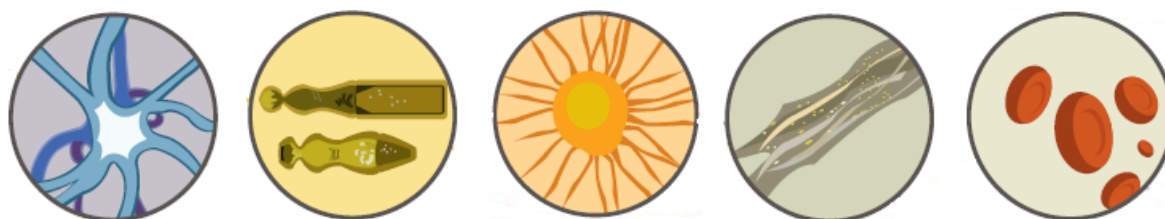


Figura 1: Substâncias específicas passam pela membrana celular
Fonte: [Tower 2010]

A figura 1 ilustra esse mecanismo de seleção, no qual algumas proteínas (1) possuem um orifício que permite a entrada e saída de tipos específicos de moléculas. Outras

são abertas em um lado da membrana (2) e fechadas na outra extremidade. Além disso, elas tem uma abertura (3) projetada para uma substância específica. Quando essa substância chega, a outra extremidade da proteína se abre e permite a passagem. [Tower 2010].

Eucarióticas: possuem núcleo verdadeiro com envoltório nuclear elaborado, por meio do qual ocorrem trocas entre o núcleo e o citoplasma. As células humanas, animais e vegetais possuem núcleo. [Tower 2010].



(a) Célula Cerebral (b) Células do Olho (c) Célula óssea (d) Célula Muscular (e) Célula Sanguinea

Figura 2: Tipos de células no corpo humano

Fonte: [Tower 2010]

Conforme apresentado na figura 2, as células com núcleo especificam-se em diferentes tipos e apresentam uma enorme diversidade de formas e funções. A união de tipos específicos de células eucarióticas constituem uma rede complexa e integrada para formar organismos complexos, como é o caso dos órgãos humanos.

Ademais, a organização molecular dos seres vivos está baseada na química do carbono. Para as diferentes moléculas, os átomos de carbono são arranjados de forma distinta. Dentre as macro-moléculas que constituem praticamente todas as estruturas celulares estão os ácidos nucleicos.

Em relação aos organismos vivos, todos contêm ácidos nucleicos na forma de DNA (ácido desoxirribonucleico) e RNA (ácido ribonucleico). Esses ácidos resultam da repetição de quatro diferentes unidades, denominadas nucleotídeos. A tabela 1 expressa as principais diferenças entre esses dois ácidos.

Tabela 1: Principais diferenças entre o DNA e o RNA

	DNA	RNA
Localização	Primariamente no núcleo da célula, mas também é encontrado nas mitocôndrias e cloroplastos	Encontrado no citoplasma, nucleólo e cromossomos
Bases Pirimídicas	Citosina e Timina	Citosina e Uracila
Bases Purínicas	Adenina e Guanina	Adenina e Uracila

Os ácidos nucleicos são compostos por uma molécula de açúcar (pentose), bases nitrogenadas (purinas e piridiminas) e ácido fosfórico. As pentoses são de dois tipos: ribose no RNA e desoxirribose no DNA. As bases nitrogenadas são também de dois tipos: piridiminas e purinas.

Ademais das diferenças químicas, esses dois ácidos se distinguem nas características estruturais e nas funções que desempenham nos organismos, sendo responsáveis por exemplo pela manutenção do material genético ou pela formação das proteínas, as quais estimulam ou diminuem a atividade de outras partes do corpo.

2.1.1.1 Ácido Desoxirribonucléico - DNA

O DNA é o principal armazenador de informações genéticas e um dos mecanismos químicos responsáveis pela transmissão das características de organismos de uma geração para a seguinte.

As moléculas do DNA contêm informações codificadas que servem de base para todas as formas de vida e asseguram a biodiversidade. Elas também contêm instruções elaboradas e detalhadas necessárias para o desenvolvimento correto dos organismos, como coordenar a reprodução e outros processos como a formação, o crescimento, a manutenção das células. Algumas sequências de letras são exclusivas dentro do DNA de cada pessoa, visto que o DNA contém as suas informações hereditárias. [Tower 2010].

O DNA é composto por duas cadeias helicoidais de polinucleotídeos com giro para a direita, formando uma hélice dupla em torno de um mesmo eixo central que lembram longas escadas em espiral. As duas fitas são antiparalelas, unidas por pontes de hidrogênio estabelecidas entre os pares de bases que podem ser comparados a "degraus" químicos. A distância fixa entre as duas moléculas de açúcar, nas fitas opostas, permite que somente os pares de bases AT e CG possam se encaixar na estrutura. Assim, um alfabeto simples de quatro letras é formado. Francis Crick, biólogo molecular e co-descobridor da estrutura de dupla hélice do DNA, sugeriu que é a *sequencia* linear dos degraus químicos que forma as instruções codificadas. [Lemos and Basílio 2000] [Tower 2015].

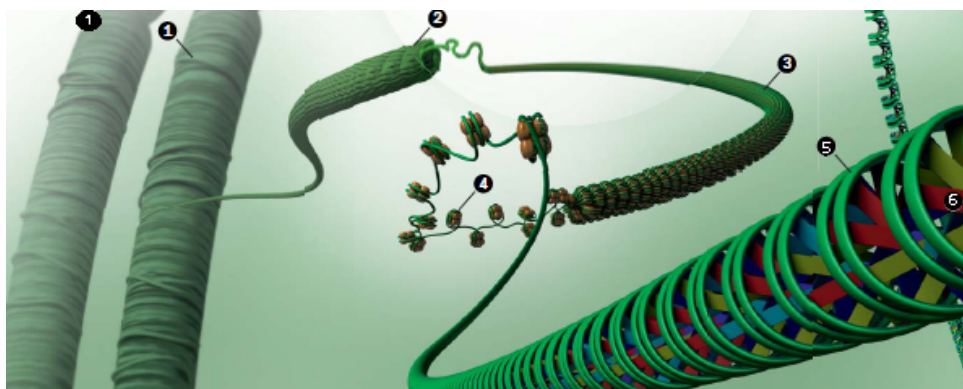


Figura 3: Armazenamento do DNA no núcleo da célula

Fonte: [Tower 2010]

Acompanhando a figura 3, observa-se que os cromossomos estão organizados em pares idênticos (1). O armazenamento do DNA é feito, ilustrativamente, como cordas em volta de toda extensão do cromossomo, onde cada corda horizontal é formada por linhas horizontais menores e dividida por linhas verticais (2). Cada linha é composta por pequenas espirais (3) entrelaçadas. Os espirais são formados por carretéis (4) enrolados pelo o DNA que tem formato parecido a uma escada de corda em espiral (5). Os degraus dessa escada são conhecidos como pares de bases (6).

A ordem em que as letras A, T, G, C aparecem no DNA forma códon, estes são organizados em uma sequência chamada genes. Cada gene contém em média 27 mil letras. Os genes e os longos espaços entre eles são organizados em agrupamentos, denominados cromossomos. São necessários 23 cromossomos para formar o genoma humano - toda informação genética do organismo. As células contém duas cópias completas do genoma, totalizando 46 cromossomos. [Tower 2010].

O DNA é duplicado para outras células quando uma célula se divide ou quando um organismo se reproduz. No caso da cópia para substituir células desgastadas ou doentes, cada nova célula terá as mesmas informações genéticas. Assim, durante esse tipo de duplicação do DNA, estruturas moleculares chamadas enzimas percorrem a sua extensão dividindo seus filamentos, depois cada um desses age como um molde para a síntese da nova cadeia complementar. Portanto, a sequência de nucleotídeos numa das fitas determina completamente a molécula de DNA. Nesse processo de cópia, o DNA é 'revisado' várias vezes e os erros detectados são corrigidos. Assim, o que permite sua auto-duplicação é a propriedade das bases de DNA serem complementares.

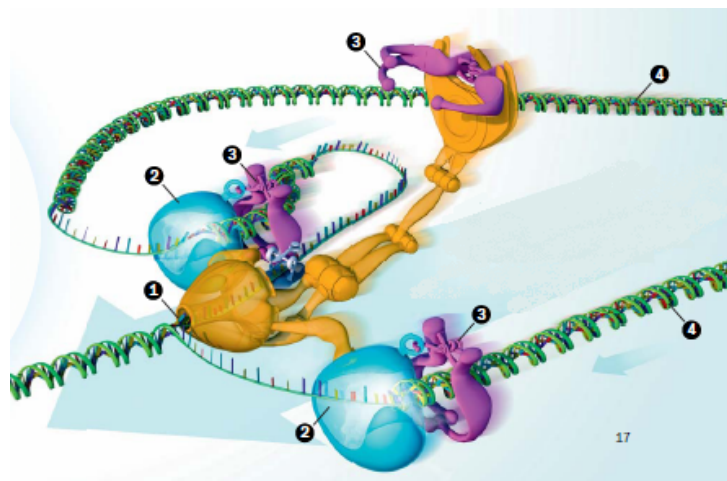


Figura 4: Máquina de Enzimas que divide o DNA em dois filamentos
Fonte: [Tower 2010]

Na Figura 4 observamos o seguinte processo de replicação do DNA: 1- A "máquina de enzimas" divide o DNA em dois filamentos. 2- Essa parte da máquina usa um único filamento do DNA como molde para criar o filamento complementar. 3- Um cursor em forma de anel guia e estabiliza a "máquina de enzimas". 4- Dois filamentos completos de

DNA são formados.

Visto que, todas as coisas vivas na Terra se reproduzem usando o DNA, o código que controla muitas de suas características herdadas é muito estável e tem garantido a distinção entre as espécies básicas de criaturas ao longo das eras.

Sobre os genes humanos, acrescenta-se que estudos foram desenvolvidos baseando-se num tipo de DNA mitocondrial, material genético transmitido apenas pela mulher. Foi feita uma comparação entre os padrões genéticos humanos ao redor da Terra e encontraram-se evidências claras de que todos os humanos têm um ancestral comum, uma fonte do DNA de todas as pessoas que já viveram e de cada um de nós. [Tierney 1988].

Na prática, os vários organismos da Terra continuam a executar suas funções especificamente projetadas na complexa teia da vida.

2.1.1.2 Ácido Ribonucléico - RNA

O RNA é uma molécula composta por uma única cadeia ou filamento, quimicamente similar à molécula do DNA. Ela está dividida em três principais classes: o RNA mensageiro (mRNA), o RNA de transferência (tRNA) e o ribossômico (rRNA). Seu trabalho é apanhar informações codificadas nos genes (DNA) e leva-las até um ribossomo para a fabricação de proteínas. Este processo é chamado de transcrição. [Tower 2010].

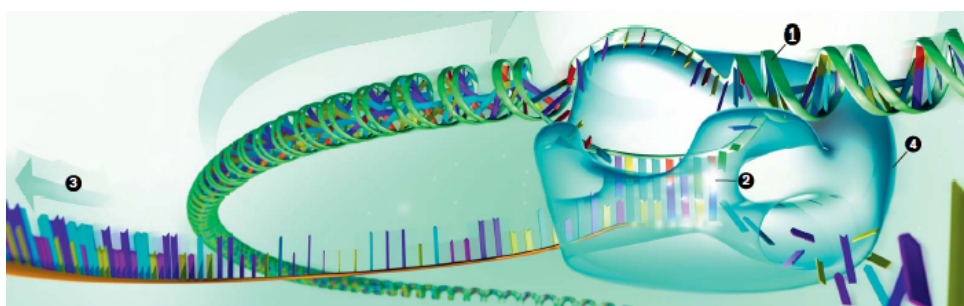


Figura 5: Um filamento exposto de DNA transmite informações para o RNA

Fonte: [Tower 2010]

Na Figura 5, o processo de transcrição pode ser descrito da seguinte forma: 1- O DNA é desenrolado e um dos filamentos transmite informações para o RNA. 2- O RNA "lê" o DNA, apanhando o código em um gene. O código do DNA diz à máquina de transcrição onde começar e onde parar. 3- Carregado de informações, o RNA deixa o núcleo da célula e se dirige a um ribossomo, onde transmitirá as instruções para a fabricação de uma proteína complexa. 4- Máquina de transcrição.

Toda proteína cumpre uma função específica, determinada pelo seu gene de DNA. Logo, o processo de transcrição é altamente seletivo, visto que copia apenas algumas informações do gene, as quais serão decodificadas para produzir uma proteína específica. O comprimento da porção codificadora de um gene depende da extensão da mensagem a

ser traduzida, isto é, o número de aminoácidos da proteína.

A transcrição inicia quando as polimerases do RNA (enzimas catalizadoras da reação) se ligam à uma sequência específica de bases de inicialização, nas extremidades do DNA. Os filamentos do DNA são desenrolados e a enzima de RNA "lê" os nucleotídeos do DNA até encontrar a sequência de terminação (bases específicas que determinam o fim da transcrição) e, produz (adiciona) as bases correspondentes no chamado mRNA, alongando sua cadeia.

Após isso, o mRNA se afasta e o DNA volta a ter seus filamentos enrolados. O tRNA reconhece o código do RNA mensageiro e apanha, com a ajuda de uma enzima, as moléculas de aminoácidos corretas para fabricar a proteína específica exigida (cada conjunto de três letras de RNA forma uma "palavra" que exige um aminoácido específico), depois o tRNA sai do núcleo da célula para o citoplasma e transporta-as até o ribossomo, onde a informação será usada para formar a proteína.

O rRNA fornece um suporte molecular para as reações químicas da montagem da proteína. Os ribossomos "leem" as sequências de nucleotídeos do rRNA e montam a ordenação específica de aminoácidos em um filamento ou cadeia. Tais aminoácidos ligam-se em ângulos específicos e formam padrões definidos; depois se enroscam e se entrelaçam criando uma estrutura tridimensional mais complexa, resultando num certo tipo de proteína. Esse processo final é chamado tradução do RNA. [Lemos and Basílio 2000] [Gonçalves 2014].

As proteínas são as mais abundantes macromoléculas presentes nas células e são responsáveis por muitas funções vitais, por exemplo: a hemoglobina que está presente nos glóbulos vermelhos do sangue para transportar oxigênio por todo o corpo; e a insulina que ajuda a metabolizar os alimentos e a controlar várias funções celulares. Assim, toda molécula tem uma função específica e um lugar designado no esquema geral, para a produção dos materiais corretos. [Tower 1999].

2.1.2 Vírus

Os vírus são entidades intra-celulares obrigatórias, cujo genoma (RNA ou DNA) é replicado no interior das células vivas por meio da máquina sintética celular, a qual sintetiza partículas que podem transferir o genoma infectado para as outras células.

O vírus não tem metabolismo, não produzem energia, não crescem e não se dividem, mas limitam-se a fornecer o material genético para a célula infectada que o replicará. Seu tamanho varia de 20nm (nanômetros) até 459 nm. A nomenclatura do vírus depende se o mesmo está dentro ou fora da célula. Quando está no meio intracelular é dito vírus, mas quando se encontra no meio extra-celular é denominado virião ou partícula viral. [Pereira 2007].

As partículas virais possuem uma cápsula protéica, chamada cápside, que liga o virião à célula-alvo e envolve o material genético viral, denominado nucleóide. Muitos vi-

riões também são revestidos por um invólucro derivado da membrana celular no processo de saída do vírus. A Figura 6 exemplifica as características morfológicas dos vírus que podem ser esféricos, cilíndricos, cúbicos e espermatozoidais.

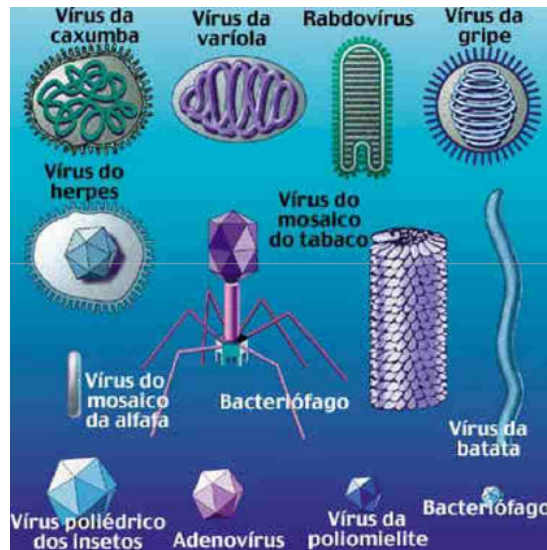


Figura 6: Alguns tipos de vírus

2.1.2.1 Ciclo Viral

O ciclo de infecção e reprodução dos vírus inclui as seguintes etapas [Pereira 2007]:

Adsorção: os viriões possuem proteínas de ligação que promovem o processo de adsorção da partícula viral pelos receptores da membrana celular. O tipo de receptor é um dos fatores que tornam as células suscetíveis à infecções e, determinam a gama de hospedeiros de cada vírus.

Penetração: pode ocorrer por 3 processos distintos. Alguns vírus penetram por translocação da partícula viral para o citoplasma, através da membrana. Outros vírus fazem-no por fusão do invólucro com a membrana celular. Outros são internalizados em endossomas, através de endocitose mediada por receptores.

Descapsidação: quando o virião já está no interior da célula, ocorre o processo de disrupção total ou parcial, em que é liberado o material genético viral. Este processo pode ocorrer no citoplasma, imediatamente após a penetração; ou junto ao núcleo, no caso de vírus de replicação nuclear, como os adenovirus ou os herpesvirus.

Fase sintética: todas as proteínas virais, sejam enzimáticas, reguladoras ou estruturais são sintetizadas e ocorre a replicação do genoma viral. Em alguns vírus, por exemplo os papovirus, esta fase é simples dado que são sintetizadas apenas as proteínas

estruturais e algumas proteínas que cooperam com as enzimas celulares para a replicação do DNA. Já nos vírus de RNA, é necessária a síntese de enzimas virais para replicarem ou transcreverem o RNA viral, função para que a célula infectada não está equipada. Em outros vírus mais complexos, como os poxvirus ou os herpesvirus, o genoma é codificado para várias enzimas homólogas de enzimas celulares e para outras funções, por exemplo enzimas envolvidas na indução da morte celular programada (apoptose) ou na defesa do vírus contra a resposta imune do hospedeiro.

Montagem e extrusão: na maioria dos vírus, os componentes das partículas são sintetizados como elementos mais ou menos complexos no local de replicação - núcleo ou citoplasma. Nos vírus de RNA de polaridade negativa e nos retrovírus, as partículas imaturas são encaminhadas para a periferia da célula junto da membrana em que foram inseridas as proteínas virais, ali ocorrem outras fases de montagem das partículas que terminam com a aquisição do invólucro e a saída da célula por gemulação.

2.1.3 Retrovírus

A família *Retroviridae* compreende vírus que utilizam RNA-dependente-DNA-polimerase, isto é, seu material genético é constituído por RNA e uma enzima chamada transcriptase reversa, que faz uma transcrição inversa produzindo uma molécula de DNA a partir do seu RNA. [Gonçalves 2014]. O processo de infecção viral e aquisição do seu material genético como parte da célula é ilustrado na figura 7.

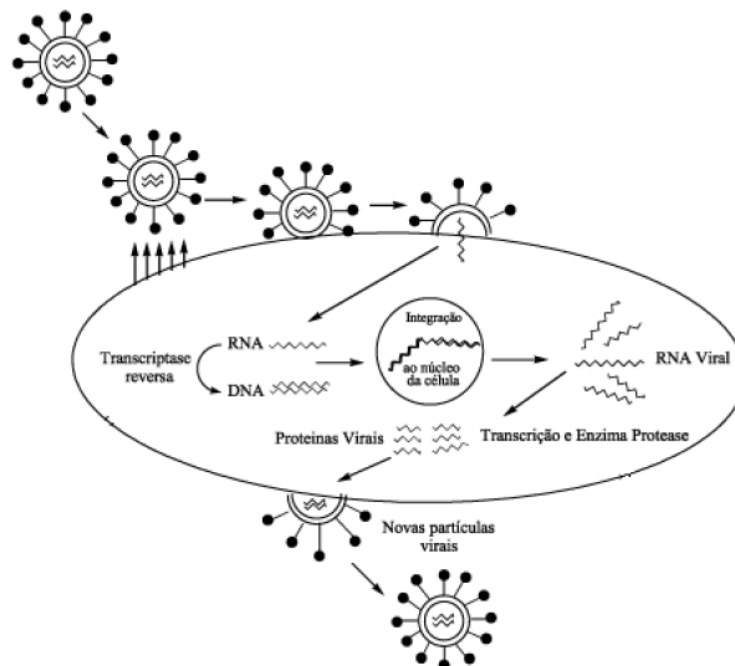


Figura 7: Replicação viral com a etapa de transcrição reversa.

Uma vez que ocorre a infecção da célula hospedeira, a enzima transcriptase reversa utiliza os nucleotídeos presentes no citoplasma da célula para montar um filamento de DNA usando como base o material genético contido no filamento de RNA infectado. Após a produção do DNA, o vírus se integra ao cromossomo da célula infectada e ocorre a síntese de proteínas virais, seguindo o processo normal da síntese proteica (DNA-RNA-proteína).

De acordo com o Comitê Internacional de Taxonomia de Vírus (ICTV), existem sete gêneros da família *Retroviridae*: Alpharetrovírus, Betaretrovírus, Deltaretrovírus, Gammaretrovírus, Epsilonretrovírus, Spumavírus e Lentivírus. Os vírus de cada gênero se diferenciam pela organização do seu genoma, presença ou ausência de oncogenes e genes acessórios, forma do capsídeo e organismo que os hospeda. [Conffin et al. 1997].

Relata-se que no genoma dos organismos eucariotos existem fragmentos de DNA chamados transposons (TE - Transposable Elements) capazes de se translocar de uma região para outra do genoma num processo chamado de transposição. [Wicher et al. 2007]. Os transposons podem se auto-replicar (retrotransposons) e se inserir em qualquer região do genoma, incluindo regiões reguladoras e codificadoras de proteínas. Atualmente, tais elementos tem papel reconhecido na evolução dos genomas hospedeiros por participarem na regulação gênica e contribuírem para o aumento de variabilidade. Entre os tipos de retrotransposons mais notáveis estão os retrovírus endógenos (ERVs - endogenous retroviruses) que possuem no seu genoma viral o gene envelope (*env*), capsídeo (*gag*), protease e polimerase (*pro-pol*) e regiões de ataque não-codificadoras denominadas LTR (do inglês Long Terminal Repeat). [Romano 2009].

Os retrovírus geralmente infectam células somáticas e possuem uma fase denominada "provírus", na qual os ERVs dos respectivos retrovírus exógenos (gênero da família) se integram às células do hospedeiro. Os provírus colonizam definitivamente seus genomas e são transmitidos aos descendentes dos indivíduos infectados. [Conffin et al. 1997] [W-H and D. 1991]. Relata-se que o genoma humano possui mais de 98.000 sequências (completas ou parciais) de retrovírus endógenos de diversas famílias. [J. et al. 2002] [S. et al. 2001].

Após a infecção, os ERVs retêm, por um tempo, a capacidade de replicação dentro do genoma do hospedeiro, a qual pode ocorrer por retro-transposição ou reinfeção. A retro-transposição ocorre na mesma célula e é realizada pela replicação ativa do genoma viral promovida pelo LTR. A reinfeção depende do gene envelope para invadir outras células. A velocidade de proliferação de um ERV tende a ser muito maior no período após a infecção inicial, declinando com o tempo devido aos mecanismos de defesa do organismo hospedeiro, que realiza substituições deletérias ao longo do genoma proviral tornando os ERVs inativos. [Conffin et al. 1997].

Todavia, uma vez que o genoma hospedeiro é invadido por retrotransposons e retrovírus endógenos, ele acaba incorporando proteínas e promotores de retroelementos na sua malha gênica, o que tem sido chamado de "domesticação molecular". [I.K. et al. 2003]. A partir disso, uma das atuações dos retrotransposons na expressão gênica seria como me-

diador de variações fenotípicas. Por exemplo, Whitelaw e Martin mostraram que a cor da pelagem nos camundongos pode ser influenciada por fatores epigenéticos, ativados pelo comportamento de certos retrotransposons. [Whitelaw and Dik 2001].

Notadamente, dentre os efeitos positivos dos ERVs no organismo hospedeiro estão a geração de novos genes, bem como o aumento da variabilidade entre os organismos. Retroelementos com integração polimórfica e solo-LTRs (LTRs solitárias geradas por recombinação) contribuem para a variação alélica na população. [Romano 2009]. Por exemplo, uma das regiões do genoma humano que apresentam maior variabilidade e densidade de retrovírus endógenos é o MHC (complexo principal de histocompatibilidade) humano. O MHC está localizado no cromossomo 6, tem extensão de aproximadamente 4×10^6 pares de bases e é a região geneticamente mais densa do genoma. Ali são encontrados ERVs de diferentes famílias em uma proporção de 10:1 em relação ao restante do genoma. [Kulski et al. 2008].

2.2 Considerações Finais

A genética e a biologia molecular modernas têm suprido ampla evidência da enorme complexidade e interdependência das relações entre DNA, RNA e proteínas. Esses achados implicam que a vida depende de ter todos esses elementos ativos *simultaneamente*. De fato, uma função-chave do DNA e do RNA é produzir a enorme variedade de proteínas necessárias para a vida.

Quanto aos vírus, embora se assemelhem mais a um conjunto químico do que a um organismo vivo, têm a capacidade de infectar todos os tipos de organismos podendo incorporar seu genoma ao da célula hospedeira, como é o caso dos vírus de RNA. Tais vírus têm sido reconhecidos como fonte de renovação da malha gênica, devido a rápida evolução que sofrem por dois fatores principais: a mutação e o ressortimento.

A mutação viral pode ocorrer por erros durante a replicação dos nucleotídeos do DNA ou RNA infectados. No ressortimento, ocorre infecções com variantes de um mesmo vírus na mesma célula. A progênie viral pode conter alguns segmentos de um parente e outros do outro parente. Este processo limita-se aos vírus com genomas segmentados, como é o caso dos vírus de RNA nos humanos. [Roehe 2015].

Tal avanço no entendimento dos mecanismos genéticos e o estudos de outros projetos da natureza, tem sido usados como fonte de inspiração - metáforas - para solucionar problemas em várias áreas. O desenvolvimento tecnológico permitiu que aplicações práticas fossem desenvolvidas por meio do estudo e copia dos detalhes de projeto de vários organismos, resultando na criação de novos produtos e no melhoramento do desempenho dos já existentes.

Assim como outras áreas do conhecimento científico, a computação tem uma linha de pesquisa denominada *Computação Natural*, que utiliza abordagens baseadas em ideias da natureza para o desenvolvimento de sistemas "artificiais", ou utiliza materiais e

mecanismos naturais para realizar computação. [Castro 2012].

A sub-área denominada *Computação Evolucionária* compreende diversos algoritmos inspirados na genética moderna, os quais fornecem um mecanismo de busca paralela e adaptativa. Alguns exemplos de algoritmos inspirados na natureza são: Algoritmos Genéticos, Programação Genética e Eletrônica Evolucionária. Estes são algoritmos gerais, probabilísticos, aplicados em problemas complexos, com grandes espaços de busca, de difícil modelagem, ou para os quais não há um algoritmo eficiente disponível.

O capítulo seguinte faz uma explanação sobre a abordagem denominada Algoritmos Genéticos. Nessa técnica de busca clássica, alguns conceitos apresentados anteriormente sobre herança genética são fortemente utilizados. Os cromossomos são palavras (binárias, vetores, matrizes etc) que codificam soluções para um problema. Operadores heurísticos baseados na teoria da evolução (seleção, reprodução, cruzamento e mutação) são aplicados por vários ciclos para melhorar as soluções geradas com o objetivo de otimizar problemas propostos.

CAPÍTULO 3

Computação Evolucionária

A computação evolucionária propõe um paradigma alternativo ao processamento de dados convencional para resolver problemas, a partir de mecanismos evolutivos encontrados na natureza, tais como a auto-organização e o comportamento adaptativo de organismos sociais.

Tais métodos heurísticos permitiram o desenvolvimento de técnicas de otimização para busca de soluções ótimas em diferentes espaços amostrais, inclusive vários com máximos e mínimos locais. Dentre essas heurísticas, os algoritmos evolucionários destacam-se em solucionar problemas de otimização de múltiplas soluções. Dentre esses, os algoritmos genéticos são vistos como eficientes otimizadores. O mecanismo de funcionamento desta técnica é explicado detalhadamente a seguir.

3.1 Introdução a Algoritmos Genéticos

O AG (Algoritmos Genéticos) é um método robusto de busca probabilística inspirado nos mecanismos de seleção natural, cujo processo biológico implica no aumento das chances de sobrevivência dos indivíduos mais fortes num ambiente competitivo específico, preservando as características hereditárias desses na geração seguinte. Com o passar das gerações, sucessivos cruzamentos e mutações ocorrem, assim os indivíduos tendem a estar cada vez mais adaptados ao meio ambiente em que vivem.

Computacionalmente, a mesma terminologia e os mesmos princípios da genética e da teoria da seleção natural são empregados. Nessa analogia com os sistemas naturais, a genética artificial caracteriza o conjunto de elementos e o conjunto de valores utilizados na resolução de um problema, conforme a Tabela 2:

Tabela 2: Analogia entre sistemas naturais e artificiais

Genética Natural	Genética Artificial
Geração	Iteração onde uma população é submetida ao processo de seleção, aplicação dos operadores genéticos e avaliação de adaptação
População	Conjunto de indivíduos
Indivíduo	Composto por um ou mais cromossomos
Cromossomo	Conjunto de genes
Gene	Caractere no cromossomo
Alelo	Valor do caractere de um gene
Locus	Posição do gene no cromossomo
Genótipo	Estrutura completa de genes, isto é, o próprio indivíduo
Fenótipo	Solução numérica real, estrutura decodificada, conjunto de parâmetros
Cruzamento	Troca de material genético entre dois indivíduos
Mutação	Inserção de material genético novo no indivíduo
Biodiversidade ou Diversidade genética	Variação alélica dos indivíduos de uma população

O AG trabalha com uma população, onde cada indivíduo representa uma possível solução codificada do problema. Cada indivíduo é composto por valores, denominados genes. A função *objetivo* (ou de desempenho) representa o ambiente no qual a população inicial encontra-se. Cada indivíduo é avaliado por essa função e o valor da aptidão de cada um é utilizado no processo de seleção para determinar quem poderá se reproduzir. Os operadores genéticos cruzamento e mutação são aplicados sobre os indivíduos selecionados. Novos indivíduos são gerados e substituem os anteriores, finalizando assim uma geração.

Por meio dos operadores genéticos, determinadas características compartilhadas entre indivíduos (sequências de genes, denominadas esquemas), serão avaliadas como desejáveis ou não e podem ser propagadas conforme a manipulação da frequência com que aparecem nas populações no decorrer das gerações. O mecanismo de avaliação prossegue ciclicamente, a partir da nova população formada e termina quando algum critério de convergência é alcançado.

Dependendo do problema, diferentes métodos podem ser usados para determinar a convergência de um algoritmo genético, os quais podem ser: um valor ótimo (previamente conhecido), uma solução aceitável encontrada, o número máximo de gerações é atingido ou não ocorrer melhora na aptidão da população dentro de um determinado número de gerações. [Lucas 2002].

Resumidamente, a sequência do processamento de um AG clássico pode ser descrita como mostra o fluxograma da Figura 8.

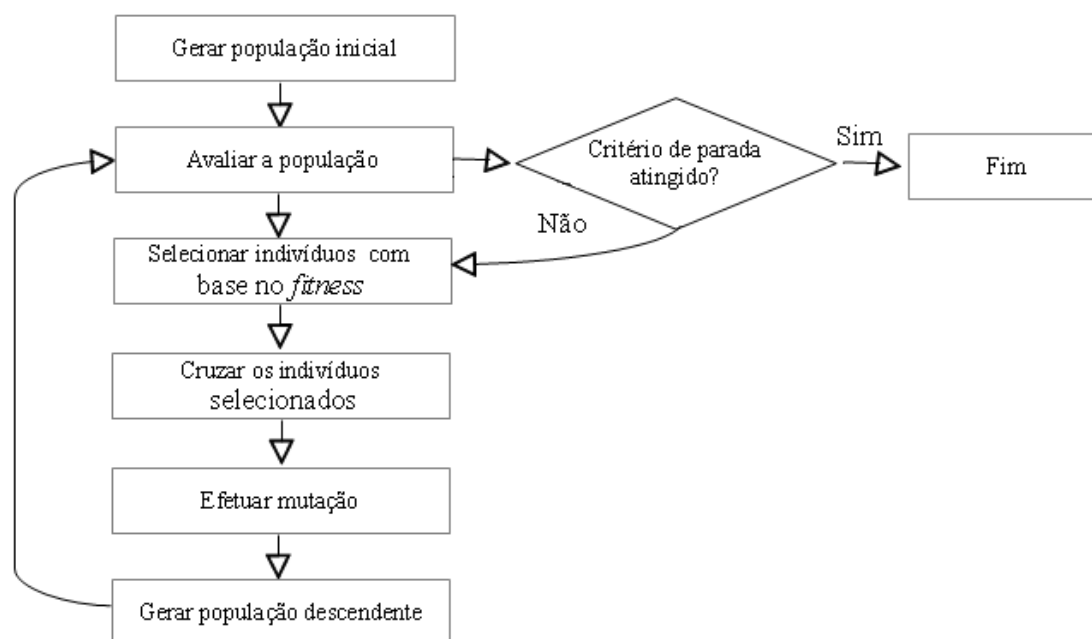


Figura 8: Fluxograma de um AG Clássico.

Deve-se determinar uma forma de avaliação da aptidão de cada indivíduo em relação ao ambiente, esta aptidão representa o quão próximo da solução ótima está a solução representada pelo indivíduo desse problema.

No AG, a característica representada pelo indicador *grau de adaptação* ou *fitness* informa o quão bem a resposta representada pelo indivíduo soluciona o problema proposto. Este indicador é calculado pela função *objetivo* (normalmente denotada $f_O(x)$). Já o indicador *grau de aptidão* está relacionado com o nível de adaptação de um indivíduo em relação a população que ele pertence. O grau de aptidão de um indivíduo x é dado por:

$$f_A(x) = \frac{f_O(x)}{\sum_{i=1}^n f_O(i)} \quad (3.1)$$

sendo n o tamanho da população.

Assim, o AG tem como objetivo fazer a população convergir para um valor ótimo de adaptação. Para isso, é calculada a *média de adaptação* da população na geração atual, a qual é comparada com a média de adaptação da população de gerações anteriores, isto é denominado *grau de convergência*. A média de adaptação é dada por:

$$MA = \frac{\sum_{i=1}^n f_O(i)}{n} \quad (3.2)$$

sendo n o tamanho da população.

3.2 Componentes do AG

3.2.1 Indivíduos

Codificam possíveis soluções para o problema a ser tratado. Tais respostas são encontradas no decorrer das gerações, por meio da manipulação ou aplicação do processo de seleção, cruzamento e mutação.

Um estudo do problema a ser otimizado é vital para definir a quantidade de indivíduos que terá a população, a formação e representação cromossômica desses indivíduos, além das probabilidades de aplicação dos operadores genéticos sobre esses, visto que a escolha da representação das soluções afeta diretamente o desempenho do programa.

Muito comum é associar a solução de um problema a um cromossomo p representado por um vetor com n posições, tal que:

$$p = (x_1, x_2, x_3, \dots, x_m) \quad (3.3)$$

Cada elemento x_i representa um gene. Tal definição implica em que um indivíduo se resume ao conjunto de genes que possui (seu genótipo). No entanto, toda avaliação de um indivíduo se baseia em seu fenótipo, ou seja, no valor observável resultante do processo de decodificação de genes. [Lucas 2002] [Falcone 2004].

Os tipos de representação mais comuns para um cromossomo são:

Representação Binária: é clássica e tem sido usada desde o início do AG.

Representação por Ponto Flutuante: apresenta maior precisão e capacidade de representação de domínios de um problema, é utilizada especialmente em problema com parâmetros reais. A utilização desta representação melhora a velocidade do processamento, a consistência dos resultados entre gerações e a precisão, se comparada com a binária. [Michalewicz 1996].

Representação por Inteiros: muito adequada para os problemas que envolvam números inteiros, tais como análise combinatória, arranjo e permutação. [Soares 1997].

A Tabela 3 apresenta o resultado do processo de decodificação de genes que será submetido à avaliação pela função objetivo, assim como a indicação da representação para os tipos de problemas a serem otimizados.

Tabela 3: Exemplos da decodificação de algumas representações de genótipos em fenótipos

Genótipo	Fenótipo	Problema
0010101001110101	10869	otimização numérica
CABDEF	comece pela cidade C, depois passe pelas cidades A,B,D,E e termine na cidade F	caixeiro viajante
$C_1R_2C_3R_4C_5R_1$	se condição 1 (C_1) execute a regra 2 (R_2), se (C_3) execute a regra (R_4), se (C_5) execute a regra (R_1)	regras de aprendizagem para agentes

3.2.2 População Inicial

A população inicial é gerada, comumente, por meio de funções geradoras de números aleatórios com distribuição uniforme. No entanto, algoritmos de busca heurística também podem ser utilizados para esta finalidade, especialmente nos problemas com alto grau de restrição em que não existem indivíduos ótimos, apenas indivíduos com algumas características desejáveis.

O desempenho do AG é afetado pelo tamanho da população, visto que uma população pequena pode reduzir o espaço de busca, diminuindo a probabilidade de o ótimo global ser alcançado. No entanto, populações grandes aumentam bastante o tempo de processamento.

Segundo Goldberg em [Goldberg 1989] e Geyer-Shultz em [Geyer-Shultz 1997], os operadores de inicialização de população mais tradicionais são:

Inicialização randômica uniforme: cada gene de cada cromossomo do indivíduo receberá como valor um alelo, sorteado de forma aleatoriamente uniforme.

Inicialização randômica não uniforme: determinados alelos tendem a ser escolhidos com mais frequência que os demais.

Inicialização randômica com "dope": indivíduos otimizados são inseridos entre a população gerada aleatoriamente. Este tipo de inicialização promove o aparecimento super indivíduos, que tendem a dominar no processo evolutivo, causando a convergência prematura.

Inicialização parcialmente enumerativa: indivíduos são inseridos na população com o objetivo de que o processo de evolução inicie com todos os esquemas possíveis de determinada ordem.

3.2.3 Avaliação

A etapa de avaliação e a escolha da representação dos indivíduos constituem as partes do AG que são mais dependentes do problema a ser otimizado. Cada indivíduo da população é avaliado para que seja determinado seu grau de adaptação, a partir daí o processo de seleção é iniciado para indicar os melhores indivíduos. Várias formas de avaliação podem ser utilizadas, dependendo do problema a ser solucionado. Vejamos:

Valor de retorno da função objetivo para otimização de funções matemáticas seria dado por

$$f_O(G) = f(d(G)), \quad (3.4)$$

onde $f_O(G)$ é a função objetivo aplicada sobre o genoma G e $d(G)$ é a função que decodifica os genes desse genoma em um valor x .

No problema do caixeiro viajante, o grau de adaptação seria dado por

$$f_O(G) = \frac{1}{\sum_{i=1}^{l_G-1} distancia(g_i, g_i + 1)} \quad (3.5)$$

onde g_i representa o i -ésimo gene do genoma G e l_G é o comprimento total desse genoma.

Para problemas de escalonamento que visam impedir a violação de restrições, pode-se atribuir um valor de penalidade para cada restrição violada. Assim, o valor de adaptação do indivíduo corresponde ao inverso da soma de todas as penalidades que ele sofreu

$$f_O(x) = \frac{1}{\sum_{j=1}^P penalidade_{ij}} \quad (3.6)$$

onde i representa o índice do indivíduo avaliado, j a penalidade associada à regra que ele violou e P representa o total de penalidades das Regras violadas.

3.2.4 Seleção

Para Dawkins em [Dawkins 1996], o mecanismo da seleção é responsável por passar as boas características dos indivíduos para as gerações seguintes. Por isso, esse processo é dirigido probabilisticamente de maneira não tão aleatória, para que os indivíduos com melhor aptidão tenham mais chances de sobreviver e se reproduzir na população em que estão inseridos.

Segundo [Michalewicz 1996] e [Goldberg 1989], dentre os métodos de seleção conhecidos, destacam-se os seguintes:

Seleção por *ranking*: os indivíduos da população são ordenados de acordo com seu grau de adaptação, então recebem uma probabilidade de escolha conforme a posição que ocupam. Uma variação deste mecanismo é simplesmente passar os N melhores

indivíduos para a próxima geração.

Seleção por giro de roleta: neste método o grau de adaptação de cada indivíduo é representado de forma proporcional ao somatório dos graus de adaptação de todos indivíduos da população, isto é, cada indivíduo recebe uma probabilidade de passar para próxima geração proporcional ao seu *fitness*. O método consiste no sorteio de um valor i pertencente ao intervalo $[0; \text{total}]$. Então, é selecionado o indivíduo x pertencente à faixa do somatório onde i está localizado.

Seleção por torneio: dois indivíduos são escolhidos aleatoriamente. Então, um número aleatório i é gerado entre 0 e 1. Se $i < k$ (parâmetro), o melhor entre os dois indivíduos é selecionado, caso contrário, o pior é selecionado.

Seleção bi-classista: são selecionados os melhores indivíduos da população dentro de um percentual determinado. E os $(100 - \text{percentual fixo})$ piores indivíduos.

Seleção *Steady-State*: apenas alguns dos piores indivíduos da população são removidos, os demais são mantidos para reproduzirem-se.

3.2.5 Reprodução

Após o processo de seleção dos indivíduos aptos para reprodução, são selecionados pares de reprodutores de acordo com uma probabilidade pré-estabelecida, denominada probabilidade de cruzamento (p_c), a fim de que algumas partes de seus genes sejam combinados para gerarem filhos. Se a p_c for muito alta, a maior parte da população será substituída, perdendo-se os indivíduos com alta aptidão. No entanto, se a p_c for muito baixa, o processo de melhoria de *fitness* será lenta. Por isso, é recomendável que a probabilidade de cruzamento esteja entre $0,5 < p_c < 0,8$. [Mitchell 1998].

A seleção dos pares de reprodutores pode seguir métodos que incluem: escolha ao acaso, indivíduos com fenótipos diferentes, combinação do mesmo indivíduo consigo mesmo, combinação de indivíduos semelhantes e, um indivíduo com alto grau de aptidão é cruzado com uma subpopulação, então os seus filhos são selecionados como pais.

Após a seleção dos pais, é aplicado o *operador cruzamento* que, basicamente, funciona como uma máscara (vetor) cujo conteúdo é aplicado, a partir de determinado gene do cromossomo ou entre intervalos de genes. Segundo [Goldberg 1989] os operadores de cruzamentos mais tradicionais para genótipos de comprimento fixo são:

Cruzamento de um ponto: dado que dois genótipos tem tamanho l_g , escolhe-se aleatoriamente um ponto p , tal que $0 < p < l_g$, de modo que o material genético além deste ponto é trocado entre os pais. Entre os operadores tradicionais, este apresenta o pior desempenho. A Figura 9 ilustra o funcionamento desse tipo de cruzamento.

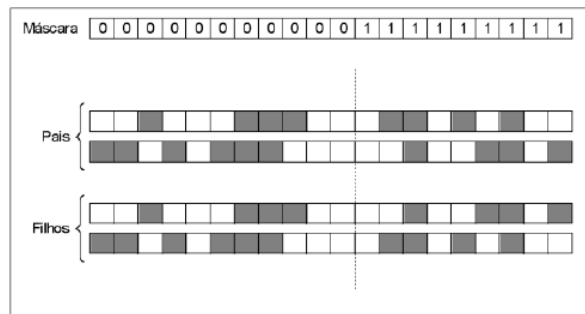


Figura 9: Cruzamento com um ponto de corte
 Fonte: [Lucas 2002]

Cruzamento multiponto: é uma generalização do operador de um ponto. São sorteados n pontos de corte e o material genético entre esses pontos é trocado. O funcionamento desse cruzamento é ilustrado na Figura 10.

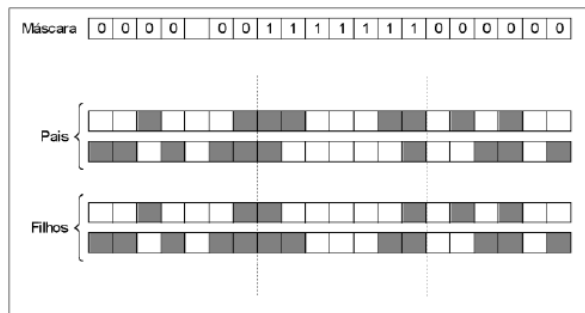


Figura 10: Cruzamento multiponto com dois pontos de corte
 Fonte: [Lucas 2002]

Cruzamento uniforme: para cada gene nos cromossomos, é sorteado qual dos pais deve fornecer o valor a ser preenchido. A Figura 11 ilustra o mecanismo de ação deste tipo de cruzamento nos indivíduos selecionados para reprodução.



Figura 11: Cruzamento uniforme
 Fonte: [Lucas 2002]

3.2.6 Mutação

Realizado o processo de cruzamento, sobre os indivíduos filhos é aplicado o operador mutação de acordo com uma probabilidade pré-determinada, denominada probabilidade de mutação (p_m), com o objetivo de inserir novo material genético na população que está sendo formada, aumentando assim a variabilidade genética e evitando a convergência prematura do AG.

O operador mutação possibilita que o processo de busca chegue a qualquer ponto do espaço de busca, pois evita que o gene selecionado fique estagnado em um valor. No entanto, a p_m não deve ser muito alta para que o processo de busca da solução não se torne completamente aleatório, pois as chances de um indivíduo que representa uma boa solução ser destruído seria muito grande. Assim, comumente defini-se $p_m = 0,1$ ou $p_m = 0,01$ ou $p_m = 0,001$.

A experimentação dos parâmetros do AG dita sua flexibilidade, poder computacional e generalidade. Os operadores genéticos de cruzamento e mutação são responsáveis por todas transformações sofridas pela população, no entanto, eles tem impactos muito distintos. Por exemplo, o operador cruzamento tem como objetivo propagar os esquemas mais adequados na população, assim a determinação do número de pontos de corte afeta diretamente quais esquemas irão sobreviver ao processo de reprodução. Por outro lado, o operador mutação garante a diversidade genética, aumentando a probabilidade de que o espaço de busca será explorado em uma parte significativa de sua extensão. Além disso, também ajuda a evitar os valores de máximo ou mínimo locais, que ocorrem quando a população se estabiliza com uma média de adaptação pouco adequada, devido à pressão evolutiva e baixa diversidade geradas pelo surgimento de um super-indivíduo que domina o processo seletivo e transmite suas características por toda população.

3.3 Comparação com Outras Meta-heurísticas

Os AGs são abordagens com aplicações em problemas com um número elevado de variáveis, não lineares e com natureza discreta. Alguns exemplos são os problemas de pesquisas numéricas, otimização de funções, aprendizagem de máquina e problemas com múltiplos objetivos, dentre outros. Apesar de não garantir o ponto ótimo global utilizado, tende a alcançá-lo desviando-se das soluções ótimas locais. Devido a sua natureza estocástica, o seu desempenho deve ser calculado considerando-se uma média do número de simulações ao invés de única rodada. [Falcone 2004].

As diferenças entre os Algoritmos Genéticos e outras meta-heurísticas incluem o fato de que estas geram uma única solução a cada iteração, enquanto que os AGs geram uma população de soluções. Ademais, nos AGs a busca da solução torna-se independente do problema original, pois uma vez codificada a solução através de um indivíduo, a série de melhorias dessa solução ocorre durante o processo evolutivo.

Segundo [Goldberg 1989], as seguintes características do AG o diferenciam de

outros processos de otimização:

1. Utilizam parâmetros codificados ao invés dos parâmetros em si, isto é, trabalham sobre as representações dos elementos do domínio do problema. No entanto, esta característica do AG impõe ao seu uso a restrição de que o conjunto de soluções viáveis de um problema deve ser codificado em uma população de indivíduos;
2. O processamento é sobre uma população de pontos ao invés de um único ponto;
3. Aplicam operadores probabilísticos e não regras determinísticas. Apesar disso, a busca não é completamente aleatória, visto que o comportamento do sistema pode ser operado com certa previsibilidade estatística por meio de tais operadores.

3.4 Considerações Finais

Os AGs são essencialmente algoritmos de busca direcionados pelo operador seleção, de acordo com o mínimo ou o máximo de uma função. O melhor indivíduo (*fitness*) é encontrado por meio de uma função de avaliação que representa o ambiente no qual a população evolui, segundo a aplicação dos operadores genéticos de cruzamentos e mutação, e o mecanismo de seleção natural. [Mitchell 1998]).

A ideia da infecção viral vem sendo utilizada em algoritmos genéticos para melhorar a taxa de busca do algoritmo superando a convergência prematura por meio do controle da diversidade da população, isso porque os vírus estão associados à fonte de inovação genética devido a alta taxa de replicação e mudança que estas estruturas apresentam. [Romano 2009].

Neste sentido, a meta-heurística denominada AGRI utiliza a infecção viral tanto como um mecanismo de recombinação genética vertical, quanto um direcionador da população de indivíduos para melhores soluções, pois utiliza o material genético dos indivíduos mais adaptados sem criar super-indivíduos. No próximo capítulo, serão apresentados os componentes do AGRI.

CAPÍTULO 4

Algoritmo Genético Retroviral Iterativo

A busca de uma fonte de inovação genética para aumentar a diversidade populacional inspirou a utilização de estrutura viral no AGRI, visto que na natureza a recombinação genética por vírus apresenta alta velocidade de replicação e frequente mutação.

Dentre os diversos tipos de vírus existentes, foi adotada como fonte de inspiração biológica a família *retroviridae*, cujos vírus são baseados em RNA e atendem à necessidade de maior variação alélica do AG, visto que o RNA não possui os mecanismos de correção em seu genoma para eliminar o material viral recombinado.

Com efeito, os vírus artificiais baseados em RNA constituem estruturas com alto índice de mutação que ao infectar os indivíduos alteram seu material genético definitivamente, gerando novos genomas computacionalmente.

No AGRI, a infecção e multiplicação viral seguem o princípio biológico, embora sejam representados como processos distintos, modularizados. Ademais, a representação da natureza é mantida por inicialmente serem criados vírus sem material genético da população de indivíduos do AG. No entanto, após o primeiro ataque viral à população somática, os vírus passam a multiplicar-se derivando uma percentagem dos genes de alguns indivíduos infectados que possuem maior grau de adaptação.

Neste algoritmo, os vírus são tratados como uma população separada da população de indivíduos. A cada infecção, o material genético viral é transmitido verticalmente entre os hospedeiros difundindo trechos de soluções dos vírus com melhor desempenho por toda população.

4.0.1 O Mecanismo Viral do AGRI

O comportamento viral no AGRI é formado por pelas estruturas lógicas denominadas vírus e população viral, além dos mecanismos da infecção viral e gerenciamento do ciclo de vida do vírus. A seguir, a descrição de cada uma dessas partes:

Estrutura Viral : no AGRI, o vírus é uma estrutura lógica com mesmo tamanho do indivíduo, mas que apresenta ausência alélica em alguns genes. É possível estipular o limite máximo de material genético que o vírus terá, no entanto, os locus dos genes que conterão informações variam probabilisticamente. Assim, os vírus formados sempre são distintos quanto a quantidade e dispersão de genoma. A figura 12 apresenta exemplos de vírus que podem ser gerados no AGRI.

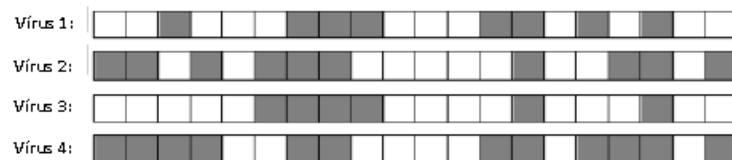


Figura 12: Representação de alguns vírus possíveis para um cromossomo de 20 posições.

População Viral : os vírus criados são organizados dentro de uma população diferente da população cromossômica. Nesta metáfora, a criação dos vírus é modelada seguindo o mecanismo natural, onde num primeiro momento os vírus existem no ambiente, mas ainda não tem o material genético da população a ser infectada. Assim, a primeira população de vírus é gerada de acordo com uma distribuição uniforme, onde cada vírus é criado da seguinte forma:

1. Um indivíduo com material genético completo é gerado;
2. A quantidade final do conteúdo genético viral é definida, aleatoriamente, dentro dos limites mínimo e máximo estabelecidos para o ensaio. Para cada vírus gerado, a taxa de material genético é novamente definida, criando DNAs virais únicos;
3. Parte do genoma do indivíduo criado é eliminado para que o vírus novo tenha somente a quantidade de genes definida para ele.

Após a primeira infecção, a criação dos próximos os vírus é inspirada no processo natural denominado transcrição reversa, típico dos Retrovírus. Assim, a população viral evolui junto com a população somática a cada geração por utilizar material genético dos melhores indivíduos. O seguinte mecanismo é aplicado:

1. Os vírus que foram mais bem sucedidos em infectar os indivíduos são preservados na população viral, de acordo com a taxa de elitismo viral;

2. Todos os vírus fora da taxa de elitismo viral são substituídos, como segue:
 - a.1. A população somática é ordenada pelo *fitness*;
 - a.2. Calcula-se a quantidade de vírus novos que substituirá os vírus fracos, representada pelo percentual de vírus fora da taxa de elitismo em relação ao tamanho da população viral;
 - a.3. Para cada vírus a ser criado, um indivíduo distinto da população somática é selecionado e um novo indivíduo é criado com material genético gerado aleatoriamente, segundo uma distribuição normal;
 - a.4. Alguns genes do indivíduo bem adaptado são selecionados para substituir os genes do novo indivíduo, nas mesmas posições;
 - a.5. O material de alguns genes é eliminado, segundo um percentual definido para esse vírus;
 - a.6. O vírus criado é adicionado à população viral.

A Figura 13 representa o processo de criação de um vírus que evolui junto com a população somática. Nesta técnica, os indivíduos com melhor *fitness* direcionam a evolução da população viral por meio da recombinação genética.

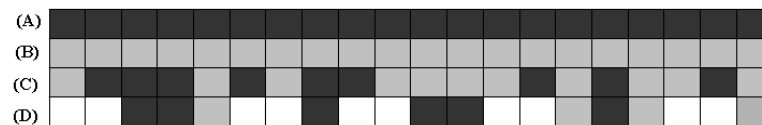


Figura 13: Processo de criação de um novo vírus, a partir da 2ª Geração: (A) Indivíduo gerado aleatoriamente (B) Indivíduo infectado selecionado da população cromossômica (C) Materiais genéticos cromossômicos mesclados (D) Novo vírus com material genético de um hospedeiro bem adaptado

Este mecanismo simula o processo da natureza em que um indivíduo infectado transmite o vírus para um indivíduo sadio. O vírus gerado possui material genético de ambos indivíduos.

Infecção : A infecção ou transcrição é o processo de sobreposição do genoma viral ao genoma do indivíduo da população, resguardando as posições em que os genes estão dispostos no vírus.

No AGRI, o processo que na natureza é denominado "transdução" não ocorre, visto que os vírus não se comportam como vetores de transferência de material genético viral e cromossômico entre dois ou mais hospedeiros infectados pelo mesmo vírus. Logo, uma dada população viral mantém seu genoma íntegro até o fim do seu ataque infeccioso à população somática. Somente após esse processo, os vírus fora da taxa de elitismo viral serão substituídos. Deste modo, o algoritmo mantém a metáfora com o mecanismo natural, pois cada modificação do genoma viral implica em um novo vírus.

O processo de infecção é orientado pelo *fitness* dos indivíduos infectados, ou seja, é feita uma comparação entre o *fitness* do indivíduo antes e depois da infecção. Caso o melhor grau de adaptação seja obtido com a infecção, então a transcrição do genoma viral é efetivada e o vírus recebe um incremento no seu número de infecções. A taxa de infecção é um parâmetro que controla a qualidade dos vírus utilizados numa geração. Caso contrário, a infecção não é efetivada no indivíduo e o vírus tem decrementado uma unidade no seu número de infecções. Este processo evita que indivíduos considerados bons, se tornem ruins, o que auxilia na busca de soluções.

Com o objetivo de maximizar o espaço de busca, cada indivíduo selecionado para o processo de infecção tem a possibilidade de ser infectado por quaisquer dos vírus da população viral. Além disso, em cada geração do AG ocorre três infecções virais sobre a população cromossômica. Assim, um mesmo vírus pode ser selecionado mais de uma vez para infectar a população cromossômica. É possível também que um vírus não seja selecionado para o processo de infecção. Logo, nesta versão do AGRI a similaridade com a natureza é representada de forma mais aproximada, pois existem indivíduos que nunca são infectados por alguns vírus e ao mesmo tempo, existem vírus que infectam mais indivíduos que outros. Computacionalmente, pode-se optar pela infecção única numa geração.

Ciclo de Vida do Vírus : na natureza, os vírus tendem a ser substituídos por outros mais fortes, com mais capacidade de propagação. Da mesma forma, no AGRI o ciclo de vida dos vírus é realizado em três etapas: surgimento, infecção e desaparecimento.

1. Surgimento: a primeira população de vírus é criada aleatoriamente. São gerados indivíduos de acordo com uma distribuição uniforme e então, são removidos alelos de seu genoma segundo o valor normalizado gerado, a partir dos limites mínimo e máximo definidos. Após a primeira infecção, a cada nova geração do AG, uma nova população de vírus é criada respeitando a taxa de elitismo viral. O percentual de novos vírus criados deriva genes de indivíduos da população cromossômica selecionados aleatoriamente. Assim, com o passar das gerações é possível que os novos vírus tenham material genético de outros vírus que infectaram os indivíduos usados na transcriptase reversa.
2. Sobrevivência: os vírus mais bem sucedidos em aumentar o *fitness* dos hospedeiros após a infecção têm maiores chances de permanecer nas gerações seguintes, pois são beneficiados pelo elitismo viral. Esses permanecerão na população viral até perder sua eficiência na infecção. Os vírus que não foram selecionados para infectar nenhum indivíduo na geração, são totalmente substituídos.
3. Desaparecimento: os vírus tendem a desaparecer naturalmente. Devido à evolução da população cromossômica, a capacidade de infecção dos vírus diminui, visto que uma infecção é efetivada apenas quando há um aumento no *fitness* do indivíduo. Assim, com o passar das gerações, os indivíduos tornam-se vacinados, ou seja, o *fitness* não tem nenhuma mudança após as tentativas de infecção. A partir disso, o vírus é denominado código viral inativo.

Para ilustrar o mecanismo viral do AGRI, o fluxograma básico da sequência do processamento é descrito na Figura 14. Um módulo de infecção viral foi adicionado à estrutura do Algoritmo Genético clássico.

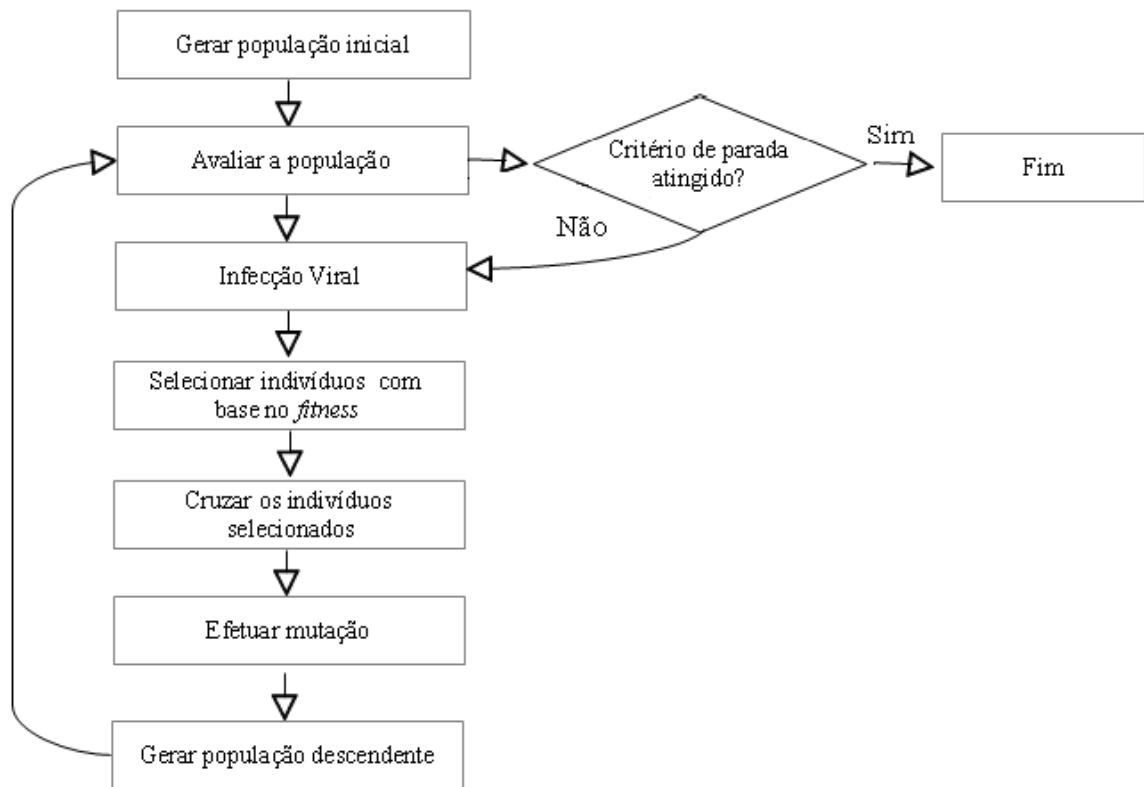


Figura 14: Fluxograma do AGRI.

Em seguida, tem-se a descrição detalhada do algoritmo de infecção viral do AGRI supra ilustrado:

1. Inicializar uma população aleatoriamente;
2. Avaliar indivíduos;
3. Infecção Viral:
 - a.1. Gera uma população viral aleatória, caso seja primeira vez;
 - a.2. Caso contrário, gera novos vírus baseados nos melhores indivíduos da população cromossômica. Os melhores vírus permanecem, o restante a população viral é gerada novamente;
- b. Propaga a infecção nos indivíduos:
 - i. Gera um número entre zero e a taxa máxima para infecção de indivíduos, e calcula o total de indivíduos que serão infectados.

- ii. Caso os indivíduos mais fracos participem do processo de infecção, a população é embaralhada para equilibrar a chance de infecção desses. Caso contrário, os indivíduos são ordenados pelo *fitness*, a fim de que os mais adaptados sejam infectados.
 - iii. Para cada indivíduo selecionado, um vírus é escolhido aleatoriamente para infectá-lo:
 - Para cada indivíduo infectado com sucesso, ou seja, quando houver aumento no seu *fitness*, incrementar 1 no nível de infectibilidade do vírus, caso contrário, decrementar 1 desse parâmetro.
 - Verificar quais foram os vírus com maior taxa de infecção e preservá-los na próxima geração viral, de acordo com a percentagem de elitismo viral.
4. Aplicar operador de Seleção;
 5. Aplicar operadores genéticos de Reprodução e Mutação;
 6. Avaliar os indivíduos e verificar a solução, se solução não for encontrada voltar ao passo 3.

A configuração dos parâmetros do AGRI afetam diretamente o desempenho e o tempo de processamento dos problemas submetidos. Para um melhor entendimento deste algoritmo é necessário entender o funcionamento de seus parâmetros.

4.0.2 Os Parâmetros do AGRI

O AGRI utiliza os bem conhecidos parâmetros do AG clássico para direcionar a busca das soluções, e utiliza os parâmetros da infecção viral inspirada no comportamento dos Retrovírus para aumentar a diversidade populacional. Abaixo a descrição dos parâmetros que controlam o processo viral [Moreira 2010]:

Taxa de Infecção Populacional : Este parâmetro é utilizado para determinar o máximo de indivíduos da população cromossômica que estará apto à infecção, isto é, a quantidade de indivíduos que podem ser infectados. Porém, não é garantido que todos o serão.

Para cada processo infeccioso, é definida a real percentagem de indivíduos que serão infectados utilizando-se uma distribuição normalizada entre zero e o máximo da taxa estipulada. Por exemplo, se essa taxa for quarenta por cento e a população é composta de cem indivíduos, não significa que os 40 indivíduos serão infectados. Supondo que o percentual 30 seja selecionado, então na realidade serão infectados 30 indivíduos. Esta taxa é diretamente proporcional à quantidade de indivíduos da população.

Taxa de Elitismo Viral : O elitismo viral tem o mesmo conceito do elitismo no AG preservando os vírus com maior nível de infectibilidade para a próxima geração, ou seja, os vírus com maiores sucessos de infecções serão passados a diante até o momento em que a população cromossômica criará imunidade a eles. A tendência é que os vírus desapareçam para manter a diversidade populacional de vírus e cromossomos.

Segundo Moreira em [Moreira 2010]: "Esta taxa é inversamente proporcional a variação da população viral, podendo levar o algoritmo a mínimo local, já que todos os indivíduos podem estar vacinados para aquela população viral específica. Quanto menor esse valor, maior variabilidade na população viral, podendo levar a uma variação muito grande dentro da população cromossômica".

Todos os vírus que não estão dentro do percentual da taxa de elitismo são substituídos quando uma nova população viral é criada ao iniciar uma outra geração.

Quantidade de Vírus : Este parâmetro define o número total de vírus que farão parte da população viral. Para cada indivíduo a ser infectado, um vírus é selecionado aleatoriamente. É possível que vários vírus da população viral não sejam selecionados para infecção, isto demonstra exatamente o que acontece na natureza onde existem vários vírus, mas nem todos infectam os indivíduos do meio em que estão. Do mesmo modo, um vírus pode infectar mais de um indivíduo de uma mesma população. Esta melhoria foi adicionada para evoluir a metáfora proposta e não modificar tanto o material genético do indivíduo infectado.

Na proposta de Moreira [Moreira 2010], a quantidade de vírus impactava diretamente no desempenho do algoritmo, visto que "todos os cromossomos sofrerão tentativa de infecção por todos os Vírus a cada geração". Isto transformava totalmente o genoma inicial do indivíduo e levava à perda prematura da diversidade populacional.

Infecção dos Mais Fracos : Este parâmetro altera o fluxo do algoritmo da infecção viral por aumentar as chances de indivíduos fracos serem infectados. É um parâmetro booleano e caso seu valor seja verdadeiro, a população cromossômica é embaralhada de acordo com o *fitness* dos indivíduos, possibilitando que os menos aptos estejam dentro da taxa de indivíduos a serem infectados. Desta forma, os indivíduos mais fracos podem se tornar mais competitivos, com a ajuda da infecção viral. Caso este parâmetro esteja inativo, a população é ordenada pelo *fitness* e os indivíduos mais aptos dentro da taxa de infecção populacional serão infectados.

O valor verdadeiro possibilita o aumento da diversidade genética exploratória. Caso este valor seja falso, o algoritmo segue de acordo com os critérios de seleção dos mais aptos, sendo orientado pelo elitismo do AG e elitismo viral, visto que os indivíduos mais fortes são privilegiados.

Infecção única : Este parâmetro é do tipo booleano e auxilia na manutenção das principais características genéticas dos indivíduos infectados, pois quando seu valor é verdadeiro, cada indivíduo é infectado apenas uma vez em cada geração, o que torna

o processamento mais rápido. Por outro lado, são necessárias mais gerações para alcançar as melhores soluções. Caso este parâmetro esteja inativo, a população cromossômica é submetida três vezes ao processo infeccioso.

Taxa de infecção interna : Este parâmetro apenas indica a percentagem máxima de material genético que os vírus de uma população podem ter. A taxa mínima para o genoma viral é 5%.

Para cada vírus criado calcula-se a quantidade de genes que ele terá, gerando-se um valor aleatório dentro do intervalo mínimo e máximo deste parâmetro. Esta melhoria aumentou a variabilidade genética da população viral, diversificando o tamanho dos trechos de soluções virais.

4.1 Considerações Finais

O AGRI é uma meta-heurística que segue os princípios biológicos em vários aspectos da infecção e multiplicação viral. Por exemplo: cria a primeira população viral sem o material genético da população somática; escolhe aleatoriamente os vírus que irão infectar um indivíduo, possibilitando que parte da população nunca seja infectada por alguns vírus e que certos vírus infectem mais indivíduos; os vírus substituídos são criados com material genético de dois indivíduos, e tem diferentes quantidades de genes.

Nesta técnica, a diversidade genética entre os indivíduos da população é mantida desde o primeiro ataque viral, visto que os vírus inserem novo material genético nos indivíduos infectados. Após esse ataque inicial, os genes dos indivíduos infectados são transmitidos para a próxima geração via recombinação genética vertical, difundindo partes das soluções por toda população.

A diversidade viral também é preservada, pois todos o vírus fora da taxa de elitismo viral são substituídos. Nesse esquema, a população viral evolui junto com a população cromossômica, pois os vírus ineficientes são substituídos por novos, criados a partir do material genético dos indivíduos mais adaptados e de outros genes novos.

A maximização do espaço de busca é realizado utilizando três mecanismos: alta variabilidade genética da população viral com tamanhos diferentes dos trechos de soluções; efetivação da infecção apenas quando há um aumento no *fitness* do indivíduo; e possibilidade de um indivíduo ser infectado por quaisquer dos vírus da população viral.

O capítulo seguinte apresenta os resultados do estudo dos parâmetros da infecção viral, a escolha da melhor configuração para os parâmetros do AGRI e a avaliação de desempenho deste algoritmo em comparação com outras meta-heurísticas considerando os problemas propostos.

CAPÍTULO 5

Resultados

O estudo do AGRI foi realizado a partir de três frentes de trabalho, a saber: análise do comportamento individual e correlacionado dos parâmetros da infecção viral, desempenho do algoritmo nos ensaios para obter as soluções quase-ótimas e comparação das soluções encontradas pelo AGRI com os resultados de outras meta-heurísticas.

A seguir são descritos os métodos aplicados e os resultados obtidos.

5.1 Estudo de Caso

Para os testes do AGRI utilizou-se como problema encontrar o mínimo das funções *benchmarking* unimodais do *CEC2005* em [Suganthan et al. 2005], que tratam do problema de otimização de variáveis reais.

Estas funções são usadas com frequência em problemas de otimização como parte das suites de testes que compõem métodos sistemáticos de avaliação de diferentes heurísticas. A avaliação dos algoritmos pode ser realizada considerando itens como: critérios de parada específicos, complexidade do problema, escalabilidade (demonstrar quanto aumenta o tempo de execução e o número de iterações, conforme o aumento do problema) e etc. [Suganthan et al. 2005].

As fórmulas matemáticas são apresentadas nas equações 5.1, 5.2, 5.3 e 5.4. Seus respectivos gráficos em 3 dimensões (quando essas funções têm 2 dimensões) estão representados, respectivamente, nas Figuras 15 , 16 , 17 e 18.

$$F_1(x) = \sum_{i=1}^D z_i^2 + f_bias_1, Z = X - O, X = [x_1, x_2, \dots, x_D] \quad (5.1)$$

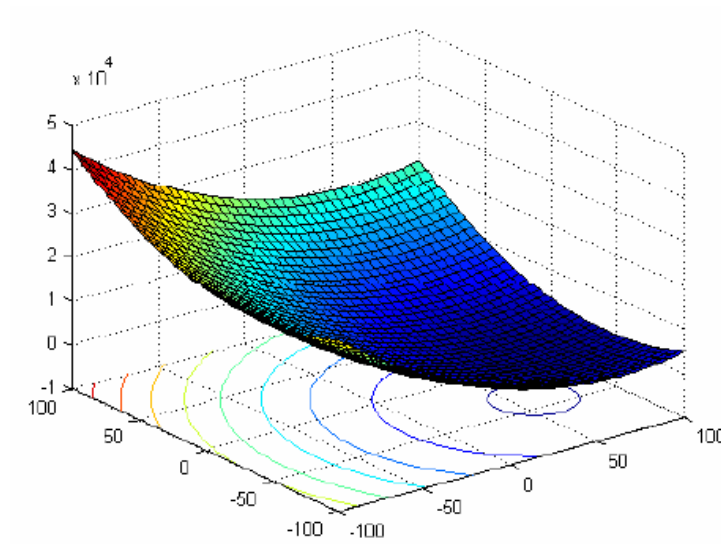


Figura 15: Shifted Sphere Function. Gráfico 3-D para 2 dimensões
Fonte: [Suganthan et al. 2005]

$$F_2(x) = \sum_{i=1}^D \left(\sum_{j=1}^i z_j \right)^2 + f_bias_2, Z = X - O, X = [x_1, x_2, \dots, x_D] \quad (5.2)$$

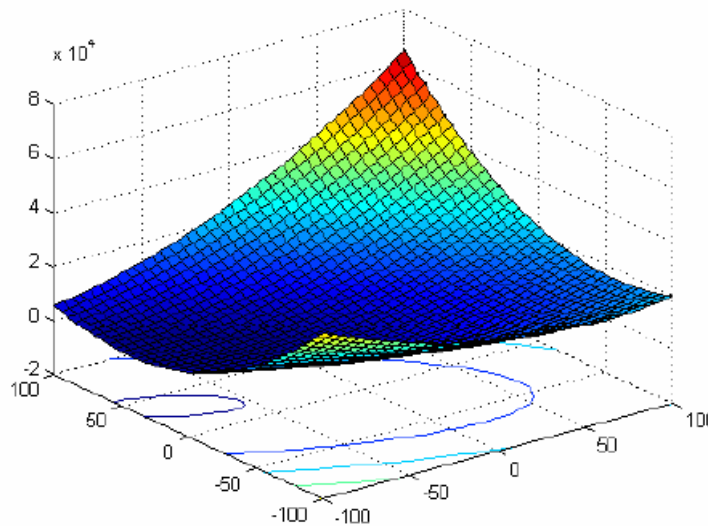


Figura 16: Shifted Schwefel's Problem 1.2. Gráfico 3-D para 2 dimensões.
Fonte: [Suganthan et al. 2005]

$$F_3(x) = \sum_{i=1}^D (10^6)^{\frac{i-1}{D-1}} z_i^2 + f_bias_3, Z = (X - O) * M, X = [x_1, x_2, \dots, x_D] \quad (5.3)$$

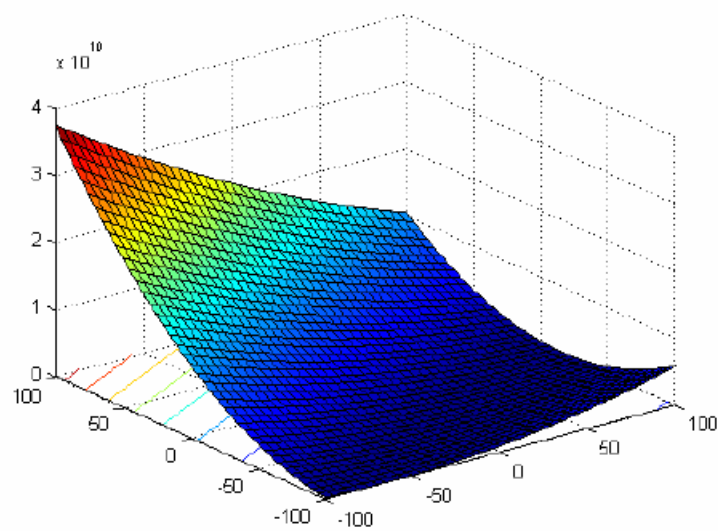


Figura 17: Shifted Rotated High Conditioned Elliptic Function. Gráfico 3-D para 2 dimensões.

Fonte: [Suganthan et al. 2005]

$$F_5(x) = \max \{|A_i x - B_i|\} + f_bias_5, i = 1, \dots, D, x = [x_1, x_2, \dots, x_D] \quad (5.4)$$

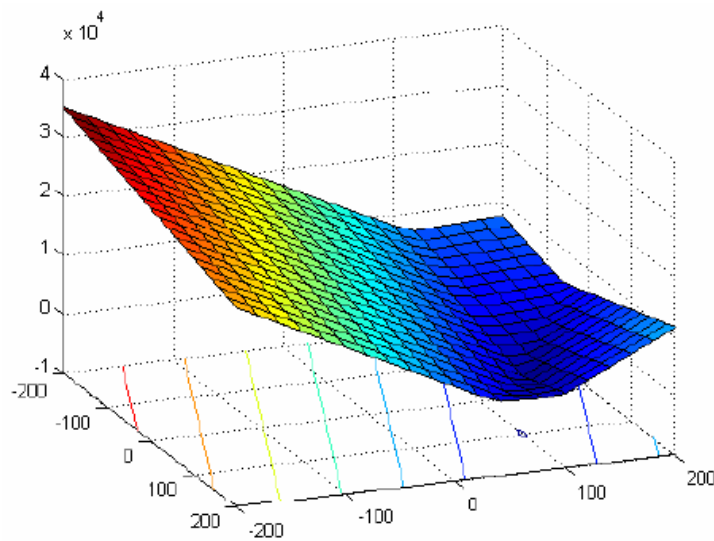


Figura 18: Schwefel's Problem 2.6 with Global Optimum on Bounds. Gráfico 3-D para 2 dimensões

Fonte: [Suganthan et al. 2005]

Para este estudo foram mantidas, como referencia, as características do *benchmark* original. A Tabela 4 exibe as notações definidas para as funções que serão usadas no decorrer dos testes. A Tabela 5 apresenta as *bias* do *fitness* para cada função $f(x^*)$, os valores dos limites mínimo e máximo da busca $[X_{min}, X_{max}]$ e os valores do ótimo global (f_{min}). [Suganthan et al. 2005].

Tabela 4: Notações definidas para os testes das funções do *benchmark*

Função	Nome
$F1$	<i>Shifted Sphere Function</i>
$F2$	<i>Shifted Schwefel's Problem 1.2</i>
$F3$	<i>Shifted Rotated High Conditioned Elliptic Function</i>
$F5$	<i>Schwefel's Problem 2.6 with Global Optimum on Bounds</i>

Tabela 5: Configurações do *benchmark* para os testes das funções

f	$f(x^*)$	Limites da Busca	f_{min}
$F1$	-450	[-100, 100]	0
$F2$	-450	[-100, 100]	0
$F3$	-450	[-100, 100]	0
$F5$	-310	[-100, 100]	0

As sessões a seguir estão divididas de maneira a apresentar os resultados de cada estudo separadamente. A primeira análise apresenta o efeito de cada parâmetro da infecção viral nos resultados do AGRI, além de estabelecer o efeito da interação entre alguns deles no desempenho do algoritmo; em seguida será determinadas algumas configurações que sejam referências de desempenho para as funções unimodais propostas e por último serão apresentados os resultados que comparam o desempenho do AGRI com outras técnicas de otimizações globais.

5.1.1 Estudo dos Parâmetros da Infecção Viral

No estudo sobre os parâmetros do AGRI foram realizados experimentos comparativos com o objetivo de analisar a contribuição de cada parâmetro da infecção viral no desempenho desse algoritmo. Um parâmetro por vez foi alterado, enquanto os demais mantinham-se fixos. Após essa análise inicial, foram realizados experimentos fatoriais considerando a combinação dos parâmetros que apresentaram maior contribuição no desempenho do AGRI, individualmente. Tanto os experimentos simples quanto os experimentos fatoriais tiveram o mesmo número de repetições [Silva 2007]. A Tabela 6 apresenta a configuração dos ensaios realizados no estudo dos parâmetros da infecção viral.

Tabela 6: Configurações do *benchmark* para os testes de avaliação dos parâmetros da infecção viral

Parâmetros	Valores
Número de Dimensões	10
Número de Execuções	30
Tamanho da População Cromossômica	100
Gerações	500

Para análise dos resultados serão utilizados, como valores de referência, a média do *fitness* da população em cada geração e a medida de diversidade genética, a qual consiste na razão entre o *fitness* médio (F_{med}) e o *fitness* máximo (F_{max}) da população, conforme descrito na equação 5.5:

$$Medida\ de\ Diversidade = \frac{F_{med}}{F_{max}} \quad (5.5)$$

A seguir, são apresentados os resultados do comportamento de cada parâmetro da infecção viral no desempenho do algoritmo, assim como as interações entre alguns deles.

5.1.1.1 Quantidade de Vírus e Infecção Populacional

Iniciando o estudo da infecção viral do AGRI, um direcionamento adotado foi verificar a interação entre os parâmetros *Quantidade de Vírus* e *Infecção Populacional*, visto que conceitualmente esses parâmetros têm uma correlação diretamente proporcional.

Os ensaios foram realizados considerando os valores da Tabela 7 para a infecção viral. A *Configuração 3* dessa tabela é utilizada como parâmetro de referencia na comparação do comportamento desses parâmetros, visto que ambos têm valores baixos.

Tabela 7: Configurações para avaliar o comportamento dos Parâmetros da infecção viral *Quantidade de Vírus* e *Infecção Populacional*

Parâmetros da Infecção Viral	Configuração 1	Configuração 2	Configuração 3
Infecção Populacional	50 %	5 %	10 %
Elitismo Viral	5 %	5 %	5 %
Infecção Interna	20 %	20 %	20 %
Quantidade de Vírus	20	100	10
Infecção dos mais fracos	Sim	Sim	Sim
Infecção Única	Não	Não	Não

Os resultados apresentados na Tabela 8 para as demais funções propostas mostram que esses parâmetros têm comportamentos opostos na manutenção da diversidade genética da população.

Tabela 8: Comparação do resultado dos Parâmetros da infecção viral *Quantidade de Vírus e Infecção Populacional*

Algoritmos	F1	F2	F3	F5
<i>Configuração 1</i>	2.006817E-22	8.182400E-24	2.907666E-75	1.039182E-13
<i>Configuração 2</i>	1.702434E-22	7.731373E-24	4.541889E-75	1.324436E-13
<i>Configuração 3</i>	1.911752E-22	8.127911E-24	9.965806E-75	1.099391E-13

A Figura 19 apresenta os resultados para a evolução do *fitness* médio. Nota-se que a *Configuração 1*, a qual expôs a população cromossômica a um ataque viral mais agressivo, apresentou o pior desempenho em comparação com as outras duas, pois induziu o algoritmo a um ótimo local.

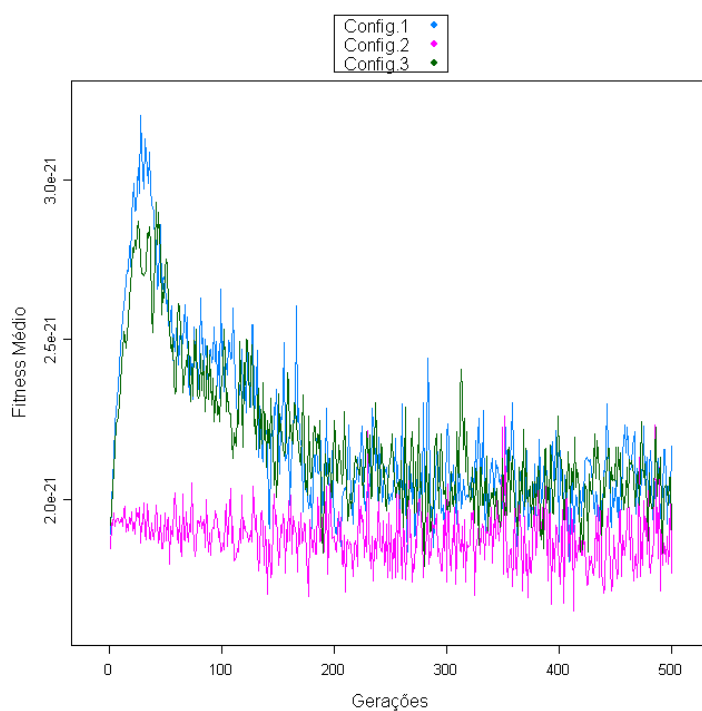


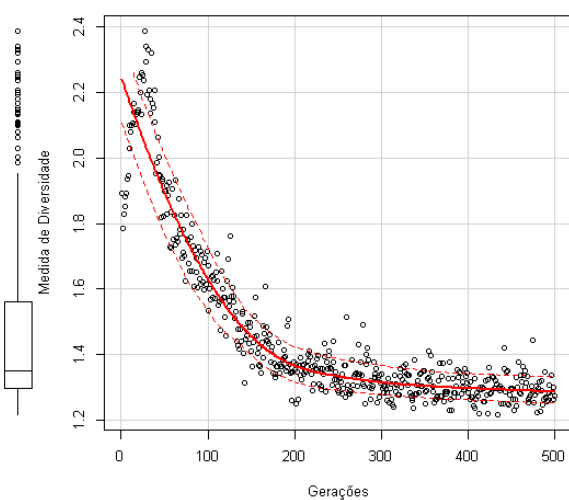
Figura 19: Função F1: Comparação do Fitness Médio ao variar os parâmetros *Quantidade de Vírus e População Infectada*

A inversão de seus valores indicou uma tendência à convergência prematura quando o parâmetro *Infecção Populacional* tem um valor muito alto, apesar de manter a diversidade genética entre os indivíduos.

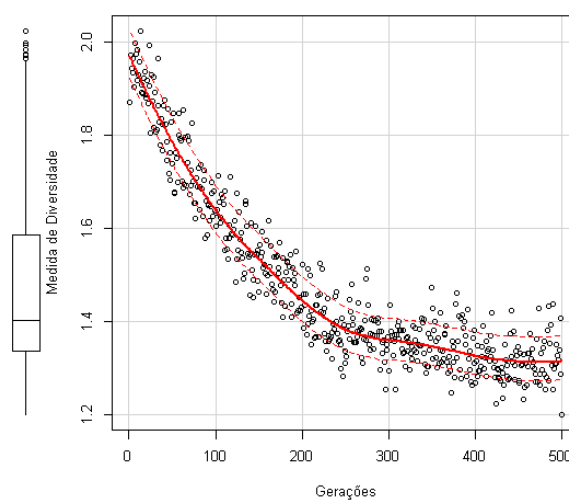
O aumento do parâmetro *Infeção Populacional* permite que mais indivíduos da população cromossômica recebam novo material genético por meio da infecção viral. Mesmo que sejam utilizados poucos vírus, o espaço de busca é aumentado. Assim, pode-se dizer que esse parâmetro representa a "porta" para infecção viral, sendo comparado ao sistema imunológico da população. Quanto maior essa taxa, mais exposta a população está ao ataque dos vírus. Por isso, este parâmetro não deve ter um valor muito alto a fim de não conferir grande estocasticidade a busca.

No AGRI, parâmetro *Quantidade de Vírus* representa diretamente todo caudal material disponível para recombinação gênica. Quando aumentamos o número de vírus da população viral, mesmo que uma pequena parte dos indivíduos seja infectada, a diversidade genética somática entre eles é mantida pela diferente quantidade de material genético que cada vírus possui.

Estudando o comportamento de ambos os parâmetros sobre a medida de diversidade genética, observa-se na comparação dos resultados das *Configurações 2 e 3* que os impactos de uma população viral muito pequena podem ser minimizados com o aumento na quantidade de indivíduos infectados. Apesar dos indivíduos em uma dada geração sofrerem infecções de mesmo material repetidamente, o AGRI consegue aumentar a diversidade genética devido seu mecanismo de transcriptase reversa que fornece genes novos aos melhores vírus. A Figura 20b e a Figura 20c apresentam os resultados da confrontação sob a perspectiva da diversidade genética.



(a) População Viral Medial



(b) População Viral Numerosa

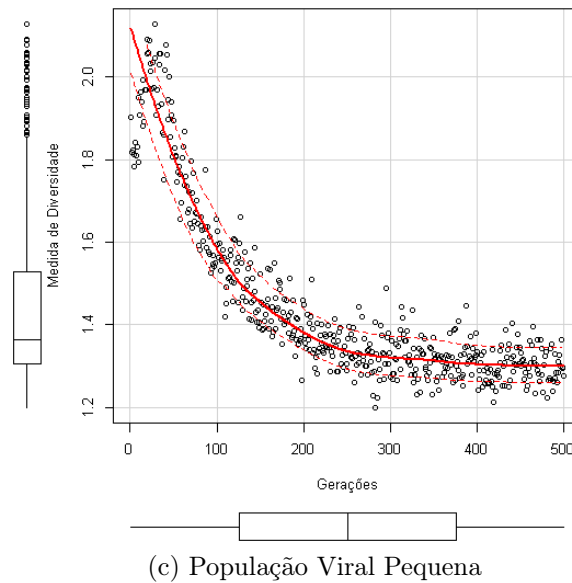


Figura 20: Função F1: Avaliação dos Parâmetros *Quantidade de Vírus* e *Infecção Populacional*- Comparação da Medida Diversidade Genética

Pode-se notar que o efeito gerado quando as taxas desses parâmetros são calibradas de forma antagônica é a perda de diversidade nas gerações iniciais quando a quantidade de indivíduos infectados é muito pequena, conforme as Figuras 20a e 20b. Observa-se também que a melhora da diversidade na *Configuração 1*, conforme a Figura 20a é devido a quantidade de vírus disponíveis não ser muito pequena. Assim, a melhora na manutenção da diversidade genética da população somática é obtida quando essa é infectada por uma população viral de tamanho médio em relação à quantidade de indivíduos.

5.1.1.2 *Infecção dos Mais Fracos*

O parâmetro *Infecção dos Mais Fracos* tem como mecanismo de funcionamento o embaralhamento da população somática, ou seja, os indivíduos com desempenhos bom e ruim são misturados antes de sofrerem a infecção viral.

Para avaliar o efeito do parâmetro *Infecção dos Mais Fracos* no desempenho do AGRI, foram criadas duas configurações virais cujos valores para os demais parâmetros são iguais. A Tabela 9 apresenta as configurações concorrentes com ativação alternada desse parâmetro.

Apesar dos indivíduos menos aptos serem recombinados com bons trechos de soluções virais, a análise mostrou que a ativação do parâmetro *Infecção dos Mais Fracos* leva o algoritmo a uma convergência prematura. Assim, o direcionamento dado pelo AG em relação ao grau de aptidão de cada indivíduo da população deve ser mantido, ou seja, não aumentar as chances dos indivíduos menos aptos no processo evolutivo, visto que

Tabela 9: Configurações para avaliar o comportamento do Parâmetro da infecção viral *Infecção dos Mais Fracos*

Parâmetros da Infecção Viral	Configuração 4	Configuração 5
Infecção Populacional	100 %	100 %
Elitismo Viral	3 %	3 %
Infecção Interna	40 %	40 %
Quantidade de Vírus	80	80
Infecção dos mais fracos	Não	Sim
Infecção Única	Não	Não

possuem baixo potencial de adaptação pós-infecção viral.

Os resultados da Configuração 5 mostram que a média de adaptação da população somática é pior do que os resultados da Configuração 4, conseqüentemente, o desempenho do algoritmo em minimizar o problema proposto foi afetado. A Figura 21 apresenta a comparação desses ensaios.

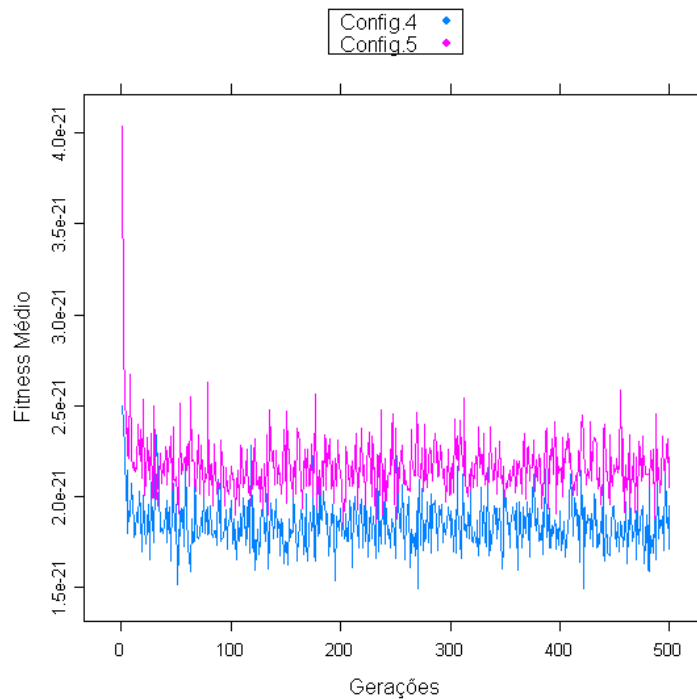


Figura 21: Função F1: Avaliação do Parâmetro *Infecção dos Mais Fracos* - Comparação do Fitness Médio

Segundo os resultados, privilegiar os mais aptos no processo de infecção melhora a precisão do algoritmo sem aumentar a estocasticidade da busca. Isso mostra que apesar da população ser infectada mais de uma vez na mesma geração, o AGRI cria trechos de soluções virais que mantêm a diversidade genética impedindo a convergência prematura. Os resultados do *fitness* médio para as demais funções propostas são apresentados na Tabela 10.

Tabela 10: Resultados para o Parâmetro da infecção viral *Infecção dos Mais Fracos*

Algoritmos	F1	F2	F3	F5
Configuração 4	7,9755E-22	2,3869E-23	1,5042E-74	2.1417E-13
Configuração 5	9,8714E-22	2,8201E-23	2,9728E-74	2,6286E-13

A avaliação do parâmetro *Infecção dos mais fracos* sob a perspectiva da medida de diversidade também mostrou que, infectar os indivíduos menos adaptados diminui a variabilidade genética, visto que o *fitness* das populações se assemelham no decorrer das gerações. Assim, ao passo que o *fitness* dos menos adaptados melhora com a infecção viral, o grau de adaptação dos indivíduos mais aptos evolui pouco devido a diminuição das chances deles participarem do processo infeccioso. As Figuras 22a e 22b apresentam a comparação da medida de diversidade entre os ensaios.

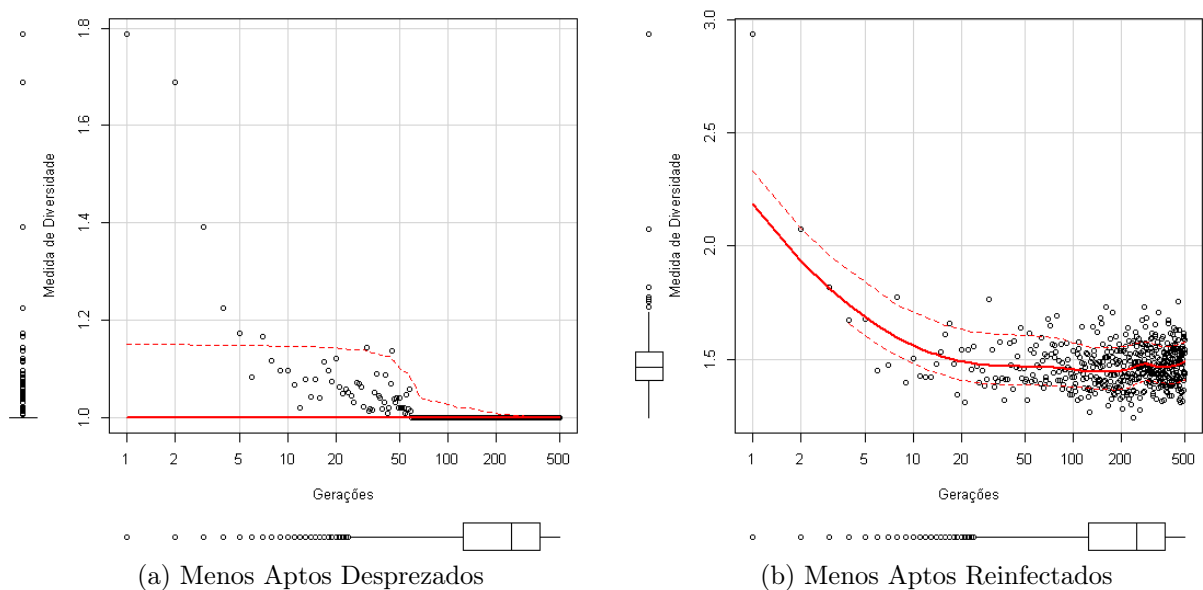


Figura 22: Função F1: Avaliação do Parâmetro *Infecção dos Mais Fracos* - Comparação da Medida de Diversidade Genética

A Figura 22a mostra uma melhora da medida de diversidade da população cromossômica em comparação com a Figura 22b, cujos indivíduos menos aptos foram beneficiados com o aumento das chances de participação no processo de busca de soluções.

Embora os indivíduos menos aptos tenham sido submetidos a várias tentativas de infecção viral, podemos observar que a recombinação genética com tais indivíduos não foi tão benéfica para o aumento da variabilidade genética da população somática no processo de adaptação, sendo indicada a desativação do parâmetro nos problemas propostos.

5.1.1.3 *Infecção Única*

O estudo sobre o parâmetro *Infecção Única* teve como objetivo entender o efeito do número de infecções numa mesma população e no desempenho do algoritmo. Além disso, buscou-se encontrar a correlação entre esse parâmetro e o parâmetro *Infecção dos mais fracos*.

O efeito do parâmetro *Infecção Única* alcança várias gerações, visto que os indivíduos infectados podem ser preservados pelo elitismo do AG. Ademais limita-se à percentagem de indivíduos definida pelo parâmetro *Infecção Populacional*. Logo, seu efeito será quase imperceptível se a percentagem de indivíduos infectados for muito baixa.

A Tabela 11 exhibe as configurações concorrentes que alternam a ativação deste parâmetro. Além disso, tomou-se como referência a Configuração 5 para comparação quando o parâmetro *Infecção Única* está inativo.

Tabela 11: Configurações para avaliar o comportamento do Parâmetro da infecção viral *Infecção Única*

Parâmetros da Infecção Viral	Configuração 6	Configuração 7
Infecção Populacional	100 %	100 %
Elitismo Viral	3 %	3 %
Infecção Interna	40 %	40 %
Quantidade de Vírus	50	50
Infecção dos mais fracos	Não	Sim
Infecção Única	Sim	Sim

Com efeito, pode-se dizer que existe uma forte interação entre os parâmetros *Infecção Única* e *Infecção dos Mais Fracos*. A Figura 23 mostra que a ativação do parâmetro *Infecção dos Mais Fracos* induziu o algoritmo a um mínimo local, visto que os indivíduos menos aptos entraram na faixa dos selecionados para infecção. Apesar dos vírus apresentarem bons trechos de solução, uma única infecção viral não foi suficiente para melhorar as chances desses indivíduos no processo evolutivo e conforme previsto, o algoritmo convergiu prematuramente

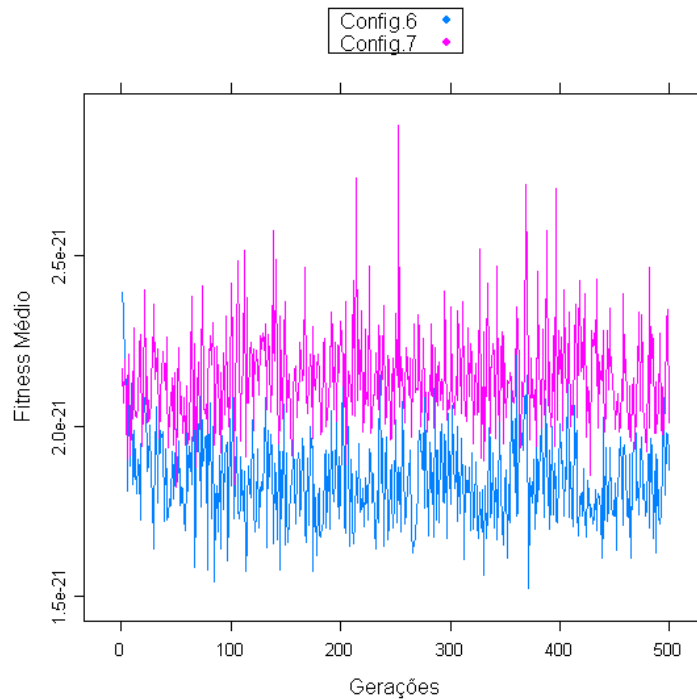


Figura 23: Função F1: Avaliação do Parâmetro *Infeção Única* - Comparação da Evolução do Fitness Médio

A Tabela 12 apresenta os resultados comparativos do *fitness* médio para os ensaios realizados com as demais funções propostas usando as Configurações 6 e 7.

Tabela 12: Resultados para o Parâmetro da infecção viral *Infeção Única*

Algoritmos	F1	F2	F3	F5
<i>Configuração 6</i>	7,4065E-22	2,5909E-23	1,8109E-74	2,3303E-13
<i>Configuração 7</i>	9,2834E-22	3,2890E-23	2,8084E-74	2,6805E-13

Ao analisar a interação entre o parâmetro *Infeção Única* e o parâmetro *Infeção dos Mais Fracos* na diversidade cromossômica, observa-se nas Figuras 24a e 24b que uma única infecção viral nos indivíduos mais aptos foi suficiente para aumentar a variabilidade genética da população somática, ou seja, os vírus criados continham bons trechos de soluções que melhoraram o grau de adaptação dos indivíduos. Logo, a infecção viral impediu que o material genético dos indivíduos se assemelhassem rapidamente.

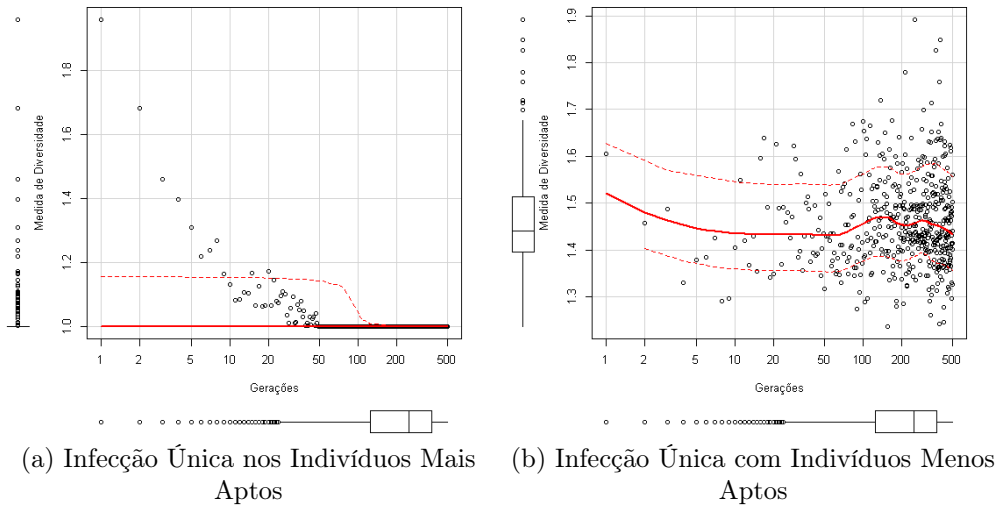


Figura 24: Função F1: Avaliação do Parâmetro *Infecção Única* - Comparação da Medida de Diversidade Genética

Em relação aos indivíduos menos aptos, pode-se observar na comparação das Figuras 22b e 24b que a diversidade populacional apresentou um melhor resultado quando esses indivíduos são submetidos a várias infecções. Para validação dessa afirmativa, utilizou-se como critério o cálculo do *Desvio-Padrão* para os ensaios da *Configuração 5* e da *Configuração 7*. As Figuras 25a e 25b apresentam o resultados dos ensaios.

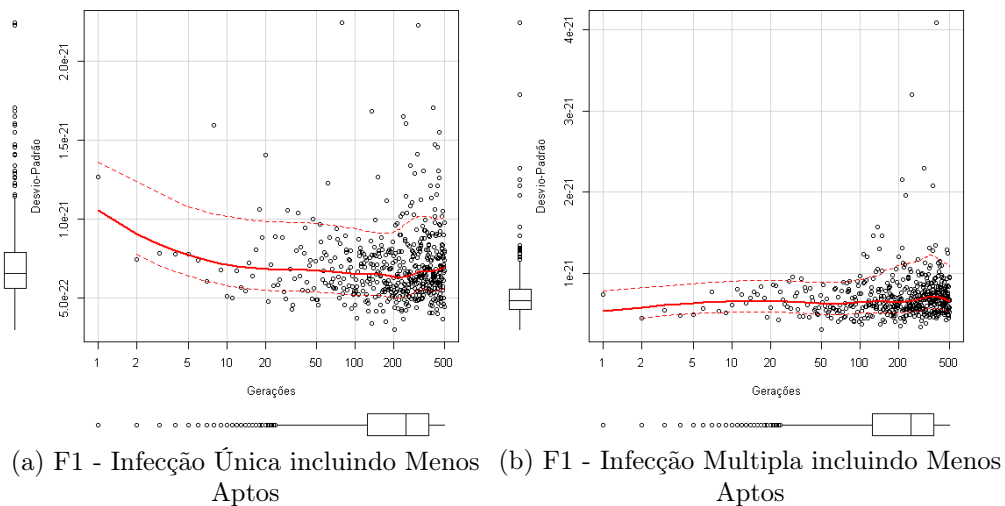


Figura 25: Função F1: Avaliação do Parâmetro *Infecção Única* - Comparação do Desvio-Padrão

Os valores medidos na Figura 25b têm uma distribuição de dispersão menor que os valores da Figura 25a, indicando maior precisão em torno da média do fitness. Apesar da melhora na medida de diversidade, o *fitness* médio não teve uma melhora significativa,

o que corrobora a conclusão de que o parâmetro *Infeção dos Mais Fracos* pode ser desativado nos problemas propostos.

5.1.1.4 *Elitismo Viral*

Sobre o estudo do parâmetro *Elitismo Viral*, observou-se um comportamento similar ao elitismo do AG. Assim, para realização de uma análise concentrada deste parâmetro utilizou-se as configurações apresentadas na Tabela 13, cuja diferença está na variação dessa taxa de preservação do material viral.

Tabela 13: Configurações para avaliar o comportamento do Parâmetro da infecção viral *Elitismo*

Parâmetros da Infecção Viral	Configuração 8	Configuração 9
Infecção Populacional	100 %	100 %
Elitismo Viral	60 %	2 %
Infecção Interna	20 %	20 %
Quantidade de Vírus	100	100
Infecção dos mais fracos	Sim	Sim
Infecção Única	Não	Não

O exame da variação do elitismo viral mostrou que este parâmetro, apesar de preservar os vírus para as próximas gerações, não é determinante nos casos de convergência prematura do AGRI. As Figuras 26a e 26b apresentam os resultados para o *fitness* médio.

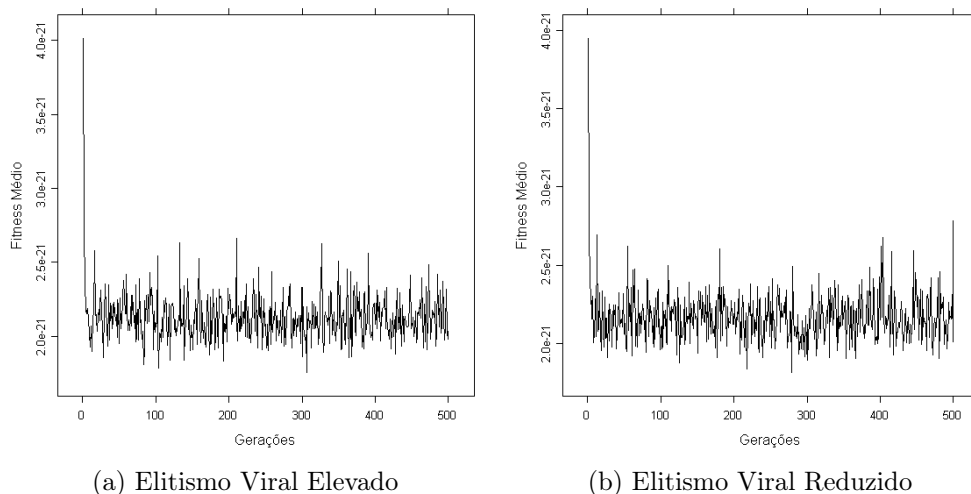


Figura 26: Função F1: Avaliação do Parâmetro *Elitismo Viral* - Comparação da Evolução do Fitness Médio

Assim como na natureza os vírus com maior poder de transmissão afetam a maior parcela de uma população, no AGRI ocorre o mesmo. A preservação de muitos vírus não garante que esses terão sucesso na infecção de indivíduos. A qualidade do material viral continua a ser o requisito para uma infecção bem-sucedida.

A taxa de elitismo reduzida é mais eficiente para o desempenho do AGRI, pois define com melhor precisão as soluções virais que devem ser preservadas. Os resultados da Figura 26b mostram que reduzir muito esta taxa não significa que bons trechos de soluções virais sejam descartados prematuramente.

Conclui-se, portanto, que a evolução do *fitness* é orientada somente pelos vírus altamente contagiosos, ou seja, com maior nível de infectibilidade no AGRI, os quais se mantêm preservados para as gerações virais posteriores. Ademais, o elitismo viral reduzido não tornou os indivíduos "vacinados", ou seja, os vírus não se tornaram inativos devido não melhorarem o *fitness* do indivíduo infectado, o que facilitaria a convergência do algoritmo para um mínimo local.

A Figura 26a mostra que elevar muito a taxa de elitismo viral não aumentou o espaço de busca. Isso significa que os vírus com baixo nível de infectibilidade preservados não conseguiram infectar as gerações seguintes, por isso devem ser descartados na geração viral em que foram criados.

Os resultados comparativos do fitness médio para a *Configuração 8* e a *Configuração 9*, considerando todas as funções propostas, é apresentado na Tabela 14.

Tabela 14: Resultados para o Parâmetro da infecção viral *Elitismo Viral*

Algoritmos	F1	F2	F3	F5
<i>Configuração 8</i>	1.002193E-21	8.124833E-24	1.000521E-74	1.012222E-13
<i>Configuração 9</i>	1.009565E-21	8.631764E-24	1.003938E-74	1.177173E-13

O estudo da diversidade populacional apresentado nas Figuras 27a e 27b mostra que essa medida não apresentou grandes variações no decorrer das gerações, apesar da enorme diferença proposta na calibragem do *Elitismo Viral*. Isto ocorre devido ao mecanismo de sorteio aleatório dos vírus e ao fato dos vírus com menor poder de transmissão se tornarem inativos.

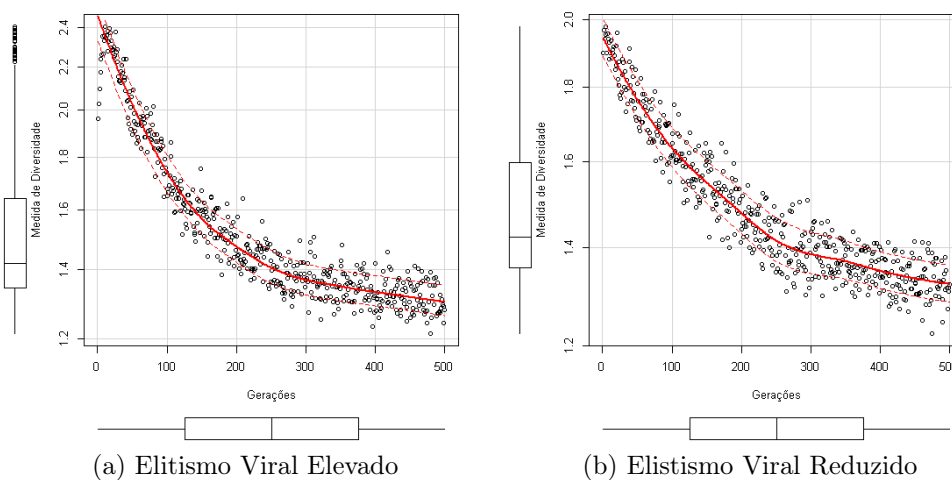


Figura 27: Função F1: Avaliação do Parâmetro *Elitismo Viral* - Confrontação da Medida de Diversidade Genética

Quando essa taxa recebe valores menores, a contribuição dos vírus pode ser orientada pelos demais parâmetros da infecção viral que direcionam o nível do ataque e a quantidade de material genético a ser gerado. Assim, é recomendado que a taxa de *Elitismo Viral* siga o percentual de *Elitismo do AG Clássico*.

5.1.1.5 Infecção Interna

Finalizando o estudo individual dos parâmetros da infecção viral do AGRI, seguiu-se o mesmo método para o exame do parâmetro *Infecção Interna*. Assim, foram aplicadas as configurações definidas na Tabela 15 para investigação dos efeitos desse na evolução do *fitness* médio e na medida de diversidade populacional.

Tabela 15: Configurações para avaliar o comportamento do Parâmetro da infecção viral *Infecção Interna*

Parâmetros da Infecção Viral	Configuração 10	Configuração 11
Infecção Populacional	100 %	100 %
Elitismo Viral	3 %	3 %
Infecção Interna	70 %	10 %
Quantidade de Vírus	100	100
Infecção dos mais fracos	Sim	Sim
Infecção Única	Não	Não

Em relação ao desempenho do *fitness* médio, o resultado dos ensaios expostos nas Figuras 28a e 28b mostram que o aumento da taxa de *Infecção Interna* não melhorou o desempenho do AGRI. Um dos motivos é que a quantidade de material genético de cada vírus é definida aleatoriamente no momento da criação dos mesmos. Como resultado a população viral tem vírus com trechos de soluções grandes e pequenas. Além disso, o mecanismo de tentativa de infecção de um indivíduo envolve o sorteio aleatório do vírus, o que beneficia equitativamente os com muito material genético e os vírus com menos genes.

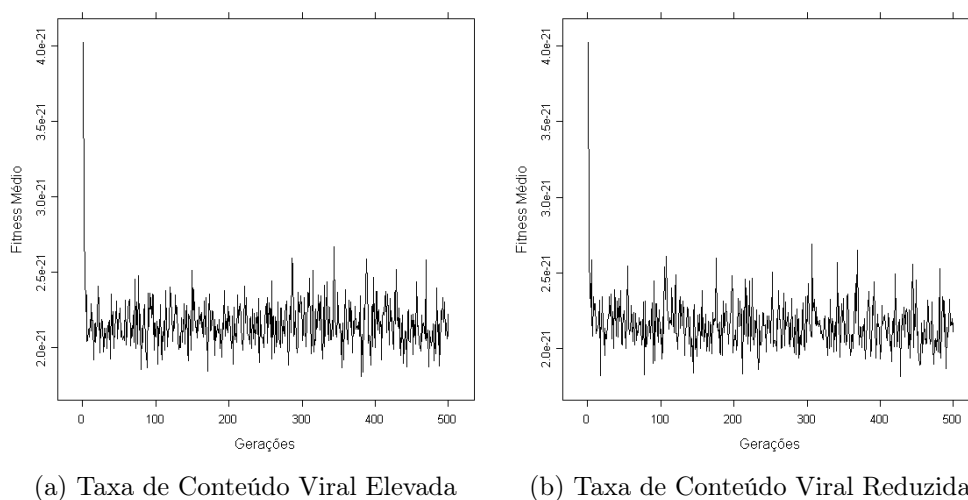


Figura 28: Função F1: Avaliação do Parâmetro *Infecção Interna* - Comparação do Fitness Médio

Apesar disso, os resultados nos levam a conclusão de que vírus com trechos de soluções grandes conduzem os resultados a mínimos locais, visto que o material cromossômico original pode ser tornar bastante reduzido.

Um ponto positivo observado quando a taxa de *Infecção Interna* está muito elevada é que o espaço de busca não aumentou drasticamente, algo que seria esperado devido a quantidade de recombinações gênicas a que foi submetida a população numa mesma geração. A Tabela 16 apresenta os resultados comparativos do fitness médio para as demais funções propostas.

Tabela 16: Comparação do resultado para o Parâmetro da infecção viral *Infecção Interna*

Algoritmos	F1	F2	F3	F5
Configuração 10	1.540927E-23	1.000129E-21	1.000309E-74	1.039182E-13
Configuração 11	0.00e+00	9.966128E-24	9.997376E-74	1.324436E-13

Focalizando a avaliação da diversidade genética cromossômica, as Figuras 29a e 29b atestam que atribuir valores altos para formação da carga genética viral não melhora a diversidade populacional. Devido a população viral conter mais vírus com carga genética

elevada e o fato de parte do material cromossômico ser utilizado para criar o percentual da população viral que será substituído, os indivíduos infectados tendem a ficar parecidos entre si após sucessivas infecções.

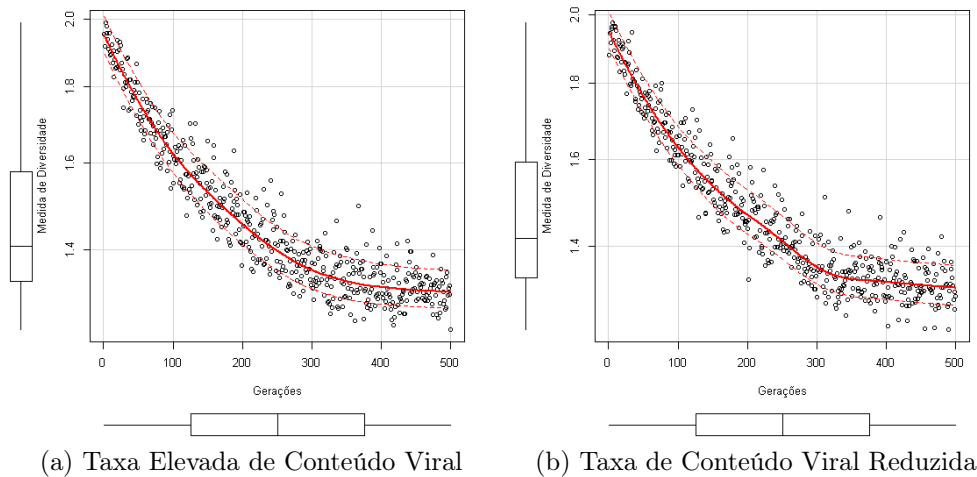


Figura 29: Função F1: Avaliação do Parâmetro *Infecção Interna* - Comparação da Medida de Diversidade Genética

Portanto, configurar o parâmetro *Infecção Interna* com valores discrepantes não aumenta muito a oscilação dessa medida, se compararmos com o efeito isolado dos outros parâmetros da infecção viral supra avaliados.

Assim, temos que o parâmetro *Infecção Interna* é o maior mantenedor da diversidade genética da população viral. Uma observação concluinte é que ao elevar muito a taxa de material genético viral, os melhores resultados para o desempenho do algoritmo foram obtidos com a ativação do parâmetro *Infecção Única*.

5.1.2 Desempenho do AGRI - Referências Iniciais

O estudo da avaliação de desempenho do AGRI apresentado a seguir tem o objetivo de definir as configurações dos parâmetros que geram boas soluções em comparação com os resultados do *benchmark* de referência. Ao final, deseja-se obter configurações que apresentem um resultado de desempenho considerado melhor, mediano e pior para os problemas propostos.

A base inicial para comparação do desempenho foi um AG com uma configuração que apresentou boa performance e um AG com taxa de mutação bastante elevada. Primariamente, a comparação entre estes visa evidenciar que uma infecção viral de qualidade do AGRI não induz a uma busca aleatória e que são gerados bons trechos de soluções, ainda que produzidos com parte do material genético dos indivíduos mais adaptados por um processo de distribuição uniforme.

Assim, o método utilizado neste estudo foi a variação dos parâmetros da infecção viral com base no resultado do comportamento desses, apresentado na sessão anterior. Cada configuração analisada foi testada com três populações iniciais fixas, como controle local e posteriormente populações iniciais variáveis, considerando o princípio da casualização para lograr sua validade [Silva 2007]. A Tabela 17 apresenta outras definições assumidas para os ensaios.

Tabela 17: Parâmetros dos Testes Artificiais

Parâmetro	Valor
Número de Dimensões	10
Tamanho da População Cromossômica	100
Gerações	500
Número de Execuções	30

O resultado do estudo é apresentado utilizando-se como referência a função $F1$, pois os resultados foram os mesmos para as demais funções. Além disso, no decorrer dos testes desta sessão, serão usadas as configurações apresentadas na Tabela 18 para os parâmetros do AG e do AG com taxa de mutação elevada.

Assim, os resultados da evolução do *fitness* médio e da medida de diversidade genética do AG Clássico e do AG com taxa de mutação elevada serão sempre comparados com as configurações virais propostas.

Tabela 18: Configuração dos parâmetros do AG clássico e AG com Taxa de Mutação Elevada

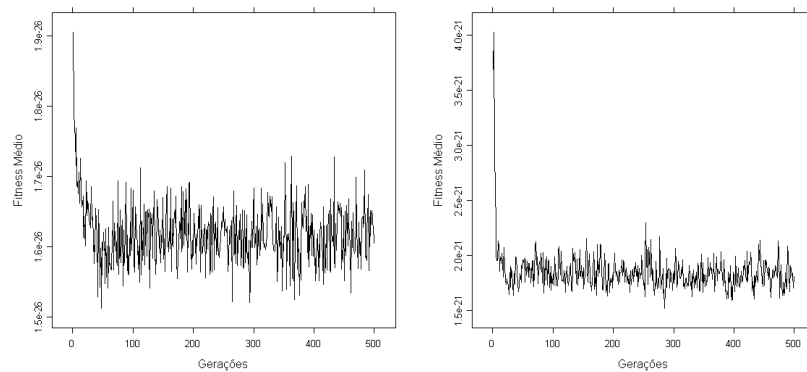
Parâmetros do AG	AGRI e AG Padrão	AG com Mutação Elevada
Taxa de Elitismo	5 %	5 %
Taxa de Mutação	0,7%	40 %
Taxa de Cruzamento	90%	90%
Tipo de Seleção	Torneio	Torneio
Tipo de Cruzamento	Ponto de Corte	Ponto de Corte

Os primeiros ensaios foram realizados com a configuração da Tabela 19 que gerou vírus de melhor qualidade para infecção da população cromossômica.

Tabela 19: Ensaio 1 - Melhor configuração dos parâmetros do AGRI

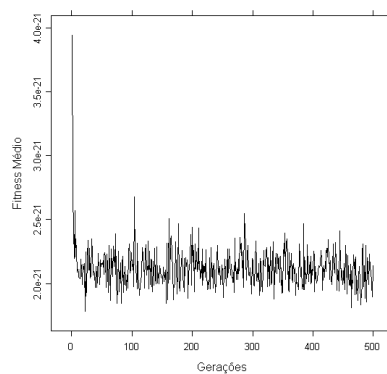
Parâmetros da Infecção Viral	Valores
Taxa de Infecção Populacional	20 %
Taxa de Elitismo Viral	50 %
Taxa de Infecção Interna	20 %
Quantidade de Vírus	100
Infecção dos mais fracos	Não
Infecção Única	Não

Considerando o resultado desses ensaios, a melhor configuração do AGRI teve um desempenho superior ao AG clássico e ao AG com alta taxa de mutação. Os resultados da evolução do *fitness* médio para essas configurações é apresentado na Figura 30.



(a) AGRI - Maior Qualidade Viral

(b) AG Canônico



(c) AG Mutação Elevada

Figura 30: Função F1: População 1 - Comparação da Evolução do Fitness - Infecção viral de qualidade, AG e AG com Taxa de Mutação Elevada

O exame da medida de diversidade genética gerou os resultados da Figura 31. O AGRI mostrou-se muito eficiente em manter a diversidade genética da população cromossômica.

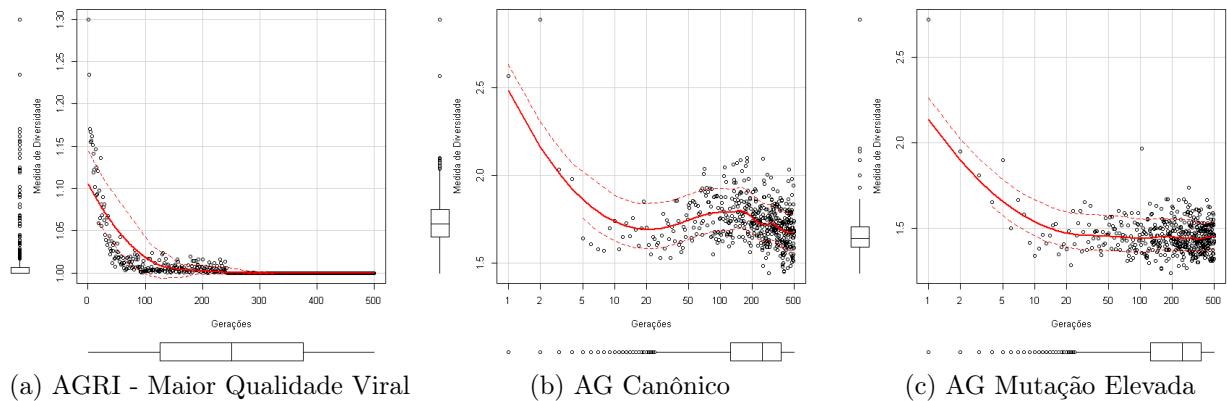


Figura 31: Função F1: População 1 - Comparação da Medida de Diversidade - Infecção viral de qualidade, AG e AG com Taxa de Mutação Elevada

Conforme os resultados do comportamento dos parâmetros da infecção viral, as proporções exibidas na Tabela 20 apresentam uma configuração viral que gera vírus de médio potencial genético.

Tabela 20: Ensaio 2 - Configuração viral com qualidade mediana

Parâmetros da Infecção Viral	Valores
Taxa de Infecção Populacional	100 %
Taxa de Elitismo Viral	3 %
Taxa de Infecção Interna	40 %
Quantidade de Vírus	80
Infecção dos mais fracos	Não
Infecção Única	Não

Populações virais pequenas, mas com muito material genético, quando infectam uma parcela significativa da população cromossômica podem gerar boas soluções, porém tendem induzir o algoritmo a um mínimo local.

As Figuras 32b e 32c se repetem, visto que são as referências para comparação e exibem o melhor *fitness* para os ensaios do AG padrão e do AG com taxa de mutação elevada, respectivamente. Nota-se na Figura 32a que apesar de ter sido configurado com um tipo de infecção viral que não promove seu melhor desempenho, o AGRI atinge uma solução melhor para o problema.

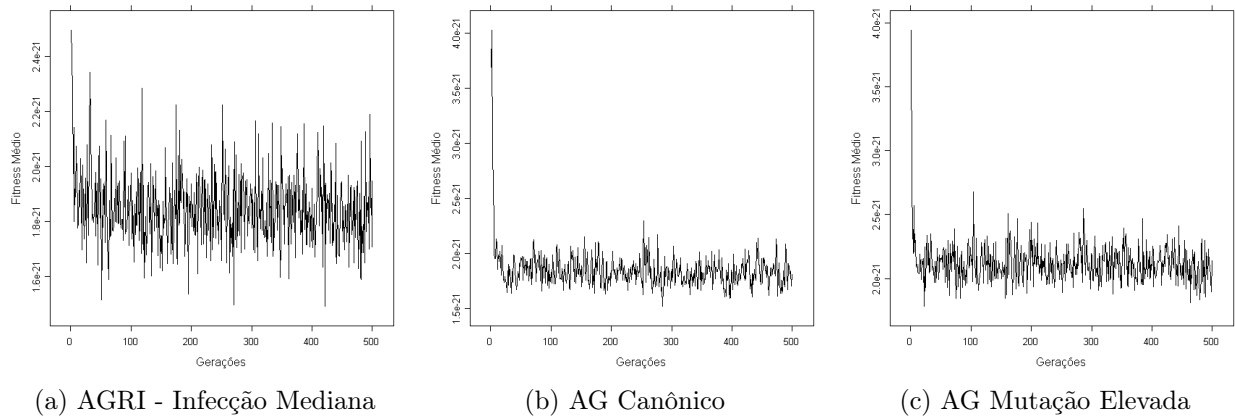


Figura 32: Função F1: População 1 - Comparação da Evolução do Fitness - Infecção viral mediana, AG e AG com Taxa de Mutação Elevada

A Figuras 33b e 33c são novamente apresentadas como medidas de referência para comparação com a diversidade do AGRI configurado com vírus medianos. A Figura 33a apresenta que a diversidade populacional melhorou no decorrer das gerações com a infecção viral, alcançando os melhores resultados.

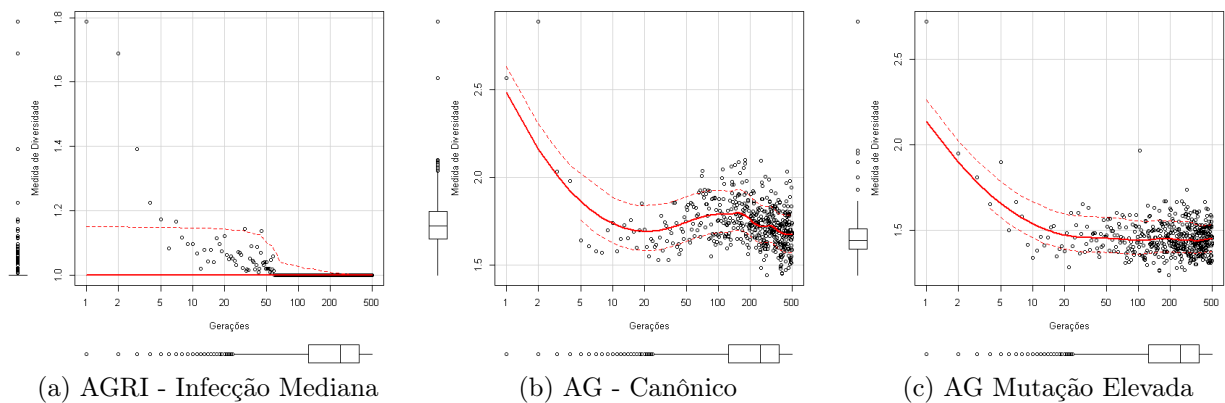


Figura 33: Função F1: População 1 - Comparação da Medida de Diversidade - Infecção viral mediana, AG e AG com Taxa de Mutação Elevada

Os piores resultados do AGRI estão relacionados à criação de populações virais com muito material genético, aumento das chances dos indivíduos classificados como menos aptos pelo AG participarem do processo evolutivo, sobrecarga de infecção, isto é, infectar muitos indivíduos da população cromossômica e população viral pequena. A Tabela 21 apresenta uma configuração viral de baixa qualidade.

Tabela 21: Ensaio 3 - Pior configuração dos parâmetros do AGRI

Parâmetros da Infecção Viral	Valores
Taxa de Infecção Populacional	100 %
Taxa de Elitismo Viral	3 %
Taxa de Infecção Interna	40 %
Quantidade de Vírus	50
Infecção dos mais fracos	Sim
Infecção Única	Sim

Nota-se na Figura 34a que gerar de vírus com muito material genético e submeter a população a uma grande variedade de recombinações pobres, induz a uma convergência prematura, prejudicando a performance do AGRI.

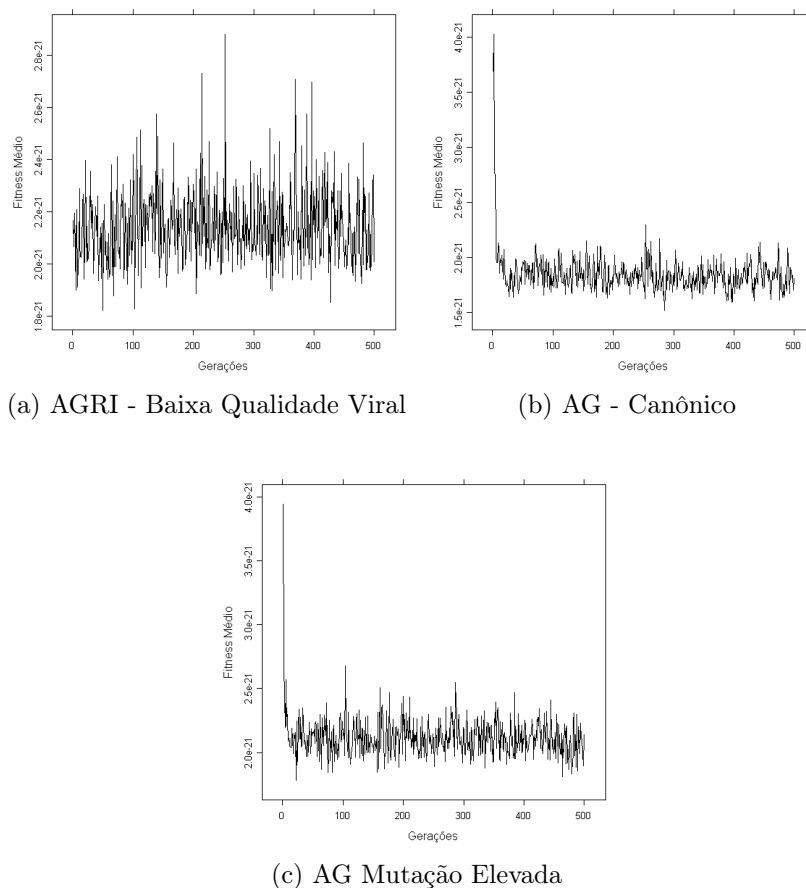


Figura 34: Função F1: População 1 - Comparação da Evolução do Fitness - Infecção viral de baixa qualidade, AG e AG com Taxa de Mutação Elevada

A Figura 35a apresenta os valores da diversidade quando vírus ruins são gerados. Com efeito, a sobrecarga viral fez com que os indivíduos perdessem muito de seu DNA original, induzindo a uma busca aleatória e a uma solução local.

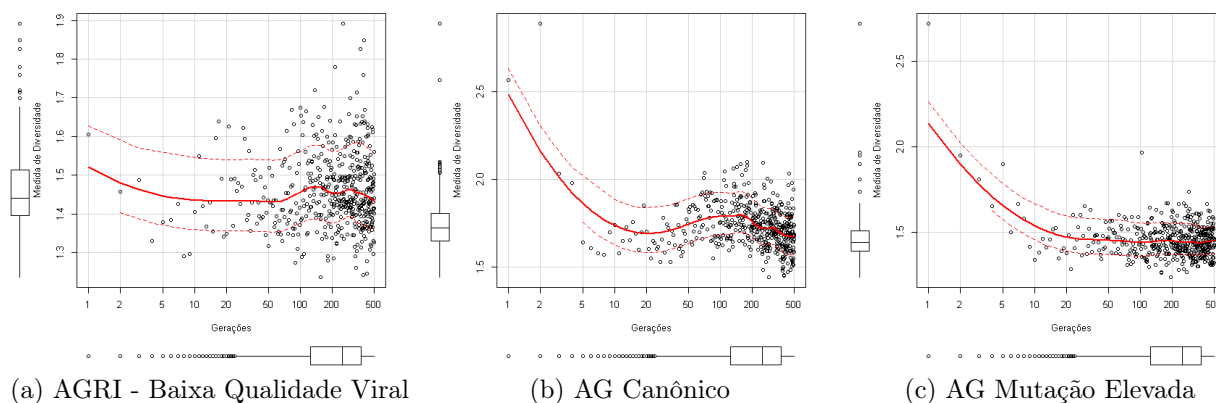


Figura 35: Função F1: População 1 - Comparação da Medida de Diversidade - Infecção viral de baixa qualidade, AG e AG com Taxa de Mutação Elevada

Ao fim, pode-se entender que embora o AGRI tenha um grande potencial em promover a manutenção da diversidade genética cromossômica, seus parâmetros devem ser calibrados de maneira a aumentar a diversidade genética do material viral - na quantidade equilibrada de possíveis soluções - e não somente no tamanho dos trechos de soluções, pois neste caso o algoritmo também seria levado a um mínimo local.

5.1.2.1 O AGRI e outras Técnicas de Otimizações Globais

A evolução dos estudos e técnicas de otimização para problemas complexos segue em ritmo acelerado com muitas propostas de adaptações das meta-heurísticas existentes. Como supra pronunciado, o AGRI mostra um grande potencial competitivo nesse caudal quantidade de algoritmos, assim, faz-se necessário a conferência de desempenho entre tais, seguindo definições de critérios de parada em comum. Neste estudo o método adotado foi a comparação direta com os resultados avaliados e publicados na literatura.

Para tanto, os estudos desenvolvidos por Yang et al. em [Yang et al. 2013], que comparam o algoritmo DE (Differential Evolution) clássico (configuração DE/rand/1/bin); e suas variações jDE (Adaptive Conventional Differential Evolution Algorithm) e PA-jDE (Population Adaptation Differential Evolution Algorithm), foram usados para confrontação com os resultados do AGRI.

Ademais, utilizou-se os estudos de Kumar et al. em [Kumar et al. 2015], que comparam o PSO (Particle Swarm Optimization Algorithm) clássico com outras duas adaptações desta técnica, chamadas de BCPSO (Blended Crossover Based Particle Swarm Optimization Algorithm) e DVPSO (Distribution Vector Particle Swarm Optimization Algorithm), a fim de comparar com desempenho do AGRI para as funções unimodais supracitadas.

Os resultados apresentados na Sessão 5.1.2 manifestam que, o AGRI pode chegar ao ótimo global ou muito próximo disso com poucas gerações e pouco material genético

viral para recombinação. Assim, a Tabela 22 apresenta a configuração adotada para o AGRI nos testes para confrontação com os resultados de outras técnicas.

Tabela 22: Configuração dos Parâmetros do AGRI

Parâmetro	Valor
Taxa de Infecção Populacional	20 %
Taxa de Elitismo Viral	50 %
Taxa de Infecção Interna	20 %
Quantidade de Vírus	100
Infecção dos Mais Fracos	Sim
Infecção Única	Não
Número de Dimensões	30
Tamanho da População Cromossômica	100
Gerações	500

A Tabela 23 apresenta os resultados dos problemas resolvidos com sucesso pelo AGRI e outros algoritmos evolucionários. Todos os resultados exibidos tiveram o mesmo número de ensaios, a saber 30 execuções distintas e 30 Dimensões.

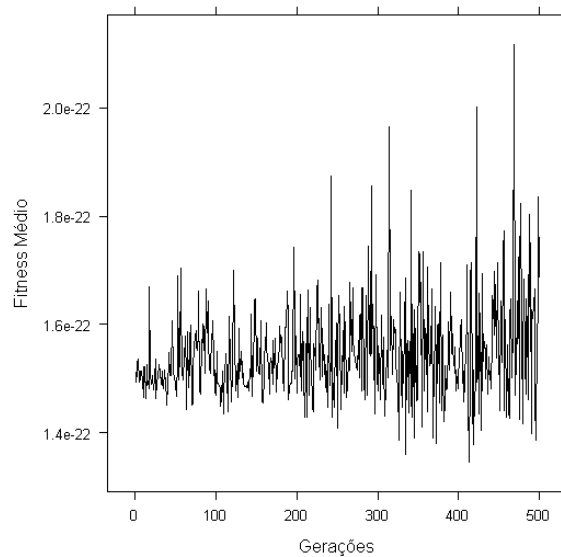
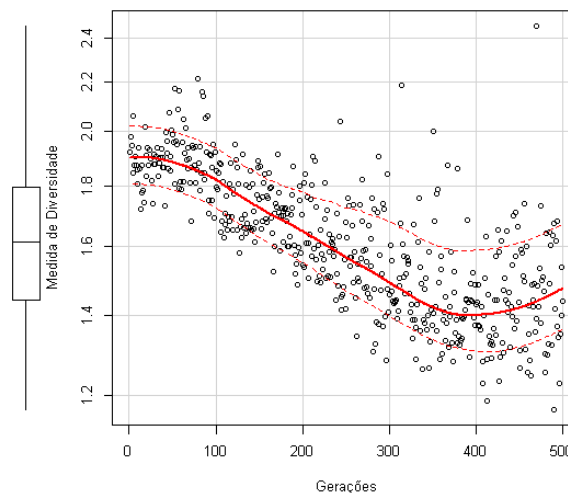
Observa-se que o AGRI teve um nível de performance similar aos demais algoritmos evolucionários avançados nos testes artificiais. Sobre o total de problemas resolvidos, o AGRI alcançou o mínimo global para uma função e obteve os melhores resultados para as outras.

Tabela 23: Comparação do resultado da melhor configuração do AGRI com outros algoritmos

Algoritmos	F1	F2	F3	F5
<i>AGRI</i>	9.565e-23	9.459e-25	0.00e+00	9.216e-14
<i>DE/rand/1/bin</i>	0.00e+00	1.90e-004	0.00e+00	3.84e-001
<i>jDE (NP=100)</i> ¹	0.00e+00	6.80e+000	0.00e+00	6.68e+001
<i>PA-jDE</i>	0.00e+00	3.44E+00	0.00e+00	6.42e-001
<i>PSO</i>	3.952e+003	1.226e+004	1.880e+007	-
<i>DVPSO</i>	2.094e+004	5.266e+004	3.992e+008	-
<i>BCPSO</i>	3.774e+003	1.470e+004	4.114e+006	-

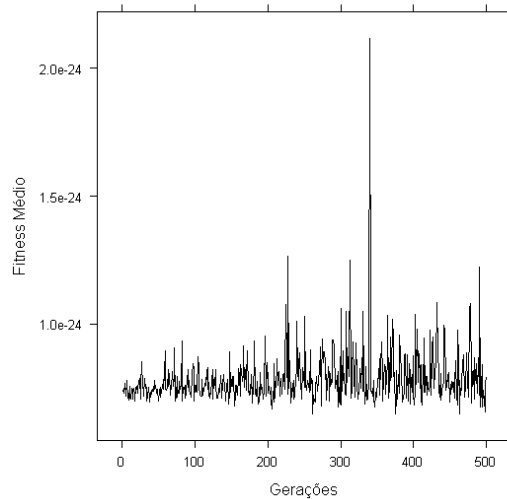
¹Onde NP = Tamanho da População

A Figura 36a apresenta o resultado da evolução do *fitness* médio da função $F1$ quando o AGRI é configurado com $30D$. Observa-se um bom desempenho do algoritmo que conseguiu valores menores que as demais heurísticas. A Figura 36b apresenta a medida de diversidade da população para a função $F1$ quando o AGRI tem indivíduos com 30 cromossomos. Podemos notar que no decorrer do ensaio a medida de diversidade apresentou melhoras com valores mais próximos de *zero*, o que indica alta variabilidade genética.

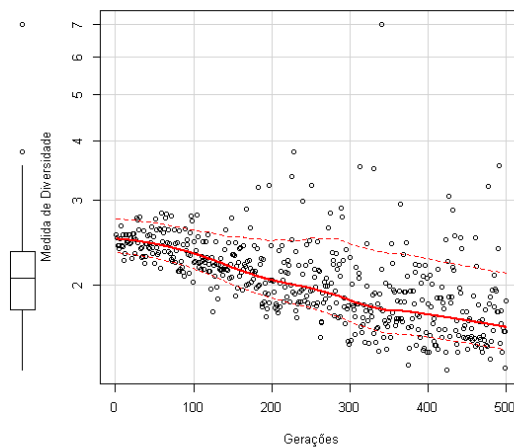
(a) Função $F1$ - $30D$ - Média do Fitness(b) Função $F1$ - $30D$ - Diversidade GenéticaFigura 36: Função $F1$: Evolução do Fitness Médio do AGRI para 30 dimensões

Os resultados da função $F2$ também atestam que o AGRI conseguiu alcançar

uma melhor solução para o problema do que seus concorrentes. A Figura 37a apresenta o resultado da evolução do *fitness* médio no decorrer de 30 ensaios. A Figura 37b apresenta a medida de diversidade da população. Os resultados mostram que durante todo o ensaio a diversidade genética de toda população cromossômica se manteve em excelentes níveis.



(a) Função $F2$ - $30D$ - Média do Fitness



(b) Função $F2$ - $30D$ - Diversidade Genética

Figura 37: Função $F2$: Evolução do Fitness Médio e Medida de Diversidade Genética do AGRI para 30 dimensões

Os resultados da função $F3$ mostram que o AGRI conseguiu solucionar esse problema. A Figura 38a apresenta o resultado da evolução do *fitness* médio no decorrer de 30 ensaios. A Figura 38b apresenta a medida de diversidade da população. Neste ensaio a diversidade genética da população também se manteve alta durante todo o ensaio indicando que a infecção viral tinha ótimos trechos de soluções.

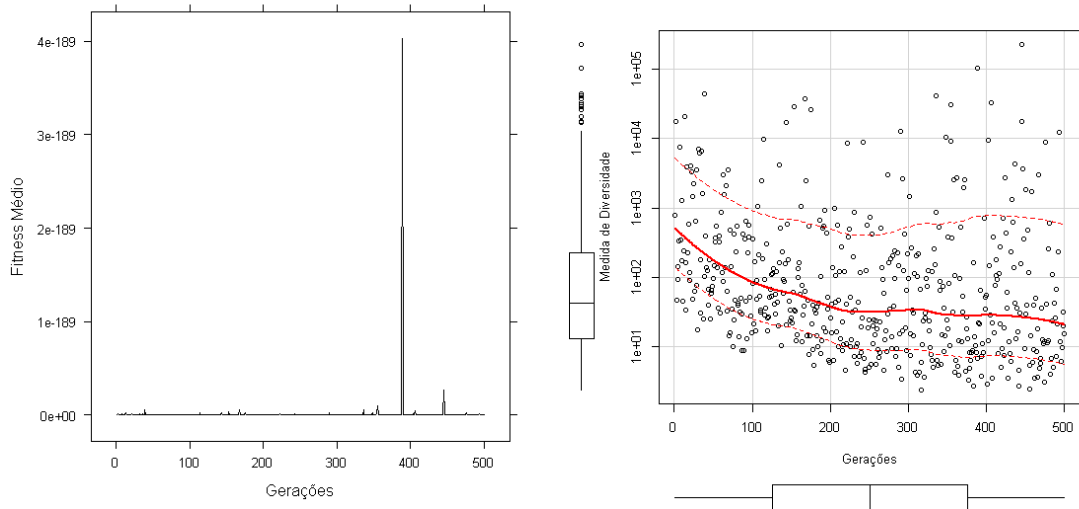
(a) Função $F3$ - $30D$ - Média do Fitness(b) Função $F3$ - $30D$ - Diversidade Genética

Figura 38: Função $F3$: Evolução do Fitness Médio e Medida de Diversidade Genética do AGRI para 30 dimensões

Os resultados da função $F5$ mostram que apesar do AGRI não ter solucionado o problema, os valores alcançados superaram os resultados das demais meta-heurísticas supracitadas. A Figura 39a apresenta que no resultado da evolução do *fitness* médio houve uma convergência prematura do AGRI. A Figura 39b mostra que a diversidade genética da população foi mantida nos testes.

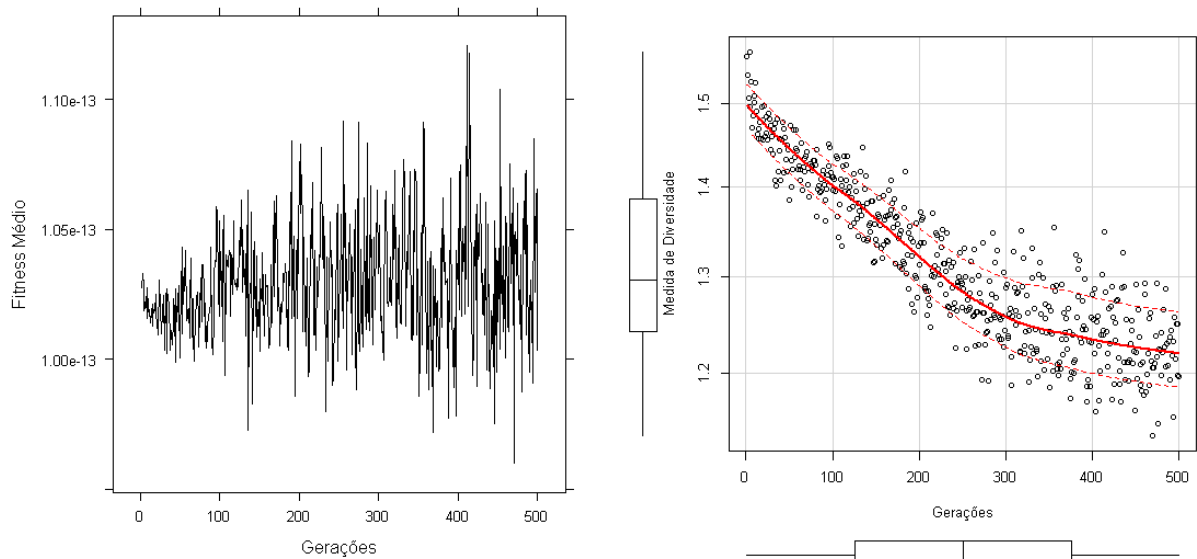
(a) Função $F5$ - $30D$ - Média do Fitness(b) Função $F5$ - $30D$ - Diversidade Genética

Figura 39: Função $F5$: Evolução do Fitness Médio e Medida de Diversidade Genética do AGRI para 30 dimensões

5.2 Conclusões do Capítulo

Gerar trechos de soluções com um tamanho que não comprometa a qualidade do material genético cromossômico, mantenha a diversidade e não induza à convergência prematura, foi o desafio impulsionador para o entendimento da melhor calibragem dos parâmetros da infecção viral do AGRI.

Este estudo mostrou que os parâmetros da infecção viral estão intrincados de maneira cooperativa e até mesmo compensatória para os casos que a infecção viral é muito agressiva. Um exemplo disso é o parâmetro *Infecção Única*, cujo comportamento permite uma calibragem da infecção com bastante material genético.

A análise dos parâmetros da infecção viral do AGRI também mostrou que existe uma compensação entre o tamanho da população viral e o número de indivíduos que serão infectados. Isso quer dizer que, o equilíbrio da diversidade genética é o ponto central do mecanismo de infecção. Esta afirmativa é corroborada pelo estudo do parâmetro *Infecção Interna*, o qual permite a criação de vírus com trechos de soluções com variadas quantidades de material genético, visando a acurácia do AGRI.

Este estudo também mostrou que a infecção viral provê um mecanismo inteligente de direcionamento da qualidade do material genético viral, pois ajuda a melhorar o *fitness* dos indivíduos, ao mesmo tempo em que nivela as oportunidades de infecção por vírus novos. Este artifício comprova que mesmo a grande quantidade de recombinações assumidas não torna a busca aleatória como no caso do AG, cuja taxa de mutação recebe um valor muito alto.

Ademais, a performance do AGRI frente a outras meta-heurísticas mostrou que para funções unimodais, como as usadas neste estudo, ele possui qualificações tão boas quanto, podendo alcançar o ótimo global ou boas soluções com poucas iterações.

CAPÍTULO 6

Conclusões

O AGRI é um algoritmo com grande potencial de contribuição à manutenção da diversidade genética da população cromossômica. Em conjunto, seus parâmetros atuam como mecanismos inteligentes em três frentes para controlar a entrada de material genético novo, a fim de sempre oferecer uma quantidade suficiente para manutenção da diversidade. Quando bem controlados, a busca não se torna exaustiva e pode chegar ao ótimo global.

Seus parâmetros *População Infectada*, *Infecção dos Mais Fracos* e *Infecção Única* gerenciam o efeito da infecção sobre a população somática ao determinar a quantidade de indivíduos que serão infectados, se os indivíduos menos aptos terão nova oportunidade dentro do processo de busca, e quantas vezes a população será infectada numa mesma geração.

Os parâmetros *Quantidade de Vírus*, *Infecção Interna* e *Elitismo Viral* tratam da qualidade da população viral que será produzida e mantida no processo infeccioso.

Diante da perda prematura de diversidade do AG canônico, este tipo de mecanismo oferece as chances de recombinações com qualidade suficiente para solucionar problemas com eficácia.

Embora este estudo seja um *overview* do comportamento global do AGRI, limitou-se ao exame de funções unimodais restando a apreciação de funções multimodais e de composição híbrida para o reconhecimento amplo do AGRI no espaço das meta-heurísticas avançadas. Ademais, um estudo interessante seria uma versão do AGRI para representações reais.

Outros trabalhos a serem realizados envolvem a simulação de mecanismos naturais restritos aos retrovírus e relacionados à reprodução dos mesmos, a saber: progênie viral resultante de ressortimento ou de recombinação por escolha-de-cópia de partes das moléculas parentais, além da mutação viral de acordo com a taxa de erro de replicação.

Além desses, pode-se simular o processo de transdução como modelo de transferência genética horizontal entre os indivíduos da população somática.

Outra frente de trabalho envolve a definição de critérios de parada para o AGRI, a fim de melhorar a performance quando são utilizadas muitas dimensões em um problema.

Referências

- [Araujo and Moreira 2009] Araujo, F. and Moreira, M. (2009). Algoritmos evolutivos com infecção viral para o problema do caixeiro viajante com coleta de premios. Disponível em: <http://www.icmc.usp.br/pessoas/alysson/files/disc/otmGrafos/R09/araujomoreiraR09.pdf>. Acessado em: 10-11-2014.
- [Castro 2012] Castro, L. (2012). Fundamentos de computação natural. Disponível em: <http://www.slideshare.net/Indecastro>. Acessado em: 29-12-2014.
- [Conffin et al. 1997] Conffin, J., Hughes, S., and editors (1997). *Retroviruses*. Plainview, NY: Cold Spring Harbor Laboratory Press.
- [Dawkins 1996] Dawkins, R. (1996). *A escalada do Monte Improvável: uma defesa da teoria da evolução*. Companhia das Letras.
- [Falcone 2004] Falcone, M. A. G. (2004). Estudo comparativo entre algoritmos genéticos e evolução diferencial para otimização de um modelo de cadeia de suprimento simplificada.
- [Geyer-Shultz 1997] Geyer-Shultz, A. (1997). *Fuzzy rule-based expert systems and genetic machine learning*. Physica-Verlag.
- [Goldbarg et al. 2006] Goldbarg, M. C., de Jesus, R. M. C. S., and Goldbarg, E. F. G. (2006). Algoritmo viral no planejamento de braquiterapia de alta dose. *VI Workshop de Informatica Medica - WIM2006*.
- [Goldberg 1989] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley.
- [Gonçalves 2014] Gonçalves, S. (2014). Transcrição. Disponível em: www.infoescola.com/genetica/transcricao. Acessado em: 07-11-2014.
- [Guedes et al. 2005] Guedes, A., J., L., and Aloise, D. (2005). Um algoritmo genético com infecção viral para o problema do caixeiro viajante.

- [I.K. et al. 2003] I.K., J., I.B., R., G.K., G., and KOONIN, E. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.*, 19:68–72.
- [J. et al. 2002] J., P., A., P., and V., P. (2002). *HERVd: database of human endogenous retroviruses*. *Nucleic Acids Res* 30.
- [Kubota and Fukuda 1998] Kubota, N. and Fukuda, T. (1998). Schema representation in virus-evolutionary genetic algorithm forknapsack problem. In *Proceedings of IEEE Internatinal Conference on Evolutionary Computation*, pages 834–839.
- [Kubota et al. 1996] Kubota, N., Shimojima, K., and Fukuda, T. (1996). The role of virus infection in virus-evolutionary genetic algorithm. In *Proceedings of IEEE Internatinal Conference on Evolutionary Computation*, pages 182–187.
- [Kulski et al. 2008] Kulski, J., Shigenari, A., and Ota, M. e. a. (2008). Human endogenous retrovirus(hervk9) strutural polymorphism with haplotypic hla-a allelic associations. *Genetics*, 182(1):445–57.
- [Kumar et al. 2015] Kumar, S., Banerjee, S., and Jana, N. D. (2015). Particle swarm optimization using blended crossover operator. Disponível em: <http://www.warse.org/pdfs/2015/icace2015sp02.pdf>. Acessado em: 17-06-2015.
- [Lemos and Basílio 2000] Lemos, M. and Basílio, A. (2000). Introdução à biologia celular e molecular. Disponível em: www.inf.puc-rio.br/melissa/publicacao/download/biologia/anexo_biologia.pdf. Acessado em: 01-02-2014.
- [Lucas 2002] Lucas, D. (2002). Algoritmos genéticos: uma introdução. Disponível em: www.inf.ufrgs.br/alvares/INF01048IA/ApostilaAlgoritmosGeneticos.pdf. Acessado em: 07-11-2014.
- [Michalewicz 1996] Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 3 edition.
- [Mitchell 1998] Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. MIT Press.
- [Moreira 2010] Moreira, R. S. (2010). Algoritmo genético retroviral iterativo.
- [Pereira 2007] Pereira, F. (2007). Microbiologia. Disponível em: www.cc04-10.med.up.pt.
- [Roehe 2015] Roehe, P. M. (2015). Curso de virologia básica. Disponível em: <http://www.ufrgs.br/labvir/material/poligrafo1.pdf>. Acessado em: 18-10-2015.
- [Romano 2009] Romano, C. M. (2009). Caracterização e dinâmica evolutiva de retrovírus endógenos da família k (erv-k) em genomas primatas. Disponível em: www.teses.usp.br. Acessado em: 10-11-14.
- [S. et al. 2001] S., L. E., L.M., L., B., B., C., N., M.C., Z., and et al., B. J. (2001). *Initial Sequencing and Analysis of the Human Genome*. Nature, 1 edition.
- [Silva 2007] Silva, J. (2007). Estatística experimental: Planejamento de experimentos.

- [Soares 1997] Soares, G. (1997). Algoritmos genéticos: Estudo, novas técnicas e aplicações. Disponível em: www.ppgee.ufmg.br/defesas/673M.PDF. Acessado em: 02-04-2014.
- [Suganthan et al. 2005] Suganthan, P. N., Hansen, N., Liang, J. J., Deb, K., Chen, Y. P., Auger, A., and Tiwari, S. (2005). Problem definitions and evaluation criteria for the cec 2005 special session on real-parameter optimization.
- [Tierney 1988] Tierney, J. (1988). The search for adam and eve. Disponível em: [http://www.virginia.edu/woodson/courses/aas102%20\(spring%2001\)/articles/tierney.html](http://www.virginia.edu/woodson/courses/aas102%20(spring%2001)/articles/tierney.html). Acessado em: 20-07-2015.
- [Tower 1987] Tower, W. (1987). *Como se iniciou a vida na Terra?* Watch Tower Bible and Tract Society of New York, INC., 1987 edition.
- [Tower 1999] Tower, W. (1999). *Uma olhada no microscópio*. Watch Tower Bible and Tract Society of New York, INC., 1999 edition.
- [Tower 2010] Tower, W. (2010). *A Origem da Vida: Cinco Perguntas que Merecem Resposta*. Watch Tower Bible and Tract Society of New York, INC., 2010 edition.
- [Tower 2012] Tower, W. (2012). *Qual é o Objetivo da Vida? Como Encontrá-lo?* Watch Tower Bible and Tract Society of New York, INC., 2012 edition.
- [Tower 2015] Tower, W. (2015). *As cápsulas - bibliotecas vivas*. Watch Tower Bible and Tract Society of New York, INC., agosto de 2015 edition.
- [W-H and D. 1991] W-H, L. and D., G. (1991). *Fundamentals of Molecular Evolution*. Sunderland: Sinauer Associates, 2 edition.
- [Whitelaw and Dik 2001] Whitelaw, E. and Dik, M. (2001). Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat Gen.*, 27:361–65.
- [Wicher et al. 2007] Wicher, T., Sabot, F., Huan-Van, A., J.L., B., Capy, P., and Chalhoub, B. (2007). *A unified classification system for eukaryotic transposable elements*. *Nat Rev Genet*.
- [Yang et al. 2013] Yang, M., Cai, Z., Li, C., and Guan, J. (2013). An improved adaptive differential evolution algorithm with population adaptation. Disponível em: <https://www.lri.fr/~hansen/proceedings/2013/GECCO/proceedings/p145.pdf>. Acessado em: 15-06-2015.