

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Análise do Algoritmo de Geração de Sequências de DNA via Códigos Corretores de Erros

Alice Noronha de Oliveira¹

Discente do Programa de Pós-Graduação em Estatística Aplicada e Biometria, Unifal, Alfenas, MG

Anderson José de Oliveira²

Professor Adjunto do curso Matemática-Licenciatura, Instituto de Ciências Exatas, UNIFAL, Alfenas, MG

1 Introdução

A teoria dos códigos corretores de erros tem como objetivo desenvolver métodos capazes de detectar e corrigir erros que possam surgir durante a transmissão ou armazenamento de dados. Embora não apresente relação aparente, estudos recentes evidenciam a utilização dos códigos corretores de erros também na transferência e armazenamento de informações genéticas. A relação entre essas duas áreas foi apresentada por [1] e [3], onde apontaram a existência dos códigos corretores de erros relacionados a sequências de DNA, bem como formas de reproduzir e classificar matematicamente essas sequências utilizando uma analogia entre um modelo do sistema de comunicação digital com o modelo de sistema de comunicação de informação genética, possibilitando assim análises mutacionais. [2] apresentaram o desenvolvimento de algoritmos e sua implementação computacional, para analisar sequências de DNA. Com isso, evidenciaram as vantagens de associar programas computacionais com a análise dessas sequências, além disso, contribuíram para a criação de banco de dados que armazena palavras-código identificadas como sequência de DNA. Para a análise das sequências de DNA utiliza-se os códigos BCH, os quais possuem características simples, porém com alto poder de detecção de erros, tornado esses códigos eficientes na transmissão de informação [4]. O estudo dos códigos corretores de erros no processo de transmissão de informação genética permite uma interessante conexão entre elementos da Álgebra, Engenharia e Biologia e possibilita que sequências de DNA com funções biológicas distintas possam ser reproduzidas matematicamente, facilitando a análise de mutações genéticas, que futuramente, pode reduzir tempo e custos laboratoriais. O objetivo deste trabalho é apresentar o detalhamento do algoritmo proposto por [1] e [3] a fim de identificar onde ocorreu uma mutação em uma molécula de DNA e como essa alteração pode modificar o processo de transmissão de informação genética.

¹alicenoronhaa@hotmail.com

²anderson.oliveira@unifal-mg.edu.br

2 Resultados e Discussões

O algoritmo citado anteriormente possibilita a identificação de uma estrutura de códigos corretores de erros nas sequências de DNA, além de possibilitar a classificação dessas sequências sob um ponto de vista matemático. A execução desse programa pode ser descrita através de 7 passos que permitem a construção dos códigos BCH sobre o corpo $GF(4^r)$ e sobre o anel \mathbb{Z}_4 , listados a seguir:

Passo 1: Gerar todos os polinômios primitivos com grau r , relacionados à extensão de Galois. *Passo 2:* Selecionar um polinômio primitivo de cada vez e localizar o grupo das unidades, denotado por $GR^*(4, r)$. *Passo 3:* Determinar o polinômio gerador e o polinômio verificação de paridade do código BCH. *Passo 4:* Determinar as permutações entre o código genético e o alfabeto do código BCH \mathbb{Z}_4 , dessa forma é realizado o rotulamento das sequências de DNA. *Passo 5:* Verificar se a sequência de DNA gerada no Passo 4 é palavra-código de acordo com os padrões de erros estabelecidos. *Passo 6:* Comparar as sequências armazenadas no Passo 5 com a sequência de DNA original, para verificar onde os erros ocorreram. *Passo 7:* Selecionar outro polinômio primitivo até que todos sejam analisados e assim é finalizado o algoritmo.

Por meio do algoritmo de geração de proteína, espera-se gerar uma sequência de nucleotídeos associados à anemia falciforme para identificar mudanças genéticas observadas nessa doença. Vale ressaltar que esta pesquisa está em andamento e acredita-se que essa análise poderá auxiliar no estudo dessa doença, caracterizando-a matematicamente por meio dos códigos corretores de erros.

Agradecimentos

À Unifal-MG e à Fapemig pelo apoio financeiro.

Referências

- [1] A. S. L. Rocha e Jr. R. Palazzo. Modelo de sistema de comunicação digital para o mecanismo de importação de proteínas mitocondriais através de códigos corretores de erros. Tese (Doutorado) - Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2010.
- [2] G. P. Diogo e Jr. R. Palazzo. Uma abordagem computacional para análise de sequências de DNA por meio dos códigos corretores de erros. Dissertação (Mestrado) - Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2014.
- [3] L. C. B. Faria e Jr. R. Palazzo. Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA, Tese (Doutorado) - Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2011.
- [4] M. C. R. Trancoso. Sistematização da codificação e discodificação de códigos BCH. Dissertação (Mestrado) - Departamento de Engenharia Eletrotécnica e de Computadores, Universidade do Porto, 1995.