

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Alúcio Cardoso Silva

Um Algoritmo Evolutivo Baseado em Heurísticas Construtivas para
Problemas de Agrupamento Aplicado à PCR Multiplex

Juiz de Fora
2019

Alúcio Cardoso Silva

**Um Algoritmo Evolutivo Baseado em Heurísticas Construtivas para
Problemas de Agrupamento Aplicado à PCR Multiplex**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. D.Sc. Carlos Cristiano Hasenclever Borges

Coorientador: Prof. D.Sc. João Batista Ribeiro

Juiz de Fora

2019

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Cardoso Silva, Aluísio.

Um Algoritmo Evolutivo Baseado em Heurísticas Construtivas para Problemas de Agrupamento Aplicado à PCR Multiplex / Aluísio Cardoso Silva. – 2019.

60 f. : il.

Orientador: Carlos Cristiano Hasenclever Borges

Coorientador: João Batista Ribeiro

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2019.

1. PCR multiplex. 2. Algoritmo genético. 3. Heurísticas construtivas. 4. Problema do empacotamento. 5. Problemas de agrupamento. I. Hasenclever Borges, Carlos Cristiano, orient. II. Ribeiro, João Batista, coorient. III. Título.

Alúcio Cardoso Silva

**Um Algoritmo Evolutivo Baseado em Heurísticas Construtivas para
Problemas de Agrupamento Aplicado à PCR Multiplex**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Aprovada em 30 de agosto de 2019

BANCA EXAMINADORA

Prof. D.Sc. Carlos Cristiano Hasenclever Borges -
Orientador
Universidade Federal de Juiz de Fora

Prof. D.Sc. João Batista Ribeiro - Coorientador
Empresa Brasileira de Pesquisa Agropecuária

Prof. D.Sc. Wagner Antônio Arbex
Universidade Federal de Juiz de Fora

Prof. D.Sc. Fabrízio Condé de Oliveira
Universidade Federal Fluminense

AGRADECIMENTOS

A Deus, causa primária de todas as coisas, por tudo aquilo que conspirou a meu favor e contribuiu para a conquista pessoal e profissional materializada neste trabalho.

Aos meus pais, Dayse e Rivelino, pelo amor e suporte incondicionais, além do imensurável esforço dedicado à minha criação e formação, o que me permitiu tal realização.

À minha namorada, Aline, pelo carinho e compreensão diante de toda a dedicação e abdicção exigidas pelo mestrado e também à sua família por todo apoio que me foi dado.

Aos meus amigos e familiares, pelos incontáveis momentos de descontração, imprescindíveis e inspiradores.

A todos os professores com os quais pude contar ao longo da minha trajetória escolar ou acadêmica que, ao compartilharem seus conhecimentos e experiências, me forneceram ferramentas fundamentais à viabilização deste trabalho.

Aos meus orientadores, Carlos Cristiano e João Batista, pela confiança, parceria e pelos valiosos direcionamentos dados durante todo o processo.

Aos membros da banca, professores Fabrízio e Wagner, pelas importantes contribuições que enriqueceram esta dissertação.

À CAPES e à UFJF, pelo apoio financeiro.

A todos que de alguma forma vibraram positivamente por mim.

“Se as portas da percepção estivessem limpas, tudo apareceria para o homem tal como é: infinito.”

William Blake

RESUMO

A reação em cadeia da polimerase (*Polymerase Chain Reaction*, PCR) é um dos métodos de biologia molecular mais utilizados em laboratórios clínicos e de pesquisa. Com a PCR é possível gerar bilhões de cópias de um determinado fragmento de DNA em um curto período de tempo. O potencial dessa reação e de suas variantes é explorado em um vasto conjunto de aplicações inseridas em diferentes campos científicos, motivando a pesquisa direcionada à otimização dos ensaios de PCR. A PCR *multiplex* consiste em uma variação da PCR convencional que permite a amplificação de múltiplos fragmentos específicos de DNA em um mesmo tubo, propiciando economia de tempo, custos e principalmente amostras do material genético. O projeto da reação é especialmente desafiador na medida em que exige a difícil tarefa de agrupar as amplificações de acordo com a compatibilidade dos componentes envolvidos. Geralmente, métodos *in silico* para essa reação são baseados no problema combinatório decorrente do agrupamento dos pares de *primers* utilizados, que consistem em sequências curtas de DNA sintetizadas para delimitar especificamente cada fragmento alvo. Neste trabalho, foi desenvolvido um modelo computacional para o problema de agrupamento da PCR multiplex, visando uma abordagem que compreenda, simultaneamente, aspectos determinantes para a aplicabilidade do modelo e a otimização eficiente da reação, a saber: o tratamento de problemas que exijam múltiplos tubos de PCR *multiplex* para a cobertura dos alvos; a minimização do número de tubos necessários; a utilização de uma estratégia de busca estocástica em oposição a algoritmos determinísticos; o desacoplamento do método de busca em relação ao conjunto de medidas de compatibilidade adotadas; e a capacidade de tratar o cenário mais complexo em que duas ou mais opções pares de *primers* são fornecidas por amplificação. O modelo é composto pela adaptação de um algoritmo evolutivo à busca de permutações dos elementos e uma heurística construtiva responsável pela decodificação das soluções mapeadas. Além disso, um processo de restrição do espaço de busca é implementado visando aprimorar o desempenho da busca. A construção da proposta foi inspirada em métodos desenvolvidos para a solução do conhecido problema do empacotamento, o que permitiu uma avaliação inicial baseada em *benchmarks* amplamente explorados na literatura. Nesse caso, a análise comparativa apresentada evidencia a competitividade do modelo desenvolvido diante dos algoritmos referenciados. Posteriormente, o modelo foi adaptado para a otimização da PCR *multiplex*. Os resultados de experimentos exploratórios realizados indicam a escalabilidade do modelo e ressaltam a relevante contribuição decorrente da amplitude da abordagem. Finalmente, é apresentada uma análise experimental comparativa com o programa MultiPLX, que baseia-se em uma formulação semelhante do problema. Os resultados superiores obtidos pelo algoritmo proposto reforçam a aplicabilidade do modelo desenvolvido.

Palavras-chave: PCR multiplex. Algoritmo genético. Heurísticas construtivas. Problema do empacotamento. Problemas de agrupamento.

ABSTRACT

Polymerase chain reaction (PCR) is one of the most widely used molecular biology methods in clinical and research laboratories. Through it, it is possible to generate billions of copies of a given DNA fragment in a short period of time. The potential of this reaction and its variants is explored in a wide range of applications in different scientific fields, which motivates research aimed at optimizing PCR assays. Multiplex PCR is a variation of conventional PCR that allows the amplification of multiple specific DNA fragments in the same assay, saving time, costs, and especially samples of genetic material. Designing this reaction is especially challenging because it requires the difficult task of efficiently grouping the amplifications according to the compatibility of their components. This task is usually abstracted to a combinatorial problem about grouping the primer pairs used, which consist of short sequences of DNA synthesized to delimit specifically each target fragment. In this work, a computational model was developed to solve the multiplex PCR grouping problem aiming at an approach capable of dealing with interests frequently approached in isolation by other works. They are: the use of a robust computational strategy as opposed to deterministic algorithms; the applicability of the model in contexts requiring grouping of amplifications in multiple multiplex PCR tubes; minimizing the number of tubes required; the dissociation of the search method from the set of compatibility measures adopted; and the ability to handle the more complex scenario in which two or more options of primer pairs are provided by amplification, expanding the universe of possibilities and potentially the quality of the results. The construction of the model was inspired by methods previously applied to the known bin packing problem. Thus, an evolutionary algorithm is adapted to the search for specific element permutations aiming at the subsequent decoding of the solutions through a building heuristic. Besides, is presented a search space restriction process that allowed to improve the optimization performance. The approach was initially adapted to the bin packing problem, which allowed an evaluation based on benchmarks widely explored in the literature. In this case, the comparative analysis presented points to the competitiveness of the developed model in relation to the referenced algorithms. Subsequently, the proposal was adapted to the multiplex PCR problem. The results of exploratory experiments conducted indicate the scalability of the model and highlight the relevant contribution arising from the breadth of the approach. Finally, it is presented a comparative experimental analysis with the MultiPLX program, which is available for multiplex PCR design and is based on a similar problem formulation. The results obtained by the developed algorithm were superior in the three considered cases, reinforcing the applicability of the proposed model.

Keywords: Multiplex PCR. Genetic algorithm. Building heuristics. Bin packing problem. Grouping problems.

LISTA DE FIGURAS

Figura 1 – Anelamento dos <i>primers</i> no DNA-molde	11
Figura 2 – Ciclo da PCR	13
Figura 3 – Crescimento exponencial das moléculas de DNA	13
Figura 4 – Eletroforese em gel de agarose	14
Figura 5 – Representação das codificações utilizadas	33

LISTA DE TABELAS

Tabela 1	– Resultados obtidos pelo AGBH para o problema do empacotamento	43
Tabela 2	– Resultados obtidos pelo AGBHcB para o problema do empacotamento	43
Tabela 3	– Resultados obtidos pelo AGBHcBD para o problema do empacotamento	43
Tabela 4	– Comparação entre AGBHcBD e algoritmos de referência para o BPP	43
Tabela 5	– Comparativo entre funções objetivo: <i>primers</i> pré-selecionados com 15 alvos	48
Tabela 6	– Comparativo entre funções objetivo: <i>primers</i> variáveis com 15 alvos	48
Tabela 7	– Comparativo entre funções objetivo: <i>primers</i> pré-selecionados com 30 alvos	48
Tabela 8	– Comparativo entre funções objetivo: <i>primers</i> variáveis com 30 alvos	48
Tabela 9	– Comparativo entre funções objetivo: <i>primers</i> pré-selecionados com 70 alvos	48
Tabela 10	– Comparativo entre funções objetivo: <i>primers</i> variáveis com 70 alvos	48
Tabela 11	– Comparativo entre funções objetivo: <i>primers</i> pré-selecionados com 150 alvos	48
Tabela 12	– Comparativo entre funções objetivo: <i>primers</i> variáveis com 150 alvos	48
Tabela 13	– Comparativo entre as buscas: <i>primers</i> pré-selecionados com 15 alvos	50
Tabela 14	– Comparativo entre as buscas: <i>primers</i> variáveis com 15 alvos . .	50
Tabela 15	– Comparativo entre as buscas: <i>primers</i> pré-selecionados com 30 alvos	50
Tabela 16	– Comparativo entre as buscas: <i>primers</i> variáveis com 30 alvos . .	50
Tabela 17	– Comparativo entre as buscas: <i>primers</i> pré-selecionados com 70 alvos	50
Tabela 18	– Comparativo entre as buscas: <i>primers</i> variáveis com 70 alvos . .	50
Tabela 19	– Comparativo entre as buscas: <i>primers</i> pré-selecionados com 150 alvos	50
Tabela 20	– Comparativo entre as buscas: <i>primers</i> variáveis com 150 alvos .	50
Tabela 21	– Resultados obtidos pelo AGBHcBD com dados biológicos	52
Tabela 22	– Resultados obtidos pelo software MultiPLX com dados biológicos	52

SUMÁRIO

1	INTRODUÇÃO	11
1.1	OTIMIZAÇÃO DA PCR <i>MULTIPLEX</i>	15
1.2	MOTIVAÇÃO	16
1.3	OBJETIVOS	17
1.4	ESTRUTURA DO TEXTO	17
2	MODELAGEM COMPUTACIONAL DA PCR <i>MULTIPLEX</i>	19
2.1	PCR <i>MULTIPLEX</i> E O PROBLEMA DO EMPACOTAMENTO	22
2.1.1	Métodos heurísticos para o problema do empacotamento	23
3	UMA ABORDAGEM COMPUTACIONAL PARA PROBLEMAS DE AGRUPAMENTO	27
3.1	ADAPTAÇÃO DO MODELO AO PROBLEMA DO EMPACOTAMENTO	27
3.1.1	Restrição do espaço de busca	28
3.1.2	Algoritmo Genético Baseado em Heurística (AGBH)	29
3.1.3	Algoritmo Genético Baseado em Heurística com Bloqueio (AGBHcB)	31
3.1.4	Algoritmo Genético Baseado em Heurística com Bloqueio Dinâmico (AGBHcBD)	32
3.1.5	Implementação modificada da heurística <i>First Fit</i>	33
4	ADAPTAÇÃO DO MODELO AO PROBLEMA DA PCR <i>MULTIPLEX</i>	34
4.1	ABSTRAÇÃO DA REAÇÃO DE PCR <i>MULTIPLEX</i>	34
4.2	RESTRIÇÃO DO ESPAÇO DE BUSCA (PCR <i>MULTIPLEX</i>)	35
4.3	PCR <i>MULTIPLEX</i> COM <i>PRIMERS</i> PRÉ-SELECIONADOS	36
4.3.1	AGBHcBD adaptado à PCR <i>multiplex</i> com <i>primers</i> pré-selecionados	36
4.4	EXTENSÃO DO MODELO PARA PCR <i>MULTIPLEX</i> COM <i>PRIMERS</i> VARIÁVEIS	38
4.4.1	AGBHcBD adaptado à PCR <i>multiplex</i> com <i>primers</i> variáveis	39
5	EXPERIMENTOS COMPUTACIONAIS	41
5.1	PROBLEMA DO EMPACOTAMENTO	41
5.1.1	Configuração experimental	41
5.1.2	Resultados e análise experimental	42
5.2	PCR <i>MULTIPLEX</i>	44
5.2.1	Dados sintéticos	44
5.2.2	Dados biológicos	45
5.2.3	Configuração experimental	46
5.2.4	Resultados e análise experimental	46

5.2.4.1	<i>Análise experimental da função objetivo</i>	47
5.2.4.2	<i>Análise experimental de qualidade e escalabilidade</i>	49
5.2.4.3	<i>Análise experimental com dados biológicos</i>	51
6	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS . .	54
6.1	PROBLEMA DO EMPACOTAMENTO	54
6.2	OTIMIZAÇÃO DA PCR <i>MULTIPLEX</i>	55
	REFERÊNCIAS	57

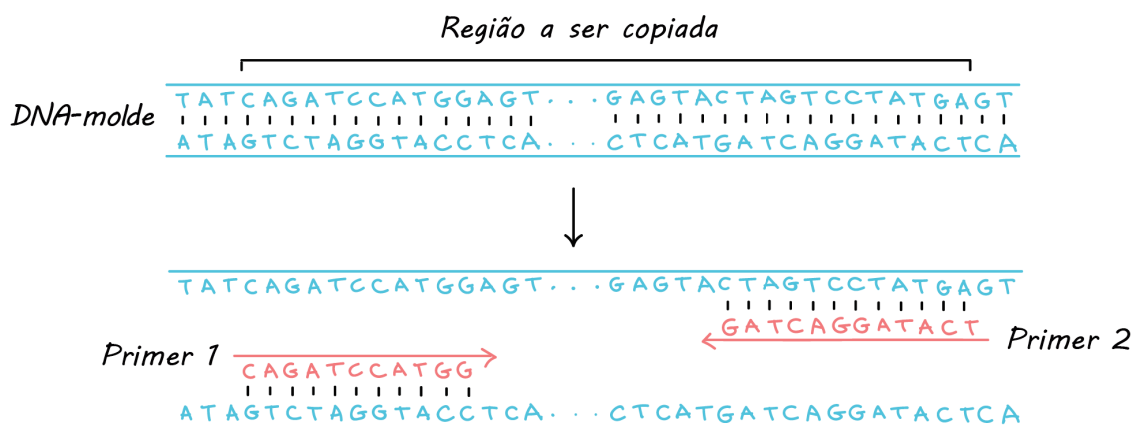
1 INTRODUÇÃO

Modelos computacionais para o projeto, a análise e a simulação de sistemas reais são frequentemente desenvolvidos com o intuito de impulsionar as pesquisas nos mais diversos campos da ciência, especialmente naquele abordado nesta dissertação, o da biologia. Definida como um campo resultante da aplicação de instrumentos da computação nas ciências biológicas, a bioinformática se tornou mais do que isso, uma grande fonte de problemas algorítmicos e estatísticos, que impactam profundamente as ciências computacionais (MANDOIU; ZELIKOVSKY, 2008). Na área da biologia molecular, por exemplo, problemas modelados computacionalmente dão origem a novas linhas de pesquisa, as quais buscam explorar e otimizar os ensaios de bancada de laboratório em proporções, muitas vezes, inviáveis na prática.

A reação em cadeia da polimerase (*Polymerase Chain Reaction*, PCR) é um método fundamental da biologia molecular para amplificação de ácido desoxirribonucleico (DNA), conhecido por propiciar uma revolução nos estudos da genética. Através dela, é possível amplificar *in vitro* um segmento específico de DNA, gerando milhões, ou bilhões, de cópias de fragmentos específicos de DNA em um único experimento, dentro de poucas horas. O crescente conjunto de aplicações da PCR inclui o estudo e a detecção de genes de resistência e mutações, diagnóstico de doenças, caracterização de organismos, testes de paternidade e impressão digital genética, além de abranger diferentes campos científicos como agricultura, medicina forense, ecologia molecular, biotecnologia e arqueologia (HENEGARIU et al., 1997; BUSTIN, 2010; BASU, 2015).

Dentre os componentes necessários para o ensaio da PCR estão o DNA-molde, extraído de uma amostra de interesse e do qual pretende-se amplificar uma determinada sequência alvo; altas concentrações do par de *primers* ou iniciadores (direto e reverso)

Figura 1 – Anelamento dos *primers* no DNA-molde



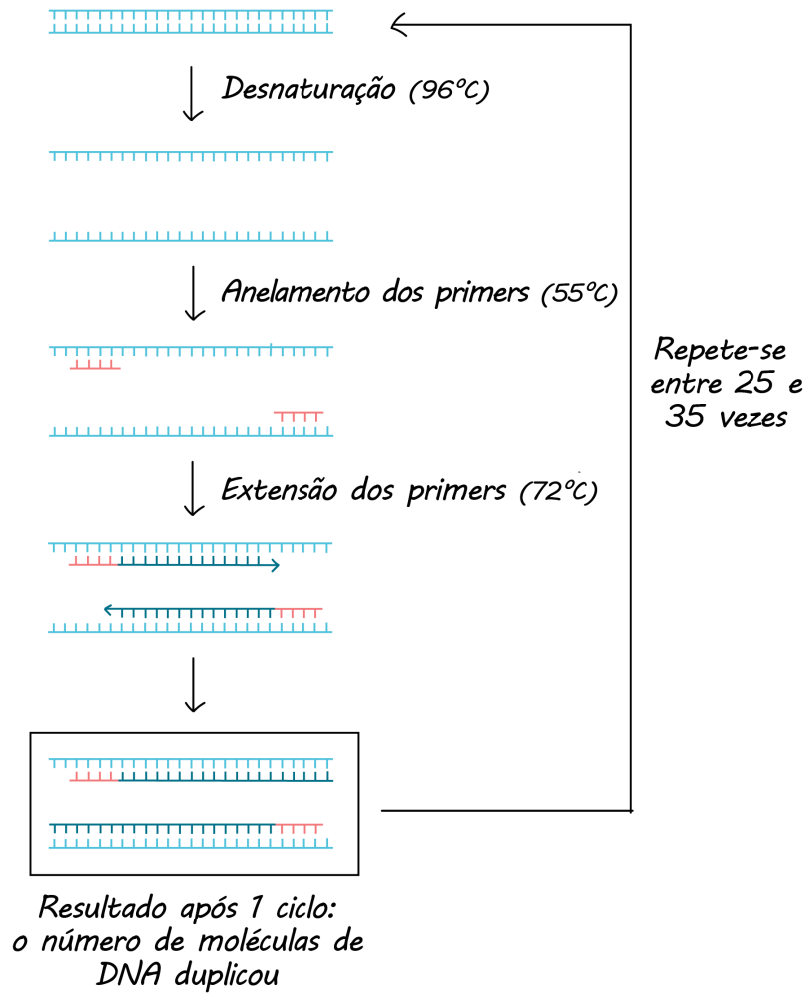
específicos em relação à sequência alvo, que constituem sequências sintéticas curtas de DNA responsáveis por delimitar o fragmento a ser replicado, como mostra a Figura 1; os dNTPs (desoxirribonucleotídeos trifosfato: A,T,C,G), que figuram como matéria prima da reação; a enzima DNA-polimerase, responsável por catalisar a reação; e reagentes químicos auxiliares como o cloreto de magnésio ($MgCl_2$) e o cloreto de potássio (KCl).

Na prática, tubos contendo os componentes supradescritos em solução são inseridos em um termociclador, que promove variações cíclicas de temperatura estimulando cada uma das fases do processo repetidamente. Cada ciclo envolve, em temperaturas específicas, a separação das fitas do DNA-molde, a hibridização ou anelamento dos *primers* aos alvos específicos e a extensão desses *primers* por meio da adição de dNTP's, promovida pela DNA polimerase, formando moléculas de DNA (Figura 2). Em condições ideais, cada ciclo permite dobrar o número dessas moléculas, promovendo um crescimento exponencial na quantidade de fragmentos replicados, também chamados de produtos de amplificação ou amplicons (Figura 3). Por fim, o material é submetido à separação eletroforética em gel, o qual é subsequentemente corado e sua imagem fotodocumentada para posterior interpretação do resultado. A Figura 4 apresenta um exemplo dos resultados da eletroforese em gel de agarose. As colunas das extremidades apresentam os marcadores de referência de tamanho dos amplicons, medidos em pares de base (pb). As colunas de 1 a 10 indicam os resultados dos ensaios que tentaram amplificar simultaneamente, através da PCR *multiplex*, os genes ermB (359pb), tetO (1.723pb) e tetM (1.862pb), a partir do DNA de amostras de leite artificialmente inoculadas com diferentes linhagens do patógeno *S. agalactiae*.

A etapa chave para o sucesso da reação é denominada por “desenho de *primers*” e consiste na definição das sequências sintéticas de DNA utilizadas como iniciadores da reação, as quais devem ser complementares em relação às extremidades da região alvo, a fim de garantir uma hibridização estável. A região alvo deve englobar a real região de interesse do experimento e, desse modo, o desenho de *primers* pode ser desempenhado através de uma busca no DNA-molde direcionada por propriedades críticas de avaliação dos *primers*. Tais propriedades incluem a temperatura de *melting* (T_m) de cada *primer* e a amplitude máxima entre elas, o percentual de guanina e citosina, o comprimento dos *primers* em número de bases, o ΔG (energia livre de *Gibbs*) calculado sobre a extremidade 3' dos *primers*, a especificidade dos *primers* em relação à sequência alvo, a possibilidade de formações indesejadas como estruturas secundárias (um *primer* se anelando nele mesmo) ou dímeros (um *primer* se anelando a outro *primer*), além de outros critérios adicionais. As faixas ótimas de valores para cada parâmetro, bem como os cálculos aproximativos utilizados para obtenção de alguns deles (como T_m e ΔG , por exemplo) podem variar em diferentes trabalhos da literatura (SHEN et al., 2009; BASU, 2015).

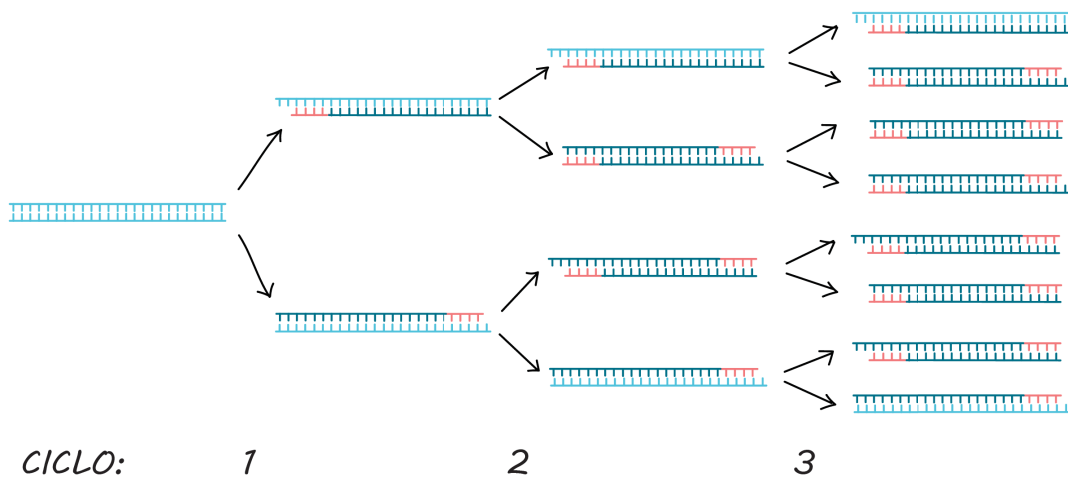
A PCR *multiplex* é uma variação da PCR convencional (ou PCR *monoplex*) e permite a amplificação simultânea de múltiplos fragmentos de DNA por meio da inclusão de diferentes pares de *primers* no tubo, cada um deles específico em relação ao seu respectivo

Figura 2 – Ciclo da PCR



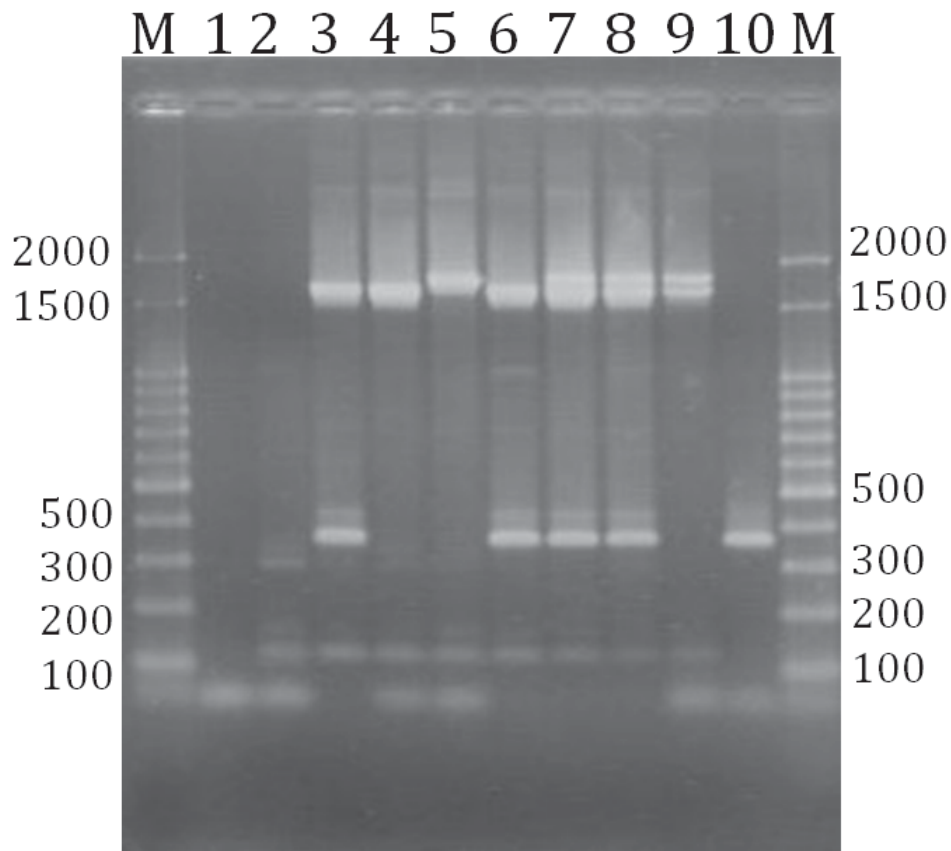
Fonte: www.khanacademy.org (acesso em agosto de 2019)

Figura 3 – Crescimento exponencial das moléculas de DNA



Fonte: www.khanacademy.org (acesso em agosto de 2019)

Figura 4 – Eletroforese em gel de agarose



Fonte: Ribeiro et al. (2015)

alvo. Dessa forma, o total de reações necessárias pode ser reduzido drasticamente, propiciando economia de tempo, reagentes químicos e principalmente de amostras do material genético, que são limitadas em muitos casos (ZANGENBERG; SAIKI; REYNOLDS, 1999). Embora aplicada com sucesso em muitas áreas de diagnóstico de DNA, a união das reações (ou amplificações) em um mesmo tubo é especialmente desafiadora, na medida em que envolve a possibilidade de interações cruzadas indesejadas em função do agrupamento de *primers* com propriedades e sequências distintas. Por esse motivo, o projeto da reação deve ser baseado em critérios que minimizem as chances dessas ocorrências e potencializem a compatibilização pretendida.

A seleção de *primers* eficientes e compatíveis entre si concentram os maiores esforços do projeto da PCR *multiplex*. Ao defini-los, o principal desafio consiste em evitar a perda dos produtos de amplificação causada pelo pareamento indesejado entre *primers*, formando dímeros (ZANGENBERG; SAIKI; REYNOLDS, 1999). Neste caso, a prevenção é geralmente baseada em uma análise da estabilidade das possíveis hibridizações entre dois *primers* quaisquer inseridos em um mesmo tubo. Adicionalmente, a diminuição dos produtos de amplificação pode resultar do anelamento dos *primers* em regiões inespecíficas do DNA-molde, ou seja, fora do alvo. Segundo Kalendar et al. (2017), a probabilidade

desta ocorrência pode ser minimizada quando há homogeneidade entre as T_m 's dos *primers* agrupados, visto que elas estão diretamente associadas à temperatura ideal de anelamento de cada *primer* e, portanto, à eficiência das amplificações. Outro problema em potencial está relacionado à garantia de interpretabilidade do resultado da eletroforese, uma vez que fragmentos de DNA com comprimentos relativamente próximos são dificilmente separáveis em gel de agarose e podem figurar como uma única banda. Deste modo, uma diferença muito pequena entre os comprimentos dos amplicons pode inviabilizar ou distorcer a interpretação da imagem resultante (SHEN et al., 2009).

1.1 OTIMIZAÇÃO DA PCR *MULTIPLEX*

A complexidade das interações envolvidas no ensaio da PCR torna o projeto manual da reação e de suas variantes altamente propenso a erros, provocando um aumento de custos e tempo necessários para sua execução bem sucedida. Esse fato, aliado à importância e popularidade do método, motivaram uma série de trabalhos voltados à otimização do projeto das reações através de inteligência computacional nos últimos anos (CHENG, 2014; GARCÍA et al., 2015). Aqueles voltados especialmente à PCR *multiplex*, se concentram predominantemente na tarefa crítica de compatibilização dos *primers* envolvidos, abstraindo a reação ao problema de agrupamento de *primers*.

De maneira geral, o projeto da PCR *multiplex* através de modelos computacionais pode ser baseado em duas abordagens distintas. Na primeira delas, a etapa de desenho dos *primers* pode ser condicionada à compatibilização das amplificações, determinando restrições que imponham a uniformidade e a compatibilidade desejada entre as propriedades dos *primers*. Entretanto, assim como apontado por Lee, Shin e Zhang (2007), trabalhos baseados nessa estratégia se limitam frequentemente a um único tubo de PCR *multiplex*, descartando os alvos que não permitem a obtenção de *primers* compatíveis. Além da possível redução na cobertura, há ainda o prejuízo relacionado à escalabilidade dessas soluções, ao executarem, em sua maioria, buscas exaustivas para obtenção dos *primers*. Essa limitação é ainda mais crítica ao considerar aplicações da PCR *multiplex*, como a genotipagem, que envolvem o interesse em amplificar mais de cem sequências alvo. Nesses casos, ainda que as restrições de compatibilidade fossem atendidas, seria inviável desenvolver um ensaio baseado em um agrupamento dessa magnitude, dado que a concentração fracionária insuficiente de cada *primer* impediria a amplificação desejada, exigindo um particionamento das amplificações em múltiplas reações de PCR *multiplex* (LIPSON, 2003).

Outra abordagem computacional amplamente empregada no projeto da PCR *multiplex* considera, para cada alvo, a utilização de *primers* originalmente projetados para a PCR *monoplex*. Tal abordagem prevê a possibilidade de agrupar as amplificações em múltiplos tubos, visando à cobertura total dos alvos pretendidos com base nas restrições

de compatibilidade impostas. Dessa forma, é possível superar as limitações de cobertura e escalabilidade da estratégia anterior. Além disso, deve-se considerar que o desenho de *primers* para a PCR *monoplex* é uma tarefa realizada por uma ampla lista de *softwares* disponíveis, alguns sofisticados e largamente configuráveis em relação aos parâmetros da reação, o que permite a obtenção de *primers* de alta confiabilidade para a amplificação do respectivo alvo (BASU, 2015). Ademais, a utilização de sequências já empregadas com sucesso anteriormente em outros trabalhos consiste em outra prática comum e potencialmente confiável para definir *primers* eficientes para um determinado alvo de PCR. Esses fatores tornam esta abordagem ainda mais relevante, na medida em que demandam a capacidade de compatibilizar *primers* pré-projetados para diferentes contextos.

1.2 MOTIVAÇÃO

Diante do potencial mencionado anteriormente, este trabalho foi baseado na abordagem de otimização da PCR *multiplex* a partir do agrupamento de *primers* pré-projetados de maneira independente. Em trabalhos anteriores, derivados da mesma abordagem, foram observadas limitações que motivaram o desenvolvimento do modelo computacional da reação proposto nesta dissertação. Tais limitações estão associadas a interesses relevantes para a qualidade dos resultados e aplicabilidade da solução, no entanto, são raramente abordados conjuntamente.

Primeiramente, deve-se destacar o interesse em obter o agrupamento mais eficiente das ampliações desejadas. Apesar da respectiva conformidade com o próprio objetivo da variação *multiplex* da PCR, algumas das ferramentas propostas anteriormente não preveem a minimização do número de tubos necessários, exigindo que esse valor seja fixado previamente. Outra limitação recorrente está associada aos algoritmos utilizados para solucionar o problema, que em muitos casos são baseados em abordagens computacionais determinísticas e conseqüentemente limitados frequentemente a soluções subótimas. Uma terceira capacidade dificilmente encontrada em outras propostas é a de lidar com múltiplas opções de pares de *primers* para amplificar cada um dos alvos desejados. A inclusão de pares candidatos amplia o universo de possibilidades e o potencial de obtenção de agrupamentos eficientes, porém, implica no aumento da complexidade do problema, principalmente diante da perspectiva de tratar simultaneamente as tarefas de seleção e agrupamento de *primers*. Ademais, as formulações encontradas para a otimização do ensaio de PCR *multiplex* apresentam uma considerável variação em relação ao conjunto de medidas de compatibilidade adotado. Esse fato motiva o desenvolvimento de um modelo flexível, que permita alterações no respectivo conjunto. Entretanto, tal flexibilidade nem sempre é encontrada nos trabalhos de referência, que em alguns casos apresentam forte acoplamento das medidas de compatibilidade em relação ao método de busca utilizado.

1.3 OBJETIVOS

Este trabalho tem como objetivo principal desenvolver um modelo computacional para o projeto da PCR *multiplex*, atendo-se a duas características principais: (i) amplitude da abordagem em relação aos interesses relevantes mencionados anteriormente; (ii) viabilidade de generalização da estratégia de busca para a solução de outros problemas de agrupamento.

O desenvolvimento do modelo teve como base referências do conhecido “problema do empacotamento”, o que permitiu enriquecer o processo construtivo e experimental. Dessa maneira, a fim de alcançar o objetivo principal, foram definidas as etapas a seguir:

- Desenvolver um modelo computacional para problemas de agrupamento inicialmente aplicado ao problema do empacotamento, com o intuito de avaliar seu desempenho diante de *benchmarks* amplamente explorados na literatura computacional.
- Implementar e avaliar um método de restrição do espaço de busca para aprimorar o desempenho do modelo. Posteriormente, avaliar comparativamente a qualidade do modelo diante de algoritmos de referência para o problema do empacotamento.
- Adaptar o modelo ao problema da PCR *multiplex* para dois cenários distintos: *primers* pré-selecionados e *primers* variáveis.
- Avaliar, através de experimentos exploratórios, a qualidade e a escalabilidade do modelo adaptado à PCR *multiplex*. Além disso, comparar experimentalmente o modelo desenvolvido com a ferramenta MultiPLX, disponível para o projeto da PCR *multiplex*.

Por fim, pretende-se evidenciar, através da argumentação teórica e dos resultados empíricos, a possibilidade de generalização do modelo desenvolvido para outros problemas de agrupamento.

1.4 ESTRUTURA DO TEXTO

Esta dissertação está organizada em 6 capítulos. Após a introdução, são apresentadas as revisões da literatura acerca da otimização do projeto da PCR *multiplex* e do problema do empacotamento, contextualizando a relação explorada entre os dois problemas no desenvolvimento do modelo computacional proposto. No capítulo 3, são apresentados os algoritmos implementados com base no modelo proposto a fim de solucionar o problema do empacotamento. No capítulo 4, são descritas as particularidades envolvidas na adaptação do modelo para o tratamento do problema da PCR *multiplex* com *primers* pré-selecionados, bem como a extensão do modelo para a solução do problema da PCR *multiplex* com

primers variáveis. Os experimentos computacionais são apresentados no capítulo 5 e, por último, no capítulo 6, são destacadas as considerações finais deste trabalho e as perspectivas de trabalhos futuros.

2 MODELAGEM COMPUTACIONAL DA PCR *MULTIPLEX*

Nicodème e Steyaert (1997) provaram que o problema de agrupamento máximo de amplificações sujeito às restrições de PCR *multiplex* é NP-completo, o que implica na ausência de um algoritmo exato conhecido para resolvê-lo em tempo polinomial. Dessa forma, métodos heurísticos ou meta-heurísticos são encorajados como a principal estratégia para a otimização do projeto da PCR *multiplex*. A seguir, serão apresentadas as principais características dos trabalhos relacionados, destacando especialmente a variação das formulações propostas para o problema e suas limitações quanto à amplitude da abordagem. Inicialmente, são descritas as propostas baseadas na etapa de desenho de *primers* e, posteriormente, os trabalhos voltados ao agrupamento de *primers* pré-projetados de maneira independente.

Liang, Lee e Wu (2005) desenvolveram um algoritmo genético (AG) adaptado ao desenho de *primers* para PCR *multiplex*. As propriedades críticas para avaliação de *primers* no contexto da PCR *monoplex* são utilizadas como restrições para a geração da população inicial. Para cada indivíduo, a busca se repete até que seja encontrado um par de *primers* factível para cada alvo. Posteriormente, assumindo a convergência da etapa inicial, o AG promove uma busca pela melhor combinação entre os *primers* encontrados. A avaliação dos indivíduos e o direcionamento da busca ocorrem através de uma função objetivo composta pela agregação de um conjunto de penalizações, que possuem pesos distintos e são baseadas em medidas de compatibilização. A proposta considera apenas uma reação de PCR *multiplex* (um único tubo) e embora os experimentos computacionais realizados para três e quatro alvos tenham encontrado combinações capazes de satisfazer todos os critérios considerados, o mesmo não ocorreu no teste com cinco alvos.

A ferramenta apresentada em (WU et al., 2007) se baseia em uma abordagem similar à anteriormente descrita. Um processo de obtenção de *primers* candidatos a amplificarem cada sequência alvo é combinado ao posterior agrupamento das reações realizado por um algoritmo genético altamente especializado. Através de uma função objetivo baseada em penalizações, o AG procura minimizar a ocorrência de dímeros e de amplicons com menos de 50 pb (pares de bases) de diferença de comprimento entre si. No entanto, a ferramenta falha nos casos em que não é possível obter *primers* que atendam às restrições da primeira etapa, além de exigir, ao fim do processo, uma verificação manual para otimizar o número de *primers* em cada tubo e evitar amplificações inespecíficas. Ademais, para prever a formação de dímeros, a ferramenta se baseia apenas na pontuação simples de complementaridade de bases entre as sequências, reconhecidamente menos confiável e com menor nível quantitativo em relação ao cálculo de estabilidade sobre parâmetros termodinâmicos, como o ΔG (VALLONE; BUTLER, 2004).

No trabalho desenvolvido por García et al. (2015), o projeto da PCR *multiplex* é

condicionado à etapa de desenho de *primers*, visto que, para cada sequência alvo, busca-se um par de *primers* que necessariamente viabilize o agrupamento das reações. Para tanto, as propriedades do par de *primers* encontrado para a primeira amplificação são utilizadas como restrições para as buscas dos pares restantes. A estratégia é baseada na união do algoritmo Simplex, responsável por buscar o primeiro par, com o algoritmo PSO (Particle Swarm Optimization), que busca os pares restantes. Experimentos *in vitro* baseados em sequências de *primers* obtidas pelo algoritmo indicaram a eficiência da estratégia para ensaios com até sete alvos, desde que uma única reação de PCR *multiplex* seja suficiente para alcançar a compatibilização das amplificações desejadas.

Abordagens que condicionam o projeto da PCR *multiplex* à etapa de desenho dos *primers* são encontradas em outros trabalhos que compartilham limitações semelhantes às destacadas até então, se mostrando pouco extensíveis a problemas maiores e descartando alvos de interesse para obter a compatibilização desejada (SCHOSKE et al., 2003; ROUCHKA; KHALYFA; COOPER, 2005; LEE et al., 2006; YAMADA; SOMA; MORISHITA, 2006). Por outro lado, como destacado no capítulo inicial, as abordagens com foco específico na etapa de agrupamento das reações podem se apresentar como uma alternativa promissora diante dessas limitações. Trabalhos desenvolvidos nesta linha são apresentados a seguir.

O trabalho de Lipson (2003) apresenta um modelo computacional para o problema de agrupamento da PCR *multiplex* baseado no problema de coloração de grafos. O cenário com *primers* pré-selecionados é abordado através da heurística *smallest last order* (SLO) que provê um agrupamento sequencial com uma garantia mínima de qualidade. Para o cenário com múltiplos *primers* candidatos por amplificação, a heurística SLO é combinada a um algoritmo genético responsável por selecionar previamente um par de *primers* por amplificação. As medidas de compatibilidade utilizadas para prevenção da ocorrência de dímeros e anelamentos inespecíficos são baseadas em pontuação simples de correspondência entre as bases com nível de rigorosidade definido por parâmetros adicionais. Assim, o ajuste do algoritmo se torna complexo, o que foi apontado como uma das possíveis justificativas para os resultados insatisfatórios obtidos em experimentos reais.

O software MultiPLX (KAPLINSKI et al., 2005) utiliza um algoritmo guloso para agrupar as reações de PCR de acordo com as restrições de compatibilidade impostas pelo usuário. As possibilidades de medidas de interação incluem o limite inferior de energia livre de ligação entre os *primers* e, opcionalmente, entre *primers* e produtos, a diferença mínima entre as temperaturas de *melting* dos *primers* e a diferença mínima entre os comprimentos dos amplicons. O programa permite que mais de um par de *primers* candidato seja fornecido para cada um dos alvos das amplificações e a seleção de um candidato ocorre simultaneamente ao processo de agrupamento. A natureza míope do algoritmo empregado torna a qualidade da busca altamente dependente do ordenamento dos dados. O programa é considerado como base comparativa em trabalhos recentes, embora, eventualmente,

possa não encontrar uma solução que atenda às restrições predefinidas pelo usuário, como reportado em (MILEWSKI et al., 2017).

Lee, Shin e Zhang (2007) propõem uma adaptação do algoritmo ϵ -MOEA para resolver o problema da PCR *multiplex* através de uma formulação baseada em otimização multiobjetivo. O algoritmo permite fornecer pares de *primers* candidatos pré-projetados de maneira independente para cada alvo e busca compatibilizá-los minimizando o número de tubos e outras duas medidas de interação entre os possíveis pareamentos. Assim como apontado pelos autores, os resultados razoáveis obtidos demonstraram a necessidade de um maior desenvolvimento do modelo. Os agrupamentos ainda apresentaram um alto índice de interações cruzadas indesejáveis ao fim do processo, indicando uma possível dificuldade em tratar o problema sem a aplicação de restrições.

A ferramenta MpPrimer (SHEN et al., 2009) integra diferentes *softwares* para obter pares de *primers* candidatos a amplificarem cada sequência alvo desejada e estabelecer uma matriz de compatibilidade entre eles. Posteriormente, a seleção de cada par é desempenhada através de um algoritmo de expansão de grafos derivado de um algoritmo guloso. Os parâmetros críticos para o sucesso da reação são considerados no processo de escolha dos candidatos, porém, a aplicação do algoritmo limita-se aos casos em que existam soluções viáveis para um único tubo de PCR *multiplex*. A ampliação da abordagem para o tratamento de casos que exijam múltiplas reações de PCR *multiplex* é destacada como interesse futuro.

O trabalho de Milewski et al. (2017) descreve uma adaptação do *Genetic Grouping Algorithm*, proposto por Falkenauer (1994) para o problema do empacotamento, o qual é modificado para o agrupamento de reações de PCR. O algoritmo é capaz de agrupar um conjunto definido de *primers* com base em um dos seguintes objetivos: maximização de energia de ligação entre possíveis pareamentos indesejáveis ou minimização da faixa de T_m dentro de cada grupo. A ferramenta não permite incluir opções de pares de *primers* para cada alvo de amplificação no processo de otimização, além de trabalhar com o número de grupos (tubos) pré-fixado pelo usuário e um único critério de avaliação por execução.

Assim como na proposta anterior, Brown et al. (2017) propõem o agrupamento de um conjunto de pares de *primers* já definidos para cada um dos alvos da PCR *multiplex*. Considerando o número de grupos pré-fixado pelo usuário, a busca é feita através da técnica de otimização matemática *hill-climbing* e direcionada por dois objetivos principais: minimizar as violações do limite de ΔG definido para possíveis pareamentos entre *primers* e evitar o problema de sobreposição de amplicons pouco abordado em outros trabalhos, denominado por *overlapping*. A proposta não atenta para a homogeneidade das temperaturas de *melting* e nem para a diferença de comprimento entre amplicons.

A modelagem computacional da PCR *multiplex* pode ser encontrada de maneiras essencialmente distintas na literatura. Nota-se que as medidas confiáveis para a otimização

do projeto da reação nem sempre são reunidas em uma mesma abordagem. Além disso, dentre os trabalhos relacionados, poucos apresentam algoritmos robustos em relação a soluções localmente ótimas, principalmente ao analisar conjuntamente os aspectos relevantes já mencionados anteriormente, como a capacidade de lidar com múltiplas reações de PCR *multiplex*, a minimização do número de reações e a manipulação de múltiplos pares de *primers* candidatos para cada amplificação.

2.1 PCR *MULTIPLEX* E O PROBLEMA DO EMPACOTAMENTO

De maneira geral, o projeto da PCR *multiplex* pode ser classificado como um problema de agrupamento clássico, definido pela tarefa de particionar um conjunto de elementos em subconjuntos mutuamente disjuntos. Formalmente, dado que S é o conjunto de todos os elementos e S_i um dos possíveis subconjuntos:

$$\cup S_i = S \quad e \quad S_i \cap S_k = \emptyset, \quad \text{onde } i \neq k \quad (2.1)$$

com cada problema derivado definindo suas próprias restrições de agrupamento e objetivos específicos.

A otimização da PCR *multiplex* está sujeita a restrições de compatibilidade entre os componentes, que visam aumentar as chances de sucesso do ensaio. Alguns dos trabalhos citados anteriormente se baseiam em algoritmos de busca que incorporam essas restrições na função objetivo através de penalizações, o que implica na difícil tarefa de definir os parâmetros de penalização. Naturalmente, dependendo das características dos dados e da eficiência do algoritmo, podem ocorrer violações das restrições na solução encontrada. Além disso, algoritmos baseados nessa abordagem possuem menor grau de flexibilidade diante do acoplamento existente entre o direcionamento da busca e as medidas de interação consideradas. Por outro lado, algumas propostas impõem as restrições de compatibilidade impedindo a formação de grupos que violem os limites definidos durante a busca. Assim, é possível obter maior independência da estratégia em relação ao conjunto de medidas adotado, além de permitir concentrar a busca na tarefa de minimizar o número de tubos necessários.

O BPP é um problema de agrupamento largamente estudado na literatura. Em sua definição clássica, um conjunto finito de objetos unidimensionais O , deve ser embalado em um número de caixas $N \leq O$, com capacidade máxima C , de maneira que N seja minimizado (FALKENAUER, 1994). O BPP e o problema da PCR *multiplex* se aproximam, principalmente, quando se considera para a PCR *multiplex* o cenário em que cada amplificação possui um par de *primers* pré-selecionado. Essa associação foi feita de maneira limitada anteriormente por Milewski et al. (2017), considerando um número fixo de tubos e aplicando variações na composição dos grupos para otimizar um único critério de compatibilidade. Essas limitações podem ser superadas através de uma adaptação com

maior potencial de otimização. Para tanto, pode-se considerar que os *primers* assumem o papel dos itens, os tubos figuram como embalagens ou caixas a serem minimizados e a restrição de capacidade passa a ser representada por conjuntos de restrições de interação entre os *primers*.

Na proposta deste trabalho, estratégias aplicadas anteriormente em trabalhos baseados no problema do empacotamento foram adaptadas à construção do modelo computacional do problema da PCR *multiplex*. Por essa razão, apresentamos a seguir uma breve revisão da literatura acerca do BPP, destacando os principais trabalhos que contextualizam e fundamentam essas estratégias.

2.1.1 Métodos heurísticos para o problema do empacotamento

Embora métodos exatos baseados em programação matemática¹ tenham sido altamente aprimorados para o problema do empacotamento, abordagens heurísticas e meta-heurísticas são fortemente encorajadas devido à sua classificação como um problema NP-completo (GARY; JOHNSON, 1979; FALKENAUER, 1994). Dentre essas abordagens, o algoritmo denominado *Genetic Grouping Algorithm* (GGA), proposto por Falkenauer (1994), é uma das meta-heurísticas mais exploradas para resolver o BPP. Resultados promissores foram alcançados quando a estrutura do algoritmo foi aprimorada por meio da adição de um processo de otimização local baseado no critério de dominância introduzido por Martello e Toth (1990), que deu origem ao denominado *Hybrid Grouping Genetic Algorithm* (HGGA) (FALKENAUER, 1996). Tais resultados são usados como base comparativa em diversos trabalhos da literatura relacionada, os quais são frequentemente desenvolvidos a partir de adaptações na proposta original de Falkenauer.

A substituição do AG por outros algoritmos evolutivos adaptados à codificação do GGA foi proposta nos algoritmos denominados *Grouping Particle Swarm Optimization* (GPSO) (KASHAN; KASHAN; KARIMIYAN, 2013) e *Grouping Evolution Strategies* (GES) (KASHAN; AKBARI; OSTADI, 2015), que obtiveram alguns resultados superiores àqueles alcançados no trabalho de Falkenauer. No algoritmo GGA-CGT (QUIROZ-CASTELLANOS et al., 2015) heurísticas eficientes e operadores altamente especializados no BPP foram introduzidos na estrutura básica do GGA, o que resultou em excelentes resultados, atribuídos principalmente ao controle de transmissão dos melhores genes durante o processo evolutivo. O trabalho descrito em (OZCAN et al., 2016) implementou um GGA de alto desempenho computacional utilizando a unidade de processamento gráfico (GPU), denominado GGA-CUDA, com o intuito de promover maior exploração do espaço de busca diante da maior eficiência computacional.

Outras duas abordagens heurísticas são frequentemente mencionadas por trabalhos

¹ Para uma revisão acerca de métodos exatos aplicados ao BPP, ver (DELORME; IORI; MARTELLO, 2016).

relacionados ao BPP devido à alta qualidade dos resultados alcançados. Na primeira delas, o algoritmo denominado HI_BP (ALVIM et al., 2004) apresenta uma heurística complexa que combina o método exato *branch-and-bound* e um processo de otimização baseado no algoritmo *Tabu Search*, além de técnicas previamente sugeridas em outros estudos, como um modelo de redução do espaço de busca e estratégias de limites superiores e inferiores. A segunda abordagem deriva da heurística *Perturbation-MBS* (FLESZAR; HINDI, 2002), que obtém novas soluções por meio da perturbação da solução atual utilizando a meta-heurística *Variable Neighborhood Search* (VNS) e estratégias com limites inferiores. Fleszar e Charalambous (2011) introduziram a este algoritmo um novo conceito, denominado “*sufficient average weighth*” (SAW), para controlar o peso médio dos itens empacotados em cada caixa, dando origem ao algoritmo *Perturbation-SAWMBS*, que superou os resultados de algoritmos considerados como o estado da arte para o BPP, dentre eles o HGGA e o HI_BP.

Na sequência, são abordadas as diferentes codificações utilizadas em meta-heurísticas desenvolvidas para solucionar o BPP, a fim de contextualizar a codificação adotada neste trabalho.

Codificações para o BPP

O processo de codificação é uma das etapas mais relevantes no desenvolvimento de uma meta-heurística evolutiva aplicada a um problema de otimização. Através dela, define-se de que maneira serão representadas as soluções do problema, denominadas como cromossomos ou indivíduos. Cada cromossomo é geralmente constituído por um vetor de símbolos, também chamados de genes, sobre os quais agem os operadores de busca. A fim de avaliar as soluções obtidas ao longo do processo, deve-se estabelecer uma função objetivo responsável por direcionar a evolução da busca determinando o valor de aptidão de cada indivíduo.

A codificação denominada *Number Encoding* (NE) pode ser considerada a representação mais direta para problemas de agrupamento, na qual um gene e um item são associados por meio de suas posições e o valor do gene indica o grupo ao qual o item pertence. Dessa forma, o cromossomo **32121** poderia codificar uma solução para um problema com cinco itens, por exemplo, onde o primeiro item pertence à caixa **3**, o segundo pertence à caixa **2**, o terceiro pertence à caixa **1** e assim por diante. *Grouping Encoding* (GE) define a codificação em que os itens alocados no mesmo grupo são mapeados por genes vizinhos no cromossomo, e a inserção de marcadores especiais delimita os subconjuntos, portanto, o indivíduo **(1)(2,4)(3,5)** poderia representar a solução codificada no exemplo anterior. Nesse caso, os valores dos genes indicam qual é o item mapeado de acordo com a ordenação de referência.

Uma codificação especial foi introduzida por Falkenauer no GGA com o objetivo

de mapear nos genes as unidades mais relevantes do problema, os grupos. Dessa forma, a ideia é proporcionar uma busca mais eficiente, por meio da meta-heurística, ao optar por uma estrutura mais representativa de codificação, superando, assim, o desempenho de outras codificações baseadas somente nos objetos. Para esse propósito, a codificação NE é incrementada com uma parcela adicional responsável por representar cada grupo em um gene. Os operadores de busca são específicos para essa forma de representação do problema do empacotamento e agem somente sobre a parcela de grupos, que pode variar de tamanho no decorrer da otimização. Assim, o exemplo descrito anteriormente poderia ser representado pelo indivíduo **32121:231**, no qual a subsequência final “**231**” indica quais os grupos compõem a solução e a primeira parcela tem o intuito exclusivo de indicar em qual grupo cada item está alocado. O problema da redundância, isto é, a possibilidade de cromossomos distintos codificarem a mesma solução, foi ressaltado por Falkenauer como uma consequência crítica de outras codificações. No entanto, a representação usada no GGA também sofre desse problema, além de exigir que o algoritmo seja capaz de manipular cromossomos de tamanhos variáveis, como mencionado em (ÜLKER; KORKMAZ; ÖZCAN, 2008; MOHAMADI, 2010).

Ülker, Korkmaz e Özcan (2008) investigaram a adaptação de um algoritmo genético baseado na representação denominada como *Linear Linkage Encoding* (LLE) para solucionar o BPP. A utilização da LLE tem o objetivo de contornar o problema de redundância de outras codificações, eliminando a simetria indesejada do espaço de busca, uma vez que cada gene representa um objeto e o seu respectivo valor indica uma ligação com o próximo objeto do grupo. Os experimentos computacionais realizados mostraram a viabilidade da proposta desde que adotados operadores evolutivos específicos para a codificação. No entanto, os resultados indicaram a necessidade de melhorias na estratégia quando comparados ao HGGA de Falkenauer.

Boa parte das diferentes representações sugeridas para o BPP exige profundas e complexas modificações no funcionamento das meta-heurísticas empregadas, dando origem a algoritmos altamente especializados e dificultando a generalização da técnica. Outra característica importante associada a essas estratégias é a dependência da adoção de heurísticas construtivas simples para corrigir inconsistências causadas pelos operadores durante a busca (FALKENAUER, 1994; QUIROZ-CASTELLANOS et al., 2015; KASHAN; KASHAN; KARIMIYAN, 2013; KASHAN; AKBARI; OSTADI, 2015). Dessa maneira, são reinseridos nas soluções os itens que eventualmente deixam de integrá-las após uma recombinação, por exemplo. Algoritmos para esse propósito são geralmente baseados em heurísticas como a *First Fit* (FF) e *Best Fit* (BF) que, através de uma abordagem gananciosa, encontram soluções com uma garantia mínima de qualidade em tempo polinomial (JOHNSON, 1973; JOHNSON, 1974).

O algoritmo FF provê o agrupamento (empacotamento) sequencial de uma lista de itens, buscando atribuir cada um deles à primeira caixa factível de acordo com as restrições

do problema. No contexto do BPP, uma caixa factível é aquela em que o agrupamento do item em questão não excede a capacidade máxima determinada. Já a heurística BF, realiza o agrupamento procurando sempre a caixa que resulte na menor sobra de capacidade, exigindo a adição de uma etapa de ordenação baseada nas ocupações das caixas antes de alocar cada item. O número de grupos, ou caixas, obtidos por essas heurísticas estão diretamente relacionados à ordenação dos itens sobre a qual a heurística opera. Portanto, embora sejam computacionalmente eficientes, estão sujeitas muitas vezes a soluções não ótimas, ou localmente ótimas, a depender do sequenciamento previamente estabelecido.

Por outro lado, meta-heurísticas adaptadas para lidar com o BPP permitem uma busca global efetiva que aumenta a chance de obter soluções ótimas ou quase ótimas, porém, geralmente com um custo computacional maior. Nessa perspectiva, algoritmos híbridos - desenvolvidos para busca de permutações ótimas de itens através de uma meta-heurística, com subsequente decodificação por meio de uma heurística construtiva - foram propostos no trabalho de Reeves (1996) e no trabalho de Gourgand, Grangeon e Klement (2014). Reeves demonstrou a viabilidade da estratégia através de experimentos computacionais que obtiveram resultados competitivos e introduziu um processo de redução do espaço de busca nessa abordagem, o qual será discutido em detalhes no próximo capítulo. Gourgand et al. destacaram as vantagens deste modelo de solução para o BPP, dentre elas, a garantia de factibilidade de qualquer solução mapeada em razão do processo de decodificação aplicado, a simplicidade das técnicas envolvidas e a facilidade de generalização dos algoritmos para a solução de outros problemas. Características estas que permitem ainda o desacoplamento dos operadores evolutivos em relação às propriedades do problema. Os autores propõem um *framework* para a implementação do modelo, sugerindo uma possível utilização de qualquer método de computação evolutiva. Para os experimentos computacionais, algoritmos de busca local foram combinados com diferentes heurísticas construtivas. Os resultados mostraram que, apesar do potencial teórico da abordagem, não houve garantia de eficiência em relação aos demais algoritmos considerados na comparação.

Neste trabalho, foi desenvolvido um modelo de representação semelhante ao descrito anteriormente e a estratégia foi aplicada em ambos os problemas: o problema do empacotamento e a otimização da PCR *multiplex*. A abordagem híbrida desenvolvida pretende explorar de maneira eficiente as características de cada técnica de solução utilizada na construção do modelo. Além do algoritmo genético e da heurística construtiva, foi introduzido um processo de restrição do espaço de busca diferente do método de redução apresentado por Reeves, capaz de aprimorar o comportamento e o desempenho do modelo híbrido. Nos próximos capítulos são descritos os detalhes de implementação dos algoritmos com as devidas adaptações para cada problema tratado.

3 UMA ABORDAGEM COMPUTACIONAL PARA PROBLEMAS DE AGRUPAMENTO

Neste capítulo e no seguinte, são apresentados os algoritmos desenvolvidos com base na abordagem computacional proposta neste trabalho, para problemas de agrupamento. O modelo é constituído basicamente pela união de um algoritmo genético adaptado a problemas de permutação e a heurística construtiva *First Fit*. Sua construção foi direcionada especificamente pela associação do problema do empacotamento ao problema de agrupamento de *primers* no projeto da PCR *multiplex*, ambos baseados em otimização combinatória com fase de agrupamento, no entanto, com restrições distintas.

De maneira geral, o objetivo em ambos os problemas abordados é obter o agrupamento ótimo dos elementos, sejam eles os itens (BPP) ou *primers* (PCR *multiplex*). Um agrupamento ótimo é aquele que resulta no menor número de grupos possível, considerando as propriedades dos elementos disponíveis e as restrições de agrupamento impostas. Como mencionado anteriormente, diferentes sequenciamentos de itens ou *primers* podem exigir diferentes quantidades de caixas ou tubos, respectivamente, dado que o agrupamento é desempenhado por uma heurística construtiva. Atendo-se a essa perspectiva, o algoritmo genético é responsável por buscar a permutação ótima dos elementos, isto é, aquela que forneça o menor número de grupos após a ação do algoritmo FF, que cumpre a tarefa de decodificar as soluções mapeadas pelo AG, impondo as restrições definidas e, assim, garantindo a viabilidade de qualquer sequenciamento.

3.1 ADAPTAÇÃO DO MODELO AO PROBLEMA DO EMPACOTAMENTO

Nesta seção, são descritos os algoritmos implementados como parte do estudo desenvolvido neste trabalho a fim de solucionar o problema do empacotamento, sua definição foi apresentada anteriormente na seção 2.1.

O primeiro algoritmo, denominado Algoritmo Genético Baseado em Heurística (AGBH), possui estrutura similar ao “GA + FF” sugerido por Reeves (1996), sendo composto por um AG adaptado a problemas de permutação que inclui os operadores clássicos de seleção, recombinação e mutação, além do operador de elitismo e da função objetivo, responsável por avaliar o agrupamento obtido pelo algoritmo *First Fit*.

Nos demais algoritmos, um processo de restrição do espaço de busca foi incorporado ao AG, adaptando sua estrutura por meio da adição de uma codificação especial e novos operadores. No primeiro caso, o algoritmo denominado Algoritmo Genético Baseado em Heurística com Bloqueio (AGBHcB) possui uma abordagem determinística, que não permite a reversão do processo de restrição do tamanho do problema. No segundo, intitulado por Algoritmo Genético Baseado em Heurística com Bloqueio Dinâmico (AGBHcBD), foi introduzido um novo operador baseado na codificação proposta, a fim de conferir maior

flexibilidade ao processo. A seguir, é apresentada uma discussão acerca do método de restrição do espaço de busca e, posteriormente, são descritos os detalhes de implementação dos algoritmos desenvolvidos.

3.1.1 Restrição do espaço de busca

Assim como no trabalho de Reeves, propomos uma intervenção no espaço de busca sempre que uma caixa atinge sua capacidade máxima, ou seja, a soma dos pesos dos itens que nela estão inseridos é igual a C . Neste momento, assume-se que uma fração do problema foi resolvida e os itens em questão são retirados do processo de busca. Em (REEVES, 1996), a **redução** é aplicada por meio da **deleção** dos itens em nível populacional. Portanto, ao identificar um agrupamento que atinge a capacidade máxima da caixa, o algoritmo deleta os itens respectivamente agrupados de todos os indivíduos da população.

Embora uma fração do problema seja solucionada ao atingir a capacidade máxima de uma caixa, a prefixação parcial de agrupamentos pode impedir a obtenção da solução ótima global. Supondo um problema em que a capacidade máxima das caixas seja igual a 10 e seis itens devam ser empacotados, tendo seus respectivos pesos iguais a 2, 4, 4, 6, 6 e 8, o mínimo de caixas necessárias é igual a 3. No entanto, caso os três primeiros itens sejam agrupados e retirados da busca, o número de caixas encontradas será obrigatoriamente 4. Reeves denomina esse problema como “garantia de subotimalidade” e, para ignorá-lo, assume que sua ocorrência é muito rara na prática.

Outra consequência crítica associada ao processo de redução sugerido por Reeves deve ser destacada. A diversidade da população pode ser duramente sacrificada ao replicar a mesma solução parcial em todos os indivíduos da população, visto que estes poderiam conduzir os mesmos itens deletados a agrupamentos distintos e, portanto, soluções diferentes, o que aumentaria as chances de evitar soluções localmente ótimas. Em síntese, ainda que a garantia de subotimalidade se apresente como um problema raro na prática, a qualidade da busca pode ser prejudicada pela deleção iterativa de itens durante o processo de busca do AG.

Nos algoritmos propostos, introduzimos a **restrição** de variáveis do problema através do conceito de **bloqueio** dos itens ao invés do método de deleção. Ao contrário do contexto de intervenção populacional associado à deleção, o processo de bloqueio atua em nível do indivíduo, bloqueando itens que formem agrupamentos perfeitos somente nos cromossomos em que tais agrupamentos são obtidos. Nesse sentido, apresentamos uma exploração do processo que permita ao algoritmo a renúncia de possíveis reduções prematuras, assegurando, dessa forma, que a possibilidade de encontrar a solução ótima global não seja anulada durante o processo de busca. Além disso, a aplicação da restrição em nível de indivíduos, apesar de mais complexa em termos de programação, evita a perda

homogênea da diversidade populacional.

No algoritmo AGBHcB, limitamos a abordagem apresentada a ação de bloquear os itens, suprimindo a possibilidade de desbloqueá-los posteriormente. Esse protótipo, apesar de conceitualmente menos restritivo em relação à proposta de Reeves, também não permite que uma redução de variáveis no indivíduo seja desfeita. Assim, pretende-se verificar se, de fato, a garantia de subotimalidade consiste em um problema raro na prática e, dessa forma, justificar empiricamente o aprimoramento proposto no algoritmo AGBHcBD.

A perspectiva de renúncia de uma restrição do espaço de busca foi incluída no algoritmo AGBHcBD através do operador de desbloqueio de itens. Tal inclusão, permite ao algoritmo contornar soluções localmente ótimas por meio da reversão de bloqueios eventualmente prematuros, sendo a principal distinção entre os algoritmos AGBHcB e AGBHcBD. Ambos não contam com o operador de recombinação na estrutura do AG, sobre o qual se discute a seguir.

A redução do espaço de busca em nível populacional, proposta por Reeves, garante ao algoritmo a conveniência de lidar sempre com indivíduos de mesmo tamanho e com os mesmos itens disponíveis após as deleções. Por outro lado, o emprego de uma restrição em nível individual, proposto neste trabalho, permite que os indivíduos evoluam de maneira independente ao longo da busca, constituindo diferentes agrupamentos e, conseqüentemente, apresentando distinções acerca dos conjuntos de itens bloqueados. Por esse motivo, a aplicação do operador de recombinação entre indivíduos é pelo menos não trivial. Tal complexidade inviabiliza a aplicação direta de um operador de recombinação conhecido que atenda ao intuito de recombinar os indivíduos sem interferir nas restrições do espaço de busca já efetivadas.

3.1.2 Algoritmo Genético Baseado em Heurística (AGBH)

Codificação:

Os cromossomos são formados por permutações de inteiros de 1 a $|S|$ que mapeiam cada item a ser empacotado, sendo S o conjunto de todos os itens. A heurística construtiva, detalhada posteriormente, é responsável por decodificar as soluções representadas.

Operador de seleção:

A seleção por ranking é utilizada para determinar quais os indivíduos serão selecionados e submetidos aos próximos operadores, dando origem a nova população. Os conceitos e detalhes de implementação deste operador são amplamente difundidos na literatura e têm por objetivo priorizar os indivíduos bem ranqueados durante o processo de seleção.

Operador de recombinação:

A população selecionada no processo anterior é submetida ao operador *Partially Mapped Crossover* (PMX), introduzido por Goldberg, Lingle et al. (1985) e um dos mais utilizados para codificações baseadas em problemas de permutação (KUMAR; GOPAL; KUMAR, 2013). De maneira geral, o processo provê a troca de uma subsequência aleatoriamente delimitada entre dois cromossomos, preservando ao máximo a ordenação dos demais alelos. Para cada dois cromossomos aleatoriamente selecionados, a recombinação é aplicada com uma probabilidade p_c , definida como parâmetro do AG. Pares de indivíduos selecionados e não submetidos à recombinação são clonados diretamente para a população seguinte.

Operador de mutação por swap:

Com probabilidade p_m , cada indivíduo da população obtida anteriormente sofre uma perturbação através de um *swap* entre dois genes aleatoriamente definidos. A aplicação do operador tem por objetivo introduzir novas informações na população e permitir que o algoritmo escape de uma possível estagnação das soluções.

Decodificação:

A aplicação da função objetivo em um cromossomo da população é precedida pela fase de agrupamento, desempenhada pelo algoritmo *First Fit*. Com base na permutação de itens mapeada pelo indivíduo, o algoritmo FF agrupa (ou empacota) cada item na primeira caixa possível, ou seja, respeitando a ordem de criação das caixas, a primeira em que a capacidade disponível seja maior ou igual ao peso do item em questão. Ao fim do processo, o total de caixas obtido e a capacidade utilizada em cada uma delas são repassados ao cálculo da função objetivo.

Função objetivo:

A função objetivo adotada é dada pela expressão definida na equação 3.1, amplamente aplicada ao BPP e introduzida por Falkenauer e Delchambre (1992). Esta função objetivo é capaz de distinguir a eficiência de duas soluções com o mesmo número de caixas, ao contrário de uma avaliação mais simples e intuitiva, na qual, o número de caixas obtido representa diretamente a aptidão do indivíduo. Além disso, o refinamento propiciado pela medida facilita a eliminação de caixas durante o processo evolutivo, dada a preferência por soluções que apresentem maior concentração do volume total dos itens e, portanto, mais caixas com níveis menores de utilização, que por sua vez, exigem menos esforço para serem esvaziadas. Para esse propósito, a busca é dirigida pela maximização da função

representada em 3.1, onde F_i representa a soma dos pesos dos itens na caixa i , C é a capacidade das caixas e M o número de caixas obtidas pela solução avaliada.

$$f_{BPP} = \frac{\sum_{i=1}^M (F_i/C)^2}{M} \quad (3.1)$$

3.1.3 Algoritmo Genético Baseado em Heurística com Bloqueio (AGBHcB)

São descritas abaixo as modificações feitas na estrutura do algoritmo AGBH para a aplicação do método de restrição do espaço de busca discutido anteriormente.

Codificação com bloqueio:

Cada cromossomo, com a mesma composição descrita no algoritmo anterior, é associado a outro vetor de inteiros de mesmo tamanho, denominado “vetor de bloqueio”, responsável por controlar quais índices do vetor solução encontram-se bloqueados. Um índice bloqueado indica que o item mapeado por ele pertence a uma caixa cheia, ou seja, a soma dos pesos dos itens agrupados é igual à capacidade máxima C . Através do processo de bloqueio, o espaço dimensional de busca sofre uma restrição, uma vez que esses itens passam a ser desconsiderados pelo operador de mutação e pelo processo de decodificação, sendo incorporados apenas no cálculo da função objetivo.

Inicialmente, todos os itens estão desbloqueados e recebem o valor 0 em sua respectiva posição no vetor de bloqueio. Durante o processo evolutivo, o operador de bloqueio é responsável por atualizar os valores do vetor.

Operador de bloqueio:

Quando um agrupamento de itens atinge a capacidade máxima da caixa, o vetor de bloqueio é atualizado: as posições associadas aos respectivos itens são definidas com o número da caixa cheia a qual pertencem. Portanto, o vetor de bloqueio também indica como estão agrupados os itens que encontram-se bloqueados.

Decodificação com bloqueio:

Assim como no algoritmo AGBH, o algoritmo FF opera a decodificação da permutação mapeada no cromossomo, agrupando os itens de acordo com a capacidade máxima das caixas para a posterior aplicação da função objetivo descrita em 3.1. Entretanto, para que o algoritmo seja capaz de lidar com o processo de restrição do espaço de busca, novas instruções foram incorporadas ao funcionamento da heurística construtiva.

Inicialmente, uma busca no vetor de bloqueio retorna o maior valor existente, que indica o total de caixas completamente preenchidas, ou seja, o número de grupos formados por itens bloqueados no respectivo (i -ésimo) cromossomo, denominado por Gb_i . Portanto, $Gb_i = 0$ na primeira iteração.

Respeitando eventuais restrições do espaço de busca já efetivadas, o algoritmo *First Fit* ignora os itens bloqueados durante o empacotamento, formando novos grupos somente com itens associados ao valor 0 no vetor de bloqueio. Além disso, para que as caixas cheias sejam contabilizadas na atual solução, o algoritmo FF numera a primeira caixa criada sempre em $Gb_i + 1$.

Operador de mutação por swap com bloqueio:

Após a seleção por ranking, descrita no algoritmo anterior, os indivíduos são submetidos ao operador de mutação por *swap* com probabilidade p_m . No entanto, apenas os índices desbloqueados em cada cromossomo, ou seja, aqueles associados ao valor 0 no vetor de bloqueio, são disponibilizados para uma possível troca.

A Figura 5 apresenta uma ilustração das codificações utilizadas, na qual, o n -ésimo item do vetor de referência é representado por i_n na permutação mapeada pelo indivíduo. No exemplo exposto, deve-se notar que, após a decodificação realizada pela heurística *First Fit*, as caixas 1 e 2 atingiram a capacidade máxima. Naturalmente, o algoritmo baseado na codificação sem bloqueio (a) continua operando sobre toda a sequência mapeada. Por outro lado, nos algoritmos baseados na codificação com bloqueio (b), a atualização do vetor de bloqueio identifica os itens que estão inseridos em caixas cheias, os quais passam a integrar uma solução parcial do problema, são eles: $i_4 = 7$, $i_6 = 6$, $i_8 = 3$, $i_3 = 2$ e $i_7 = 2$. Neste caso, o algoritmo passa a operar a busca somente sobre os itens desbloqueados, isto é, a permutação contida na subsequência: $[i_5, i_1, i_2] = [3, 9, 1]$.

3.1.4 Algoritmo Genético Baseado em Heurística com Bloqueio Dinâmico (AGBHcBD)

Os algoritmos AGBHcBD e AGBHcB compartilham estruturas idênticas, com exceção do operador adicional descrito a seguir, que confere dinamismo à restrição do espaço de busca, permitindo que uma eventual restrição subótima seja desfeita.

Operador de desbloqueio:

Com probabilidade p_d , cada cromossomo da população com no mínimo uma caixa cheia é submetido ao operador de desbloqueio. Uma das caixas cheias existentes, selecionada aleatoriamente, tem seus itens desbloqueados e reinseridos no espaço de busca. Para

Figura 5 – Representação das codificações utilizadas

Itens (pesos): [9, 1, 2, 7, 3, 6, 2, 3] Capacidade máx. = 10

Ex. de permutação = 7 6 3 2 3 9 2 1

Decodificação: **caixa1** = {7;3}, **caixa2** = {6;2;2}, **caixa3** = {3;1},
caixa4 = {9}a) Codificação sem bloqueio
(AGBH)

i_4	i_6	i_8	i_3	i_5	i_1	i_7	i_2
-------	-------	-------	-------	-------	-------	-------	-------

b) Codificação com bloqueio
(AGBHcB, AGBHcBD)

i_4	i_6	i_8	i_3	i_5	i_1	i_7	i_2
1	2	1	2	0	0	2	0

Fonte: Elaborado pelo autor (2019).

isso, suas respectivas posições no vetor de bloqueio são definidas em 0. Deve-se notar que a aplicação do operador de desbloqueio não implica diretamente que a caixa cheia desbloqueada será desfeita. Porém, permite que os itens associados à caixa desbloqueada sejam manipulados pelo operador de mutação e pelo processo de decodificação.

3.1.5 Implementação modificada da heurística *First Fit*

Uma das perspectivas de aprimoramento do modelo apresentado pode ser associada a avanços na decodificação provida pela heurística construtiva, principalmente com o intuito de diminuir a possibilidade de agrupamentos localmente ótimos. Dessa forma, nos protótipos adaptados ao BPP, a decodificação executada pelo algoritmo *First Fit* foi modificada com o objetivo de evitar a formação de caixas com um conjunto de itens de pesos menores. Isso é interessante no sentido de que esses itens podem ser importantes para ajustar caixas com pouca capacidade ociosa.

A heurística *First Fit* modificada é construída em duas etapas: (i) em primeiro lugar, os itens maiores que a metade da capacidade de armazenamento são pré-empacotados em caixas individuais; (ii) em seguida, a clássica heurística *First Fit* é aplicada para finalizar a composição dos grupos. O empacotamento inicial dos itens maiores, que devido à restrição de capacidade não poderiam ser alocados juntos, diminui as possibilidades de uma formação desnecessária de caixas compostas somente por itens de menor peso, provocada por um eventual sequenciamento desfavorável. Essa abordagem foi usada com sucesso em (QUIROZ-CASTELLANOS et al., 2015) para gerar a população inicial.

4 ADAPTAÇÃO DO MODELO AO PROBLEMA DA PCR *MULTIPLEX*

O problema de agrupamento inerente à PCR *multiplex* é abordado neste trabalho em dois cenários distintos. No primeiro, cada sequência alvo de amplificação possui um par de *primers* pré-selecionado, assim, todos os pares de *primers* disponíveis devem ser agrupados no menor número de grupos possível, analogamente ao BPP. No segundo cenário, cada região alvo possui dois ou mais pares de *primers* candidatos à amplificação. Neste último caso, o processo de otimização também deve ser responsável pela escolha do *primer* candidato, de preferência concomitantemente com o objetivo central de minimizar o número de grupos.

4.1 ABSTRAÇÃO DA REAÇÃO DE PCR *MULTIPLEX*

O agrupamento das reações de PCR é limitado pelos critérios de compatibilidade entre as propriedades dos *primers*, o que decorre do intuito de potencializar as chances de sucesso da reação de PCR *multiplex*. Diferentemente do BPP, no qual os agrupamentos são limitados pela capacidade máxima de cada caixa. Dessa forma, as reações, ou amplificações, são abstraídas aos pares de *primers* utilizados.

As principais medidas de compatibilidade frequentemente consideradas em outros trabalhos concentram-se basicamente em evitar três ocorrências indesejadas durante a reação, a saber:

- Primer vs. Primer
 - Dimerização: pareamento entre dois *primers* quaisquer e a consequente perda do produto de amplificação.
- Primer vs. DNA-molde
 - Anelamento inespecífico: um *primer* se anelando em uma região não alvo da sequência genômica, o que pode levar à distorção dos resultados e à perda de produto.
- Produto vs. Produto
 - Dificuldade ou inviabilidade de interpretação dos resultados devido à diferença insuficiente entre o comprimento de dois ou mais produtos.

De acordo com as referências citadas nos capítulos iniciais e visando à consideração de cada uma dessas possibilidades, os protótipos desenvolvidos se apoiam nas seguintes diretrizes:

- Impedir o agrupamento de *primers* capazes de formar dímeros estáveis, que excedam um limite mínimo definido de ΔG para o pior caso de anelamento;

- Impedir o agrupamento de *primers* com T_m 's que excedam um limite máximo de amplitude, visando à estabilidade da hibridização dos *primers* em relação aos seus alvos durante as variações cíclicas de temperatura;
- Impedir o agrupamento de *primers* que produzam amplicons com uma diferença entre seus comprimentos menor do que o limite definido.

Considerando as propriedades dos *primers* e os critérios de compatibilidade abordados neste trabalho, foi desenvolvida uma função booleana responsável por determinar se dois pares de *primers* podem ou não ser agrupados e, portanto, se duas amplificações quaisquer podem ocorrer no mesmo tubo de PCR *multiplex* com base no modelo de abstração da reação. Para tanto, a função de compatibilidade se baseia nos limiares parametrizados para cada uma das três medidas.

A adaptação de qualquer um dos algoritmos apresentados para o problema de empacotamento seria construída de maneira similar para o problema da PCR *multiplex*. Optou-se, porém, pela descrição da adaptação do algoritmo mais completo (AGBHcBD), visto que seus componentes englobam os outros protótipos.

4.2 RESTRIÇÃO DO ESPAÇO DE BUSCA (PCR *MULTIPLEX*)

Dois fatores devem ser considerados ao discutir a adaptação do processo de restrição do espaço de busca apresentado anteriormente (seção 3.1.1) para o problema da PCR *multiplex*.

I - Assim como descrito para o BPP, a restrição do espaço de busca é promovida através do bloqueio de índices já agrupados, reduzindo o número total de possíveis combinações. No entanto, a principal diferença entre o agrupamento de itens (BPP) e o agrupamento de reações (PCR *multiplex*) consiste na ausência de um valor prático de capacidade máxima de reações por tubo, o que permitiria um critério para identificar uma solução parcial do problema. Possivelmente, a não consideração desse valor em outros trabalhos se deve à tendência natural de que a incompatibilidade se apresente como principal fator limitante na tentativa de obter agrupamentos de muitas reações. Devido a essa característica, a dinâmica do processo de bloqueio foi adaptada para que a definição de uma solução parcial do problema seja baseada na priorização dos agrupamentos que contém o maior número de reações e, portanto, contribuem em maior grau para a minimização do número de tubos necessários.

II - Uma vez que o número ótimo de reações por tubo não é definido previamente, a ação de bloquear os índices de um agrupamento não impede que novas reações sejam adicionadas posteriormente, garantindo apenas que aquelas já agrupadas permaneçam

unidas. Do contrário, o processo de restrição do espaço de busca poderia forçar o aumento desnecessário do número de tubos.

4.3 PCR *MULTIPLEX* COM *PRIMERS* PRÉ-SELECIONADOS

Como descrito em (LIPSON, 2003), o problema de agrupamento de *primers* pré-selecionados para a PCR *multiplex* pode ser definido formalmente da seguinte maneira: dado que S é o conjunto de sequências alvo; $pp(s)$, o par de *primers* pré-selecionado para amplificar cada sequência $s \in S$; D , a sequência genômica; e $B_D : (\bigcup_{s \in S} pp(s)) \times (\bigcup_{s \in S} pp(s)) \rightarrow \{0, 1\}$, a função booleana de compatibilidade entre os pares de *primers*, o objetivo é obter um agrupamento $\phi : S \rightarrow R$, sendo R o número de reações de PCR *multiplex*, tal que:

- Se $\phi(s_k) = \phi(s_l)$ então $B_D(pp(s_k), pp(s_l))$ é satisfeito.
- O número de tubos (reações) de PCR *multiplex* é mínimo.

4.3.1 AGBHcBD adaptado à PCR *multiplex* com *primers* pré-selecionados

A seguir, são descritas as adaptações de implementação promovidas na estrutura do algoritmo AGBHcBD para aplicação no problema da PCR *multiplex* com *primers* pré-selecionados. A codificação com bloqueio e os operadores de seleção e mutação por *swap* com bloqueio não sofreram alterações em relação às suas implementações descritas para o problema do empacotamento e, por essa razão, foram suprimidos nesta seção.

A exemplo do vetor de itens utilizado como referência para o mapeamento das soluções do BPP, as reações são previamente numeradas de 1 até $|S|$ e mapeadas por esses valores nos cromossomos do AG.

Decodificação com bloqueio (PCR multiplex com primers pré-selecionados):

Semelhante ao mesmo processo descrito para o BPP, o algoritmo *First Fit* é responsável por agrupar as reações mapeadas por uma permutação e, assim, definir a composição de cada tubo de PCR *multiplex*. Inicialmente, uma varredura no vetor de bloqueio retorna o maior inteiro encontrado, que indica o total de grupos com índices bloqueados Gb_i . Em seguida, o algoritmo FF executa o agrupamento sobre os índices desbloqueados, isto é, aqueles associados ao valor 0 no vetor de bloqueio. Ressalta-se que neste caso, como discutido anteriormente na seção 4.2, um índice desbloqueado pode ser agrupado a índices bloqueados, desde que haja compatibilidade mútua entre tais reações mapeadas.

A compatibilidade entre duas reações a e b ocorre quando a função de compatibilidade B_D indica que os *primers* podem ser agrupados de acordo com os limiares

estabelecidos, isto é, quando $B_D(pp(a), pp(b)) = 1$. No entanto, para inserir um par de *primers* $pp(k)$ em um grupo G , $pp(k)$ deve ser compatível com todos os pares de *primers* eventualmente já agrupados em G , portanto, $B_D(pp(k), \cup_{n \in G} pp(n)) = 1$.

Operador de bloqueio (PCR multiplex com primers pré-selecionados):

Uma vez que os tubos não são limitados por uma capacidade máxima de reações, o critério de bloqueio foi estabelecido por uma meta progressiva de grupos a terem seus índices bloqueados, definida por Gb'_i . Em cada cromossomo, o valor de Gb'_i é inicializado em 1 na primeira iteração e, para garantir a progressão da restrição do espaço de busca, $Gb'_i = Gb_i + 1$ em cada nova iteração. Portanto, na primeira iteração $Gb_i = 0$ e $Gb'_i = 1$.

Eventualmente, pode ocorrer de, após a etapa de decodificação, o número de grupos obtido ser inferior à meta de grupos bloqueados Gb'_i . Nesse caso, a restrição do espaço de busca é reinicializada, definindo-se novamente $Gb'_i = 1$ e desbloqueando eventuais índices bloqueados anteriormente.

Assim, a atualização do vetor de bloqueio tem seu comportamento condicionado ao valor de Gb'_i e o processo pode ser resumido pelas seguintes instruções:

- Os grupos formados pela execução do algoritmo *First Fit* são ordenados, em ordem decrescente, de acordo com a quantidade de reações agrupadas.
- Se $Gb'_i >$ total de grupos obtido:
 - Somente os elementos do primeiro grupo e, portanto, o maior deles, são bloqueados. Para isso, seus respectivos índices são definidos em 1 no vetor de bloqueio e os demais definidos em 0.
- Se $Gb'_i \leq$ total de grupos obtido:
 - Os primeiros Gb'_i grupos tem seus elementos bloqueados: as reações agrupadas em cada um deles tem seus respectivos índices no vetor de bloqueio atualizados com os valores de 1 a Gb'_i , indicando a qual dos agrupamentos com bloqueio cada reação pertence.

Função objetivo:

Analogamente à função objetivo aplicada ao problema do empacotamento (seção 3.1.2), buscou-se desenvolver uma medida refinada de avaliação das soluções para o agrupamento de *primers* na PCR *multiplex*, evitando que a otimização seja direcionada somente pelo número de tubos obtido em cada solução. Em caso de empate, o cálculo descrito abaixo prioriza as soluções com uma distribuição mais desbalanceada da quantidade

de reações por tubo, ou seja, aquelas em que a probabilidade de esvaziamento de um ou mais tubos seja maior devido à existência de tubos com baixo número de reações. Para tanto, o algoritmo busca a minimização da função abaixo, onde N_t é o número de tubos (grupos) obtido e $Np(i)$ o total de possíveis pareamentos entre dois *primers* no i -ésimo tubo, calculado através do coeficiente binominal, no qual, P_i é o número de *primers* no i -ésimo tubo :

$$f_{PCR_m} = N_t + \frac{1}{\sum_{i=1}^{N_t} Np(i)}, \quad \text{onde } Np(i) = \binom{P_i}{2} \quad (4.1)$$

A seguir, é apresentado um exemplo a fim de ilustrar o comportamento da função objetivo proposta. Para tanto, são supostos três indivíduos mapeando agrupamentos distintos de 6 reações (12 pares de *primers*) em três tubos. Para cada agrupamento, é discriminado o cálculo da função objetivo (f.o.):

- 1 - Tubo1 (2 reações), Tubo2 (2 reações), Tubo3 (2 reações), f.o.: $3 + \frac{1}{6+6+6} = 3.056$
- 2 - Tubo1 (3 reações), Tubo2 (2 reações), Tubo3 (1 reação), f.o.: $3 + \frac{1}{15+6+1} = 3.046$
- 3 - Tubo1 (4 reações), Tubo2 (1 reação), Tubo3 (1 reação), f.o.: $3 + \frac{1}{28+1+1} = 3.033$

No decorrer da busca, espera-se que o indivíduo 3 tenha maior probabilidade de obter um agrupamento capaz de compatibilizar as reações em menos de 3 tubos, em razão do potencial esvaziamento do Tubo2 ou do Tubo3. Nessa perspectiva, a solução representada pelo indivíduo 3 é melhor avaliada em relação às demais.

Experimentos computacionais foram realizados a fim de comparar a função proposta com a simples adoção do número de tubos obtido como aptidão dos indivíduos. Os resultados são apresentados no capítulo 5.

4.4 EXTENSÃO DO MODELO PARA PCR *MULTIPLEX* COM *PRIMERS* VARIÁVEIS

Neste capítulo, é descrita uma extensão do modelo computacional apresentado anteriormente, visando ao tratamento de problemas de agrupamento com múltipla escolha, especificamente o agrupamento ótimo de reações de PCR quando são fornecidas opções de pares de *primers* para cada amplificação. Neste caso, além buscar o agrupamento ótimo dos pares de *primers* de acordo com a abstração da reação de PCR *multiplex* apresentada na seção 4.1, o algoritmo deve lidar com a tarefa de selecionar um único par candidato por amplificação. Dessa forma, busca-se selecionar os candidatos que permitam reduzir o total de tubos necessários devido à sua compatibilidade em relação aos demais.

Os termos “seleção” e “agrupamento” são utilizados a seguir para distinguir duas tarefas: (i) a seleção de um dos pares de *primers* candidatos por amplificação; (ii) o agrupamento das reações de PCR. A etapa de seleção foi inserida na otimização por meio de adaptações no algoritmo genético que permitem uma busca simultânea pela seleção e o agrupamento ótimos. Dessa forma, pretende-se obter uma exploração eficiente de todo o espaço de busca, considerando as possibilidades combinatórias adicionais advindas da adição dos pares de *primers* candidatos.

Lipson (2003) define formalmente o problema de agrupamento de reações de PCR com pares de *primers* candidatos da seguinte forma: dado que S é o conjunto de sequências alvo, $PP(s) = (pp_1(s), pp_2(s), \dots, pp_{m_s}(s))$ o conjunto de pares de *primers* candidatos a amplificar cada sequência $s \in S$; D , a sequência genômica; e $B_D : (\bigcup_{s \in S} PP(s)) \times (\bigcup_{s \in S} PP(s)) \rightarrow \{0, 1\}$, a função booleana de compatibilidade entre os pares de *primers*, o objetivo é obter:

- um conjunto de representantes $pp(s) \in PP(s)_{s \in S}$,
- um agrupamento $\phi : S \rightarrow R$, sendo R o número de reações de PCR *multiplex*.

tal que:

- Se $\phi(s_k) = \phi(s_l)$ então $B_D(pp(s_k), pp(s_l))$ é satisfeito.
- O número de tubos (reações) de PCR *multiplex* é mínimo.

4.4.1 AGBHcBD adaptado à PCR *multiplex* com *primers* variáveis

A seguir, são descritas as adaptações de implementação promovidas na estrutura do algoritmo AGBHcBD para aplicação no problema da PCR *multiplex* com *primers* variáveis. A função objetivo 4.1 e os operadores de seleção e bloqueio não sofreram alterações em relação às suas implementações descritas para o problema da PCR *multiplex* com *primers* pré-selecionados (seção 4.3.1) e, por essa razão, não serão revistos.

Codificação com bloqueio (PCR multiplex com primers variáveis):

Neste modelo, a representação de uma solução é construída de forma a codificar conjuntamente as duas etapas do processo de otimização. Para tanto, o cromossomo é composto por duas sequências de inteiros com tamanho igual a $|S|$, sendo $|S|$ o número total de reações:

- Seleção: é representada na primeira parte, onde cada alelo i mapeia um valor k_s que varia de 1 a m_s ($s \in S$), indicando o par de *primers* $pp_{k_s}(s)$ selecionado para amplificar a sequência s .

- Agrupamento: mapeada na segunda metade do cromossomo, é composta pela permutação dos inteiros de 1 a $|S|$ que indica a ordem em que as reações serão agrupadas pela heurística construtiva.

O vetor de bloqueio, também de tamanho $|S|$, é responsável por controlar a restrição do espaço de busca bloqueando os índices de ambas as partes do cromossomo simultaneamente. Dessa forma, um índice bloqueado, $i_b \neq 0$ no vetor de bloqueio, estabelece que:

- Seleção: o par de *primers* selecionado no i -ésimo alelo está bloqueado e, portanto, não será alterado, isto é, não será disponibilizado para a ação do operador de mutação.
- Agrupamento: a reação mapeada na i -ésima posição está bloqueada e, portanto, está fixada no agrupamento indicado pelo valor de i_b .

Na prática, os alelos bloqueados em ambas as partes do cromossomo são desconsiderados pelo operador de mutação e constituem uma solução parcial do problema quando da aplicação do processo de decodificação.

Operador de mutação com bloqueio (PCR multiplex com primers variáveis):

Cada indivíduo sofre a ação do operador de mutação nas duas partes do cromossomo de acordo com as probabilidades, opcionalmente distintas, p_{ma} para agrupamento e p_{ms} para seleção:

- Seleção: um dos alelos desbloqueados, aleatoriamente escolhido, é definido com um valor aleatoriamente selecionado entre 1 e m_s .
- Agrupamento: assim como nas outras versões do AGBHcDB, aplica-se um *swap* entre dois índices desbloqueados aleatoriamente selecionados.

Deve-se destacar que o operador de mutação é aplicado de maneira independente em cada uma das partes do cromossomo, o que permite que a perturbação ocorra eventualmente em somente uma delas.

Decodificação com bloqueio (PCR multiplex com primers variáveis):

Assim como nas versões do algoritmo apresentadas anteriormente, o algoritmo *First Fit* é responsável por agrupar as reações, isto é, os pares de *primers* que as representam, com base na sequência fornecida pela parte de agrupamento do cromossomo, respeitando possíveis bloqueios já efetuados. Para efeitos da verificação de compatibilidade, o algoritmo FF recorre à parte de seleção do cromossomo e identifica qual o par de *primers* $pp_{k_s}(s)$ encontra-se assinalado para cada sequência alvo s .

5 EXPERIMENTOS COMPUTACIONAIS

Neste trabalho, foram desenvolvidos inicialmente três algoritmos para solucionar instâncias do problema do empacotamento, sendo eles: Algoritmo Genético Baseado em Heurística (AGBH), Algoritmo Genético Baseado em Heurística com Bloqueio (AGBHcB), Algoritmo Genético Baseado em Heurística com Bloqueio Dinâmico (AGBHcBD). Posteriormente, o algoritmo AGBHcBD foi adaptado para solucionar o problema de agrupamento de *primers* da PCR *multiplex* em dois cenários: (i) pares de *primers* pré-selecionados para cada amplificação e (ii) múltiplos pares de *primers* candidatos para cada amplificação. Os algoritmos foram desenvolvidos na linguagem própria do software MATLAB, versão 9.0.0.

5.1 PROBLEMA DO EMPACOTAMENTO

Para a avaliação dos algoritmos implementados para o BPP, foram utilizadas instâncias de problemas disponíveis em *benchmarks* amplamente reconhecidos pela comunidade científica. As quatro primeiras classes de problema foram geradas no trabalho de Falkenauer (1996): “u120”, “u250”, “u500”, “u1000”. Cada uma delas contém 20 instâncias com 120, 250, 500 e 1000 itens respectivamente, nas quais o peso dos itens variam uniformemente entre 20 e 100, com a capacidade das caixas igual a 150. O conjunto definido como “Hard10” foi proposto por Scholl, Klein e Jürgens (1997) e contém 10 instâncias consideradas difíceis, principalmente por conta dos pesos que variam amplamente entre 20.000 e 35.000. Cada instância contém 200 itens para serem empacotados em caixas com capacidade igual a 100.000. O último conjunto de problemas, denominado “Hard28”, foi apresentado por Schoenfeld (2002) e considerado muito difícil por trabalhos recentes. São 28 instâncias com diferentes quantidades de itens (160, 180 ou 200), pesos variando de 1 a 800 e a capacidade das caixas igual a 1.000.

5.1.1 Configuração experimental

Os parâmetros do algoritmo genético foram configurados com base em testes preliminares e valores estabelecidos em outros AG’s adaptados ao BPP referenciados na literatura. Os valores foram os mesmos para todos os algoritmos desenvolvidos: população de 100 cromossomos, número de indivíduos de elite igual a 10, probabilidade de mutação e recombinação (ou desbloqueio) igual a 80% e 50%, respectivamente, e critério de parada estabelecido em 10.000 gerações ou quando o mínimo global de caixas fosse encontrado. Cada problema foi executado com 5 diferentes sementes sequenciais geradoras de números aleatórios, fixadas no tipo “*default*” da linguagem, iniciando em 1, conforme relatado nos trabalhos de Alvim et al. (2004), Fleszar e Charalambous (2011) e Quiroz-Castellanos et al. (2015). Cada algoritmo totalizou, portanto, 590 execuções.

5.1.2 Resultados e análise experimental

Nas Tabelas 1, 2 e 3 são apresentados os resultados dos algoritmos implementados para solução do BPP. As colunas indicam o conjunto de dados (Dados); o número de instâncias de cada conjunto (Instâncias); o número de instâncias resolvidas por cada algoritmo, isto é, quando encontram o mínimo global de caixas (Soluções encontradas); o número de execuções que encontraram a solução ótima (Execs. ótimas); número médio de gerações até a convergência, ainda que tenha ocorrido prematuramente (Média de convergência); e o número médio de gerações até atingir um dos critérios de parada (Média de gerações). A coluna “Média de convergência” permite avaliar o esforço necessário até encontrar a solução ou a última melhora na resposta dentro do limite de gerações. Já a coluna “Média de gerações” representa o esforço total empregado nas buscas.

Para as primeiras quatro instâncias, os números de respostas encontradas pelo AGBHcBD foram substancialmente melhores em relação aos outros dois algoritmos, além de apresentar uma média de esforço computacional significativamente menor quando observadas as médias de gerações. Os resultados obtidos pelo AGBHcB indicam a ocorrência do problema de garantia de subotimalidade, mencionado na seção 3.1.1 e considerado como “possivelmente raro” no trabalho de Reeves (1996). O algoritmo converge prematuramente, visto que encontra um número consideravelmente inferior de soluções ótimas quando comparado aos outros algoritmos, com uma média de convergência também inferior.

No caso das instâncias dos conjuntos difíceis (*Hard*), os resultados não apresentaram o mesmo padrão. Nesses casos, os três algoritmos encontraram o mesmo número de respostas, além de encontrá-las praticamente na mesma fração da quantidade total de execuções, com exceção da instância Hard28, na qual o AGBHcB deixou de encontrar a melhor solução em uma execução a mais do que os outros dois algoritmos. Os valores associados às médias de convergência e o esforço empregado pelos algoritmos foram muito semelhantes nestes últimos casos, de modo que a leve vantagem obtida pelo AGBH nesses quesitos não foi considerada suficiente para sua utilização na solução de um problema desconhecido.

A Tabela 4 mostra uma comparação entre o AGBHcBD, algoritmo desenvolvido que obteve o melhor resultado geral, e outros algoritmos considerados como o estado da arte para o problema do empacotamento ou referenciados por trabalhos recentes em função da alta qualidade demonstrada, todos mencionados na seção 2.1.1. Para cada classe de problemas tratada neste trabalho, apresentamos o número de instâncias não resolvidas pelos algoritmos expostos. A ausência deste valor indica que o trabalho não apresentou experimentos para o respectivo conjunto de instâncias.

O algoritmo desenvolvido neste trabalho foi capaz de encontrar todas as respostas para as classes “u120” e “u1000”, assim como a maior parte dos outros algoritmos, com

Tabela 1 – Resultados obtidos pelo AGBH para o problema do empacotamento

Dados	Instâncias	AGBH			
		Soluções encontradas	Execs. ótimas	Média de convergência	Média de gerações
u120	20	17	91	491	1389
u250	20	14	62	1142	4735
u500	20	11	41	4028	8280
u1000	20	0	0	6735	10000
Hard10	10	8	40	941	2857
Hard28	28	5	25	37	8223
TOTAL	118	55	259	13374	35484

Tabela 2 – Resultados obtidos pelo AGBHcB para o problema do empacotamento

Dados	Instâncias	AGBHcB			
		Soluções encontradas	Execs. ótimas	Média de convergência	Média de gerações
u120	20	13	78	28	2226
u250	20	13	42	67	5839
u500	20	3	6	118	9411
u1000	20	0	0	257	10000
Hard10	10	8	40	1189	3083
Hard28	28	5	25	38	8299
TOTAL	118	42	191	1697	38858

Tabela 3 – Resultados obtidos pelo AGBHcBD para o problema do empacotamento

Dados	Instâncias	AGBHcBD			
		Soluções encontradas	Execs. ótimas	Média de convergência	Média de gerações
u120	20	20	98	233	432
u250	20	17	82	381	2166
u500	20	19	91	765	1642
u1000	20	20	96	2036	2385
Hard10	10	8	40	1334	3175
Hard28	28	5	24	42	8227
TOTAL	118	89	431	4791	18027

Tabela 4 – Comparação entre AGBHcBD e algoritmos de referência para o BPP

Dados	HGGA	GPSO	GES	HI_BP	Pert.-SAWMBS	GGA-CGT	GGA-CUDA	AGBHcBD
u120	2	0	0	0	0	0	-	0
u250	2	-	-	0	1	0	-	3
u500	0	-	-	0	0	0	-	1
u1000	0	-	-	0	0	0	-	0
Hard10	-	2	2	0	0	0	-	2
Hard28	-	-	-	23	23	12	23	23

exceção do HGGA. Testes extras realizados para as instâncias “u250” e “u500” revelaram que um aumento de 30 % no número máximo de gerações promove uma redução no número de respostas não encontradas para essas classes, passando para 1 e 0 respectivamente. Entretanto, optou-se por adotar os resultados obtidos com até 10.000 gerações, por ser considerado o limite máximo aplicado em AGs adaptados ao BPP no trabalho de Quiroz-Castellanos et al. (2015).

Para os conjuntos considerados difíceis, de maneira geral, o algoritmo reproduziu a qualidade obtida por alguns dos algoritmos listados, especialmente ao considerar os problemas e os algoritmos individualmente. Deve-se destacar especialmente a instância Hard28, que é fortemente encorajada em trabalhos de referência como o conjunto a ser testado por trabalhos que apresentam algoritmos de alta qualidade. Nessa classe, o algoritmo foi capaz de encontrar o mesmo número de respostas apontadas por dois dos três melhores algoritmos e também igualar a qualidade obtida pelo recente GGA-CUDA.

O número médio de gerações necessárias até a estagnação das respostas variou entre 42 e 2036, indicando que os resultados encontrados exigiram um percentual relativamente baixo do esforço computacional empregado. Por outro lado, as respostas não encontradas apontam a necessidade de aprimoramentos no algoritmo no intuito de melhorar o desempenho computacional no que tange ao processo da busca, consequentemente escapando de soluções localmente ótimas. Ressalta-se que, nesses casos, o número de caixas encontrado ficou sempre em apenas uma unidade acima das soluções ótimas.

5.2 PCR *MULTIPLEX*

A fim de avaliar os algoritmos desenvolvidos para otimização do projeto da PCR *multiplex* foram utilizados dois conjuntos de dados.

5.2.1 Dados sintéticos

Para o cenário com *primers* pré-selecionados para cada amplificação, a função “*randseq*” do MATLAB foi utilizada para gerar randomicamente 150 pares de sequências de DNA, variando aleatoriamente o tamanho entre 18 e 30 bases. Outros quatro conjuntos de 150 pares de sequências foram obtidos com diferentes sementes geradoras para o cenário com *primers* variáveis, totalizando 5 pares de *primers* candidatos a amplificarem cada uma das 150 supostas regiões alvo.

Para cada cenário, três subconjuntos de dados foram retirados do total de 150 supostas amplificações, dando origem a três problemas menores com 15, 30 e 70, alvos respectivamente.

As medidas utilizadas na função de compatibilidade foram construídas da seguinte

forma:

- Energia livre de *Gibbs*: as medidas de ΔG calculadas sobre o pareamento entre dois *primers* quaisquer foram obtidas através do software MultiPLX. Assim como em outros trabalhos, a ferramenta subdivide esse tipo de interação em três categorias, de acordo com a região em que o anelamento ocorre:
 - Anelamento entre a extremidade 3' de ambos os *primers* (considerado o mais crítico);
 - Anelamento entre a extremidade 3' de um *primer* e uma região interna do outro;
 - Anelamento entre as regiões internas de ambos os *primers*.

O valor de ΔG foi calculado para o pior caso em cada tipo, isto é, quando a hibridização é mais estável.

- Temperatura de *melting*: por conta da aleatoriedade na composição das sequências geradas, o cálculo da T_m através da ferramenta MultiPLX retorna valores ineficazes para *primers* reais, o que não ocorreu com a medida de ΔG . Por essa razão, foram gerados valores randômicos de T_m variando aleatoriamente entre 58°C e 65°C para cada *primer*, faixa ideal sugerida em (SHEN et al., 2009). Uma restrição adicional foi imposta para que as T_m 's do *primer* direto e reverso de um mesmo par não apresentassem mais do que 5°C de diferença, amplitude máxima estabelecida em boa parte dos trabalhos voltados ao desenho de *primers*.
- Comprimento dos amplicons: os comprimentos dos supostos amplicons também foram gerados randomicamente, variando entre 100 pb e 2000 pb, limites comumente considerados na literatura relacionada.

5.2.2 Dados biológicos

Para a análise do desempenho do algoritmo a partir de dados biológicos reais, foram obtidas através do GenBank (NCBI), as sequências de 30 genes identificados como marcadores de resistência a antibióticos em patógenos causadores da mastite bovina. A ferramenta Primer3 (UNTERGASSER et al., 2012), uma das mais populares para o desenho de *primers*, foi utilizada para desenhar 5 pares de *primers* candidatos a amplificarem cada um dos genes. Outros dois subconjuntos, com 10 e 20 alvos selecionados aleatoriamente, foram retirados do conjunto de dados original, dando origem a mais dois problemas menores.

A medida de T_m dos *primers* e o ΔG para os três tipos de pareamentos entre dois *primers* quaisquer foram calculadas através do software MultiPLX. Os amplicons produzidos pelos *primers* projetados variam de 131 a 1781 bases de comprimento.

5.2.3 Configuração experimental

Parâmetros PCR *multiplex*:

- ΔG - Os limites mínimos de energia livre entre possíveis pareamentos foram definidos com base no perfil “*high*” do software MultiPLX, que representa alta rigorosidade quanto à possibilidade de formação de dímeros, sendo definidos em:
 - -2 kcal/mol para anelamentos entre a extremidade 3’ de ambos os *primers*;
 - -6 kcal/mol para anelamentos entre a extremidade 3’ de um *primer* com a região interna de outro *primer*;
 - -6 kcal/mol para anelamentos entre regiões internas de ambos os *primers*.
- T_m - O limite de amplitude máxima entre as T_m ’s dentro de um mesmo tubo foi estabelecido em 5°C, valor comumente indicado como referência.
- Mobilidade eletroforética - O limite mínimo de diferença entre os comprimentos dos amplicons produzidos em um mesmo tubo foram definidos em 150 pb. Este valor leva em conta as chances de distinção das bandas eletroforéticas na ampla faixa de comprimentos considerada (100 pb - 2000 pb), principalmente para fragmentos acima de 1000 pb que exigem uma diferença maior do que os 50 pb comumente estabelecidos, como apontado em (SHEN et al., 2009).

Parâmetros do algoritmo genético: com base em testes preliminares, os parâmetros do AG foram definidos nos seguintes valores: população de 100 cromossomos, número de indivíduos de elite igual a 10, probabilidade de mutação e desbloqueio igual a 60% e critério de parada estabelecido em 10.000 gerações. Cada problema foi executado com 30 diferentes sementes sequenciais geradoras de números aleatórios, fixadas no tipo “*default*” da linguagem, iniciando em 1.

5.2.4 Resultados e análise experimental

Experimentos preliminares realizados com as adaptações dos algoritmos AGBH e AGBHcB para o problema da PCR *multiplex* demonstraram um comportamento semelhante acerca dos resultados em relação ao mesmo comparativo entre os protótipos aplicados ao BPP, os quais apontaram a superioridade do AGBHcBD. Dessa forma, optou-se por suprimir tais experimentos nesta dissertação e assumir a adoção do AGBHcBD como método principal de solução.

Os resultados para os experimentos realizados com o AGBHcBD aplicado à otimização da PCR *multiplex* são divididos em três comparativos. No primeiro deles foi feita uma avaliação da função objetivo proposta em relação à aptidão mais simples e direta,

que utiliza somente o número de grupos obtido. No segundo momento a escalabilidade do algoritmo foi investigada juntamente à qualidade da busca, ao executar os mesmos experimentos com uma abordagem totalmente randômica que substituiu a ação do algoritmo genético. Em ambos os comparativos iniciais, foram utilizados os 8 conjuntos de dados sintéticos, envolvendo ambos os cenários abordados. No terceiro comparativo, os dados biológicos são utilizados a fim de comparar o principal algoritmo desenvolvido - o AGBHcBD baseado na extensão do modelo para PCR *multiplex* com *primers* variáveis - com a ferramenta MultiPLX.

5.2.4.1 *Análise experimental da função objetivo*

As Tabelas 5 a 12 expõem os resultados para a avaliação da função objetivo desenvolvida neste trabalho (F.o. proposta), calculada através da equação 4.1 com o intuito de fornecer uma medida mais informativa e eficiente do que a adoção direta do número de tubos obtido como aptidão de um indivíduo (F.o. simples). Os 8 conjuntos de dados sintéticos foram utilizados para este comparativo.

Boa parte dos resultados apresentados não demonstram diferenças significativas em termos de qualidade, avaliada em função da média de tubos obtida por cada composição da aptidão. Nos experimentos em que a qualidade superior foi obtida pela utilização exclusiva do número de tubos (Tabelas 6, 8 e 11), a diferença entre as médias de tubos foi sempre menor do que 0.5. Nesses casos, o esforço computacional necessário para obtenção das respostas (indicado pela “Média de convergência”) foi predominantemente maior por parte da função objetivo simplificada.

Por outro lado, nos experimentos em que a função proposta neste trabalho foi superior (Tabelas 9, 10 e 12), a diferença chegou a atingir uma média de 2 tubos a menos do que nos resultados encontrados com a aptidão simples. Essa vantagem ocorreu principalmente nos dois problemas mais difíceis, *primers* variáveis com 70 e com 150 alvos. Destaca-se ainda que, nesses dois casos, as soluções encontradas nas melhores execuções de ambas as funções tiveram uma diferença de 2 tubos entre si, com vantagem para a função proposta. No caso do problema com 70 alvos, o esforço computacional foi ainda ligeiramente menor por parte da f.o. proposta.

Embora não tenha ocorrido uma superioridade predominante acerca da qualidade dos resultados discutidos, a análise apresentada demonstra uma vantagem levemente mais expressiva nos resultados das buscas direcionadas pela função desenvolvida em relação à aptidão simples. Com base na tendência observada, optou-se pela utilização da função objetivo proposta nos demais experimentos.

Tabela 5 – Comparativo entre funções objetivo: *primers* pré-selecionados com 15 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	8,000	8,000
Desvio padrão	0	0
Melhor	8	8
Pior	8	8
Média de convergência	1,000	1,000

Tabela 6 – Comparativo entre funções objetivo: *primers* variáveis com 15 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	4,133	4,000
Desvio padrão	0,346	0
Melhor	4	4
Pior	5	4
Média de convergência	484,633	561,800

Tabela 7 – Comparativo entre funções objetivo: *primers* pré-selecionados com 30 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	17	17
Desvio padrão	0	0
Melhor	17	17
Pior	17	17
Média de convergência	12,033	1,433

Tabela 8 – Comparativo entre funções objetivo: *primers* variáveis com 30 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	7,667	7,233
Desvio padrão	0,479	0,430
Melhor	7	7
Pior	8	8
Média de convergência	1733,367	3561,333

Tabela 9 – Comparativo entre funções objetivo: *primers* pré-selecionados com 70 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	33	33,100
Desvio padrão	0	0,305
Melhor	33	33
Pior	33	34
Média de convergência	201,100	94,667

Tabela 10 – Comparativo entre funções objetivo: *primers* variáveis com 70 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	17,067	18,433
Desvio padrão	0,365	0,504
Melhor	16	18
Pior	18	19
Média de convergência	3585,100	3747,133

Tabela 11 – Comparativo entre funções objetivo: *primers* pré-selecionados com 150 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	58,367	58,267
Desvio padrão	0,556	0,521
Melhor	57	57
Pior	59	59
Média de convergência	1810,267	2654,633

Tabela 12 – Comparativo entre funções objetivo: *primers* variáveis com 150 alvos

(Núm. de tubos)	F.o. proposta	F.o. simples
Média	37,167	39,167
Desvio padrão	0,648	0,531
Melhor	36	38
Pior	38	40
Média de convergência	5387	4526,967

5.2.4.2 *Análise experimental de qualidade e escalabilidade*

Com o intuito de investigar a qualidade da busca promovida pelo algoritmo genético desenvolvido, os conjuntos de dados sintéticos foram novamente utilizados para uma busca randômica não otimizada. Para tanto, 1 milhão de indivíduos factíveis foram gerados com valores aleatórios e repassados sequencialmente à etapa de decodificação desempenhada pelo algoritmo *First Fit*, que agrupou as reações de acordo com sequência apresentada em cada indivíduo gerado. Após a decodificação, os indivíduos foram avaliados através da função objetivo desenvolvida (equação 4.1). Os resultados são expostos da Tabela 13 à Tabela 20, juntamente com os resultados obtidos pelo AGBHcBD quando direcionado pela mesma função objetivo.

Neste caso, para obter a média de convergência foi considerado que cada conjunto de 100 indivíduos constitui uma de 10.000 gerações randômicas, ou seja, o mesmo total de gerações produzidas pelo AGBHcBD. Embora a busca randômica não garanta uma otimização progressiva das respostas, a média de convergência é capaz de indicar a dificuldade do problema e o esforço computacional necessário para resolvê-lo através de uma abordagem “cega”.

Os resultados obtidos para os experimentos com *primers* pré-selecionados para 15 e 30 alvos (Tabelas 13 e 7) indicaram que esses problemas, de menor magnitude, foram facilmente resolvidos, visto que as soluções (provavelmente globalmente ótimas) foram encontradas predominantemente entre a primeira e a terceira geração, por ambas as buscas. Para o caso com 30 alvos, a média de convergência igual a 12, apresentada pelo o AGBHcBD, foi impactada negativamente por uma única execução das 30 promovidas, que exigiu 322 gerações para encontrar uma permutação que resultasse no agrupamento com 17 tubos necessários. Apesar da relativa facilidade encontrada também no experimento para *primers* pré-selecionados com 70 alvos (Tabela 17), nota-se que o esforço computacional exigido pela busca randômica foi em média aproximadamente 6 vezes maior para encontrar exatamente as mesmas respostas. Nesse caso, os valores de convergência não apresentaram excepcionalidades isoladas.

Considerando um aumento na complexidade do problema nos demais experimentos (*primers* pré-selecionados para 150 alvos e *primers* variáveis para 15, 30, 70 e 150 alvos), os resultados obtidos demonstraram vantagens expressivas para o AGBHcBD. A diferença entre a média de tubos encontrada pelas duas buscas variou de pouco mais de 1, para *primers* variáveis com 15 alvos (Tabela 14), até quase 21 tubos, para *primers* variáveis com 150 alvos (Tabela 20). O esforço computacional utilizado em cada busca reforça a vantagem por parte do AGBHcBD, visto que em alguns casos a qualidade superior foi acompanhada por uma média de convergência menor, como para *primers* pré-selecionados com 150 alvos (Tabela 19) e *primers* variáveis com 15 e 30 alvos (Tabelas 14 e 16).

Uma análise preliminar acerca da escalabilidade do AGBHcBD pode ser baseada na

Tabela 13 – Comparativo entre as buscas: *primers* pré-selecionados com 15 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	8	8
Desvio padrão	0	0
Melhor	8	8
Pior	8	8
Média de convergência	1	1

Tabela 14 – Comparativo entre as buscas: *primers* variáveis com 15 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	4,133	5,167
Desvio padrão	0,346	0,379
Melhor	4	5
Pior	5	6
Média de convergência	484,633	3600,600

Tabela 15 – Comparativo entre as buscas: *primers* pré-selecionados com 30 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	17	17
Desvio padrão	0	0
Melhor	17	17
Pior	17	17
Média de convergência	12,033	1,267

Tabela 16 – Comparativo entre as buscas: *primers* variáveis com 30 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	7,667	11,800
Desvio padrão	0,479	0,407
Melhor	7	11
Pior	8	12
Média de convergência	1733,367	1756,767

Tabela 17 – Comparativo entre as buscas: *primers* pré-selecionados com 70 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	33	33
Desvio padrão	0	0
Melhor	33	33
Pior	33	33
Média de convergência	201,100	1239,533

Tabela 18 – Comparativo entre as buscas: *primers* variáveis com 70 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	17,067	28,300
Desvio padrão	0,365	0,596
Melhor	16	27
Pior	18	29
Média de convergência	3585,100	3067,767

Tabela 19 – Comparativo entre as buscas: *primers* pré-selecionados com 150 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	58	62,400
Desvio padrão	0,556	0,563
Melhor	57	61
Pior	59	63
Média de convergência	1810,267	2685,433

Tabela 20 – Comparativo entre as buscas: *primers* variáveis com 150 alvos

(Núm. de tubos)	AGBHcBD	Busca randômica
Média	37,167	58,100
Desvio padrão	0,648	0,803
Melhor	36	56
Pior	38	59
Média de convergência	5387	3261,533

relação de proporcionalidade entre o número de alvos de amplificação e a quantidade média de tubos encontrada em cada problema. Para o cenário com *primers* pré-selecionados, nota-se que essa relação se mantém próxima a 50% para os três problemas menores e apresenta uma melhora para o problema mais complexo com 150 alvos, quando se aproxima de 38%, indicando que o agrupamento foi mais eficiente nesse caso. Para o cenário com *primers* variáveis, a relação de proporção se manteve próxima a 25% para todos os casos. Embora outros fatores possam interferir nessa proporção, tais como as propriedades dos dados adicionados em cada expansão do problema e o limite de gerações definido, a vantagem progressiva observada em relação à busca randômica e a consistência dos resultados obtidos nos experimentos realizados neste trabalho se apresentam como evidências empíricas relevantes acerca da escalabilidade do AGBHcBD.

Por fim, uma análise importante sobre os experimentos discutidos nesta seção revela uma acentuada redução no número de tubos necessários para agrupar as reações quando opções de pares de *primers* são fornecidas para cada amplificação. As relações de proporção citadas anteriormente demonstram que o total de tubos é reduzido aproximadamente pela metade na maioria dos casos, proporcionando um impacto importante na qualidade das soluções diante da perspectiva de economia exposta nos capítulos iniciais. Os resultados evidenciam ainda a capacidade do algoritmo de lidar simultaneamente com a seleção de um único par por amplificação e o agrupamento eficiente das reações.

5.2.4.3 *Análise experimental com dados biológicos*

Os três conjuntos de dados biológicos foram utilizados para avaliar o comportamento do AGBHcBD em relação a propriedades calculadas a partir de sequências de *primers* reais, que podem apresentar uma distribuição diferente daquela encontrada nos dados gerados sinteticamente. Neste caso, somente o cenário com *primers* variáveis foi considerado, devido à maior complexidade da busca e à vantagem relatada anteriormente para esse contexto. Os mesmos dados biológicos foram utilizados para projetar reações de PCR *multiplex* através do software MultiPLX. A ferramenta foi utilizada como base comparativa em trabalhos recentes (BROWN et al., 2017; MILEWSKI et al., 2017) e, além disso, constitui uma das poucas alternativas disponíveis para abordar o projeto da PCR *multiplex* visando ao agrupamento eficiente das amplificações com opções de pares de *primers* por amplificação.

Os resultados obtidos pelo AGBHcBD são expostos na Tabela 21, seguindo o mesmo padrão apresentado para os experimentos anteriormente analisados, sendo uma coluna para cada um dos conjuntos de dados biológicos: 10, 20 e 30 alvos. As execuções realizadas para o menor conjunto resultaram em uma pequena variação em relação ao número mínimo de tubos necessários, em função de quatro das trinta execuções terem encontrado o mínimo de 5 tubos e não os 4 tubos encontrados pelas demais. Nesses casos,

Tabela 21 – Resultados obtidos pelo AGBHcBD com dados biológicos

(Núm. de tubos)	AGBHcBD		
	10 alvos	20 alvos	30 alvos
Média	4,133	6	9
Desvio padrão	0,346	0	0
Melhor	4	6	9
Pior	5	6	9
Média de convergência	43,533	986,733	1361,167

Tabela 22 – Resultados obtidos pelo software MultiPLX com dados biológicos

	MultiPLX		
	Número de tubos		
Ordenação	10 alvos	20 alvos	30 alvos
<i>File</i>	7	14	22
<i>Random</i>	7	14	22
<i>Friends</i>	6	10	13

diante da média de convergência apresentada, que ficou muito abaixo do limite de 10.000 gerações, fica evidente que o algoritmo não foi capaz de evitar a estagnação da busca em regiões localmente ótimas, apesar do tamanho reduzido do problema. Tal variação não foi observada para os dois conjuntos maiores, com 20 e 30 alvos. Possivelmente, as propriedades dos *primers* e os valores de restrição impostos resultaram em um problema mais difícil para o conjunto com 10 alvos, no qual um agrupamento factível em 4 tubos depende de um número reduzido de combinações e permutações específicas.

Os resultados obtidos pelo software MultiPLX são expostos na Tabela 22, que apresenta, para cada conjunto, o número de tubos necessários após o agrupamento baseado nos três parâmetros distintos de ordenação:

- *File* - As reações são agrupadas com base na sequência fornecida pelo arquivo de entrada;
- *Random* - Uma ordenação aleatória é aplicada sobre os dados. Nesse caso, um parâmetro adicional define o número de ordenações aleatórias a serem testadas, retornando o agrupamento que exigiu o menor número de grupos ao fim das iterações. Esse limite foi definido em 1 milhão, acompanhando o total de avaliações efetuadas pelo AGBHcBD;
- *Friends* - Os alvos são ordenados de acordo com a compatibilidade dos *primers* em relação a todos os demais, iniciando pelo menos compatível. Assim como apontado pelos autores, esta ordenação geralmente encontra as melhores respostas.

Os resultados obtidos pela ferramenta no melhor caso (ordenação “*friends*”) foram inferiores àqueles obtidos pelo AGBHcBD, mesmo ao considerar as piores execuções do algoritmo desenvolvido. Essa desvantagem pode ser justificada pela característica determinística da abordagem que o software MultiPLX emprega no agrupamento das reações, principalmente quando utilizados os parâmetros de ordenação “*file*” e “*friends*”. Uma variação nos resultados era esperada ao definir a ordenação em “random” com 1 milhão de iterações, o que não ocorreu possivelmente por conta da estratégia de seleção do par de *primer* para cada amplificação, que permaneceu determinística, direcionada pelo candidato menos incompatível em relação a todos os demais.

Deve-se mencionar que nas versões iniciais do modelo desenvolvido foram testadas funções objetivo baseadas na agregação de penalizações para as violações dos limites de compatibilidade estabelecidos. Essa abordagem é comumente utilizada nos trabalhos relacionados e foi brevemente analisada na seção 2.1. No entanto, os resultados obtidos pelos modelos iniciais foram, por vezes, inferiores àqueles obtidos pela ferramenta MultiPLX, motivando a estratégia empregada no modelo final apresentado de imposição das restrições de compatibilidade.

6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Este trabalho apresentou um modelo computacional capaz de solucionar problemas de otimização combinatória com fase de agrupamento. A solução foi desenvolvida e adaptada, inicialmente, para o problema do empacotamento e, posteriormente, para a otimização do projeto da PCR *multiplex*.

O modelo foi construído por meio de uma proposta híbrida, na qual um algoritmo genético foi adaptado à busca de permutações ótimas dos elementos, baseadas no processo de decodificação realizado pela heurística construtiva *First Fit*. Além disso, um processo inovador de restrição do espaço de busca foi incorporado de maneira eficiente ao principal algoritmo apresentado, permitindo aprimorar o desempenho do modelo. Tal processo motivou o desenvolvimento de operadores genéticos específicos de bloqueio e desbloqueio, projetados para refinar busca realizada.

6.1 PROBLEMA DO EMPACOTAMENTO

Quanto à adaptação do modelo para o problema do empacotamento, foram realizados experimentos computacionais com *benchmarks* frequentemente explorados na literatura do BPP. Inicialmente, foram comparados os três algoritmos desenvolvidos a partir do modelo proposto. Os resultados evidenciaram, de maneira geral, o aperfeiçoamento pretendido com a incorporação do processo de restrição do espaço de busca. Posteriormente, os resultados obtidos pelo protótipo de melhor desempenho foram comparados aos de algoritmos de referência para o BPP, o que indicou o potencial da abordagem híbrida explorada. A qualidade similar obtida para os conjuntos de instâncias testadas, incluindo o mais difícil deles (“Hard28”), sugere que o modelo pode ser competitivo, principalmente ao considerar uma perspectiva de aperfeiçoamentos futuros.

Uma melhor compreensão do mecanismo de empacotamento em relação à obtenção de ótimos locais pode contribuir para o aperfeiçoamento dos operadores de modo a torná-los mais eficazes, promovendo um desempenho ainda melhor para a meta-heurística híbrida. Nesse sentido, esforços futuros podem ser direcionados ao aprimoramento do operador de desbloqueio, permitindo que informações relevantes acerca dos dados sejam utilizadas na etapa de escolha da caixa a ser desbloqueada. A título de exemplificação, pode ser investigada a predileção pelo desbloqueio de caixas que contenham mais itens de menor peso, importantes para o preenchimento de caixas com pouca capacidade ociosa. Além disso, deve ser considerada a utilização de outras heurísticas construtivas em substituição ao algoritmo *First Fit*.

6.2 OTIMIZAÇÃO DA PCR *MULTIPLEX*

Acerca do problema de otimização da PCR *multiplex*, foram realizados experimentos exploratórios, baseados em dados sintéticos, com o intuito inicial de avaliar a função objetivo proposta para o problema. Os resultados indicaram a necessidade de maior aprimoramento da função desenvolvida, visando promover uma melhora consistente na qualidade dos resultados quando comparados aos de uma busca baseada apenas no número de grupos exigidos por cada solução. Para tanto, novas formas de penalização e distinção de soluções com o mesmo número de grupos devem ser estudadas.

No segundo momento, o comparativo apresentado entre o desempenho do algoritmo genético desenvolvido e uma abordagem randômica cumpriu a expectativa de investigar preliminarmente a qualidade da adaptação do modelo ao problema da PCR *multiplex*. Além disso, a escalabilidade do algoritmo também foi avaliada no respectivo comparativo, que apontou a estabilidade na qualidade das soluções obtidas diante das variações no tamanho do problema.

Foi apresentada ainda, uma análise experimental comparativa entre o cenário com *primers* pré-selecionados e o cenário com *primers* variáveis por amplificação. A melhora substancial na qualidade dos resultados obtidos com *primers* variáveis aponta a relevância de abordagens computacionais capazes de lidar com esse cenário, que exige o tratamento do respectivo aumento na complexidade do problema. Tal capacidade é raramente encontrada em propostas anteriores para o projeto da PCR *multiplex* e, dessa forma, os resultados apresentados corroboram uma importante contribuição deste trabalho ao demonstrarem a viabilidade de aplicação do protótipo em ambos os casos.

Finalmente, foram utilizados dados biológicos reais para uma análise experimental comparativa entre o algoritmo desenvolvido e o programa MultiPLX que, dentre as referências ora descritas, é a que apresenta a modelagem do problema mais próxima daquela sustentada neste trabalho. Os resultados alcançados pelo algoritmo proposto foram superiores aos da ferramenta referenciada em todos os casos considerados.

O modelo computacional desenvolvido para a otimização da PCR *multiplex* permite lidar com boa parte das limitações observadas em trabalhos anteriores, uma vez que possibilita o agrupamento eficiente das reações de PCR, contemplando a minimização do número de tubos necessários e o atendimento das restrições de compatibilidade, tanto para casos com *primers* pré-selecionados quanto para *primers* variáveis. Além disso, o desacoplamento da função objetivo em relação às restrições do problema permite que o conjunto de medidas de compatibilidade seja alterado sem a necessidade de adaptação na função objetivo. Desse modo, o sucesso retratado na análise experimental realizada evidencia a aplicabilidade da estratégia apresentada neste trabalho.

Futuramente, deve ser avaliada uma forma de lidar com a variação não linear da

restrição associada à mobilidade eletroforética dos amplicons que, apesar de comumente fixada em um único valor, teve sua complexidade abordada em maior amplitude no trabalho de Shen et al. (2009).

Outra possibilidade de aprimoramento pode ser desenvolvida a partir da adoção simultânea da imposição de restrições e a agregação de penalizações baseadas nas medidas de compatibilidade. Dessa forma, poderia ser estudada a viabilização de uma busca que atenda aos limites definidos para cada critério e, posteriormente, procure permutar as reações entre os tubos formados com intuito de otimizar as avaliações de compatibilidade além dos limites estabelecidos. Essa ideia é especialmente relevante ao levar em conta que os limiares de compatibilidade entre componentes da PCR são baseados em cálculos aproximados, assim como as propriedades das sequências de nucleotídeos. Tal característica, decorrente da complexa natureza biológica do problema, implica em algum nível de incerteza quanto aos critérios de previsão de interações indesejadas. Nesse caso, deve ser avaliado se uma possível intensificação das chances de sucesso da reação aliada a garantia de atendimento das restrições justificaria a dependência da função objetivo em relação às medidas de compatibilidade.

Por fim, a aplicação do modelo computacional desenvolvido em ambos os problemas abordados, o problema do empacotamento e o problema da PCR *multiplex*, sustenta a viabilidade de generalização da estratégia apresentada. Estudos adicionais devem incluir a perspectiva de novas adaptações para outros problemas de agrupamento conhecidos.

REFERÊNCIAS

- ALVIM, Adriana CF; RIBEIRO, Celso C; GLOVER, Fred; ALOISE, Dario J. A hybrid improvement heuristic for the one-dimensional bin packing problem. **Journal of Heuristics**, Springer, v. 10, n. 2, p. 205–229, 2004.
- BASU, Chhandak. **PCR primer design**. [S.l.]: Springer, 2015.
- BROWN, Silas S.; CHEN, Yun-Wen; WANG, Ming; CLIPSON, Alexandra; OCHOA, Eguzkine; DU, Ming-Qing. Primerpooler: automated primer pooling to prepare library for targeted sequencing. **Biology Methods and Protocols**, v. 2, n. 1, p. bpx006, 2017. Disponível em: <<http://dx.doi.org/10.1093/biomethods/bpx006>>.
- BUSTIN, Stephen A. **The PCR revolution: basic technologies and applications**. [S.l.]: Cambridge University Press, 2010.
- CHENG, Yu-Huei. Computational intelligence-based polymerase chain reaction primer selection based on a novel teaching-learning-based optimisation. **IET nanobiotechnology**, IET, v. 8, n. 4, p. 238–246, 2014.
- DELORME, Maxence; IORI, Manuel; MARTELLO, Silvano. Bin packing and cutting stock problems: Mathematical models and exact algorithms. **European Journal of Operational Research**, Elsevier, v. 255, n. 1, p. 1–20, 2016.
- FALKENAUER, Emanuel. A new representation and operators for genetic algorithms applied to grouping problems. **Evolutionary computation**, MIT Press, v. 2, n. 2, p. 123–144, 1994.
- FALKENAUER, Emanuel. A hybrid grouping genetic algorithm for bin packing. **Journal of heuristics**, Springer, v. 2, n. 1, p. 5–30, 1996.
- FALKENAUER, Emanuel; DELCHAMBRE, Alain. A genetic algorithm for bin packing and line balancing. In: IEEE. **Proceedings 1992 IEEE International Conference on Robotics and Automation**. [S.l.], 1992. p. 1186–1192.
- FLESZAR, Krzysztof; CHARALAMBOUS, Christoforos. Average-weight-controlled bin-oriented heuristics for the one-dimensional bin-packing problem. **European Journal of Operational Research**, Elsevier, v. 210, n. 2, p. 176–184, 2011.
- FLESZAR, Krzysztof; HINDI, Khalil S. New heuristics for one-dimensional bin-packing. **Computers & operations research**, Elsevier, v. 29, n. 7, p. 821–839, 2002.
- GARCÍA, Liliana Torcoroma; CRISTANCHO, Laura Maritza; VERA, Erika Patricia; BEGAMBRE, O. A new multiplex-pcr for urinary tract pathogen detection using primer design based on an evolutionary computation method. **Journal of microbiology and biotechnology**, v. 25 10, p. 1714–27, 2015.
- GARY, Michael R; JOHNSON, David S. **Computers and Intractability: A Guide to the Theory of NP-completeness**. [S.l.]: WH Freeman and Company, New York, 1979.
- GOLDBERG, David E; LINGLE, Robert et al. Alleles, loci, and the traveling salesman problem. In: LAWRENCE ERLBAUM, HILLSDALE, NJ. **Proceedings of an international conference on genetic algorithms and their applications**. [S.l.], 1985. v. 154, p. 154–159.

- GOURGAND, Michel; GRANGEON, Nathalie; KLEMENT, Nathalie. An analogy between bin packing problem and permutation problem: A new encoding scheme. In: SPRINGER. **IFIP International Conference on Advances in Production Management Systems**. [S.l.], 2014. p. 572–579.
- HENEGARIU, O; HEEREMA, NA; DLOUHY, SR; VANCE, GH; VOGT, PH. Multiplex pcr: critical parameters and step-by-step protocol. **Biotechniques**, Future Science, v. 23, n. 3, p. 504–511, 1997.
- JOHNSON, David S. **Near-optimal bin packing algorithms**. Tese (Doutorado) — Massachusetts Institute of Technology, 1973.
- JOHNSON, David S. Fast algorithms for bin packing. **Journal of Computer and System Sciences**, Elsevier, v. 8, n. 3, p. 272–314, 1974.
- KALENDAR, Ruslan; TSELYKH, Timofey V; KHASSENOV, Bekbolat; RAMANCULOV, Erlan M. Introduction on using the fastpcr software and the related java web tools for pcr and oligonucleotide assembly and analysis. In: **PCR**. [S.l.]: Springer, 2017. p. 33–64.
- KAPLINSKI, Lauris; ANDRESON, Reidar; PUURAND, Tarmo; REMM, Mairo. Multiplx: automatic grouping and evaluation of pcr primers. **Bioinformatics**, v. 21, n. 8, p. 1701–1702, 2005. Disponível em: <<http://dx.doi.org/10.1093/bioinformatics/bti219>>.
- KASHAN, Ali Husseinzadeh; AKBARI, Ali Akbar; OSTADI, Bakhtiar. Grouping evolution strategies: An effective approach for grouping problems. **Applied Mathematical Modelling**, Elsevier, v. 39, n. 9, p. 2703–2720, 2015.
- KASHAN, Ali Husseinzadeh; KASHAN, Mina Husseinzadeh; KARIMIYAN, Somayyeh. A particle swarm optimizer for grouping problems. **Information Sciences**, Elsevier, v. 252, p. 81–95, 2013.
- KUMAR, Rakesh; GOPAL, Girdhar; KUMAR, Rajesh. Novel crossover operator for genetic algorithm for permutation problems. **International Journal of Soft Computing and Engineering (IJSCE)**, v. 3, n. 2, p. 252–258, 2013.
- LEE, Chungnan; WU, Jain-Shing; SHIUE, Yow-Ling; LIANG, Hong-Long. Multiprimer. **Applied bioinformatics**, Springer, v. 5, n. 2, p. 99–109, 2006.
- LEE, In-Hee; SHIN, Soo-Yong; ZHANG, Byoung-Tak. Multiplex pcr assay design by hybrid multiobjective evolutionary algorithm. In: SPRINGER. **International Conference on Evolutionary Multi-Criterion Optimization**. [S.l.], 2007. p. 376–385.
- LIANG, Hong-Long; LEE, Chungnan; WU, Jain-Shing. Multiplex pcr primer design for gene family using genetic algorithm. In: ACM. **Proceedings of the 7th annual conference on Genetic and evolutionary computation**. [S.l.], 2005. p. 67–74.
- LIPSON, Doron. **Optimization problems in design of oligonucleotides for hybridization based methods**. [S.l.]: Technion-Israel Institute of Technology, Faculty of Computer Science, 2003.
- MANDOIU, Ion; ZELIKOVSKY, Alexander. **Bioinformatics algorithms: techniques and applications**. [S.l.]: John Wiley & Sons, 2008. v. 3.

MARTELLO, Silvano; TOTH, Paolo. Lower bounds and reduction procedures for the bin packing problem. **Discrete applied mathematics**, Elsevier, v. 28, n. 1, p. 59–70, 1990.

MILEWSKI, Marek C.; KAMEL, Karol; KURZYNSKA-KOKORNIAK, Anna; CHMIELEWSKI, Marcin K.; FIGLEROWICZ, Marek. Evoligo: A novel software to design and group libraries of oligonucleotides applicable for nucleic acid-based experiments. **Journal of computational biology : a journal of computational molecular cell biology**, v. 24 10, p. 1014–1028, 2017.

MOHAMADI, N. Application of genetic algorithm for the bin packing problem with a new representation scheme. **Mathematical Sciences**, v. 4, n. 3, p. 253–266, 2010.

NICODÈME, P; STEYAERT, Jean-Marc. Selecting optimal oligonucleotide primers for multiplex pcr. v. 5, p. 210–3, 02 1997.

OZCAN, Sukru Ozer; DOKEROGLU, Tansel; COSAR, Ahmet; YAZICI, Adnan. A novel grouping genetic algorithm for the one-dimensional bin packing problem on gpu. In: SPRINGER. **International Symposium on Computer and Information Sciences**. [S.l.], 2016. p. 52–60.

QUIROZ-CASTELLANOS, Marcela; CRUZ-REYES, Laura; TORRES-JIMENEZ, Jose; GÓMEZ, Claudia; HUACUJA, Héctor J Fraire; ALVIM, Adriana CF. A grouping genetic algorithm with controlled gene transmission for the bin packing problem. **Computers & Operations Research**, Elsevier, v. 55, p. 52–64, 2015.

REEVES, Colin. Hybrid genetic algorithms for bin-packing and related problems. **Annals of Operations Research**, Springer, v. 63, n. 3, p. 371–396, 1996.

RIBEIRO, João Batista; SILVA, Márcio R.; GUIMARÃES, Alessandro de Sá; MENDONÇA, Letícia C.; SOUZA, Guilherme N. de; ALMEIDA, Paula A. A.; LANGE, Carla Christine; BRITO, Maria Aparecida V. P. Padronização de um método baseado em pcr para detecção de genes de resistência a eritromicina e tetraciclina diretamente em amostras de leite bovino. In: **VI Congresso Brasileiro De Qualidade Do Leite**. [S.l.: s.n.], 2015.

ROUCHKA, Eric C; KHALYFA, Abdelnaby; COOPER, Nigel GF. Mprime: efficient large scale multiple primer and oligonucleotide design for customized gene microarrays. **BMC bioinformatics**, BioMed Central, v. 6, n. 1, p. 175, 2005.

SCHOENFIELD, Jon E. Fast, exact solution of open bin packing problems without linear programming. **Draft, US Army Space and Missile Defense Command, Huntsville, Alabama, USA**, 2002.

SCHOLL, Armin; KLEIN, Robert; JÜRGENS, Christian. Bison: A fast hybrid procedure for exactly solving the one-dimensional bin packing problem. **Computers & Operations Research**, Elsevier, v. 24, n. 7, p. 627–645, 1997.

SCHOSKE, Richard; VALLONE, Pete M; RUITBERG, Christian M; BUTLER, John M. Multiplex pcr design strategy used for the simultaneous amplification of 10 y chromosome short tandem repeat (str) loci. **Analytical and bioanalytical chemistry**, Springer, v. 375, n. 3, p. 333–343, 2003.

SHEN, Zhiyong; QU, Wubin; WANG, Wen; LU, Yiming; WU, Yonghong; LI, Zhifeng; HANG, Xingyi; WANG, Xiaolei; ZHAO, Dongsheng; ZHANG, Chenggang. Mpprimer: a program for reliable multiplex pcr primer design. In: **BMC Bioinformatics**. [S.l.: s.n.], 2009.

ÜLKER, Özgür; KORKMAZ, Emin Erkan; ÖZCAN, Ender. A grouping genetic algorithm using linear linkage encoding for bin packing. In: SPRINGER. **International Conference on Parallel Problem Solving from Nature**. [S.l.], 2008. p. 1140–1149.

UNTERGASSER, Andreas; CUTCUTACHE, Ioana; KORESSAAR, Triinu; YE, Jian; FAIRCLOTH, Brant C; REMM, Mairo; ROZEN, Steven G. Primer3—new capabilities and interfaces. **Nucleic acids research**, Oxford University Press, v. 40, n. 15, p. e115–e115, 2012.

VALLONE, Peter M; BUTLER, John M. Autodimer: a screening tool for primer-dimer and hairpin structures. **Biotechniques**, Future Science, v. 37, n. 2, p. 226–231, 2004.

WU, Li-Cheng; HORNG, Jorng-Tzong; HUANG, Hsi-Yuan; LIN, Feng-Mao; HUANG, Hsien-Da; TSAI, Meng-Feng. Primer design for multiplex pcr using a genetic algorithm. **Soft Computing**, Springer, v. 11, n. 9, p. 855–863, 2007.

YAMADA, Tomoyuki; SOMA, Haruhiko; MORISHITA, Shinichi. Primerstation: a highly specific multiplex genomic pcr primer design server for the human genome. **Nucleic acids research**, Oxford University Press, v. 34, n. suppl_2, p. W665–W669, 2006.

ZANGENBERG, G.; SAIKI, R.K.; REYNOLDS, R. 6 - multiplex pcr: Optimization guidelines. In: INNIS, Michael A.; GELFAND, David H.; SNINSKY, John J. (Ed.). **PCR Applications**. San Diego: Academic Press, 1999. p. 73 – 94. ISBN 978-0-12-372185-3.

Disponível em:

<<http://www.sciencedirect.com/science/article/pii/B9780123721853500079>>.