Modelos de Codificação Genética e Genômica gerados através de Códigos Corretores de Erros

Débora E. C. Matoso[1], *Cátia R. O. Quilles Queiroz*[2†]

[1] *Universidade Federal de Alfenas*

[2] *Docente do Instituto de Ciências Exatas, Universidade Federal de Alfenas*

Resumo: *Este trabalho apresenta um estudo de códigos corretores de erros com aplicações em modelos de codificação genética e genômica apresentados em Rocha (2010) e Faria (2011).*

Palavras-chave: Códigos corretores de erros, sequências de DNA, codificação genética e genômica.

Abstract: *This paper presents a study of error-correcting codes with applications in models of genetic and genomic coding presented in Rocha (2010) and Faria (2011).*

Keywords: Error-correcting codes, DNA sequences, genetic and genomic codification.

Introdução

A transferência da informação é uma preocupação tanto para a teoria de comunicações quanto para a genética, embora aparentemente não apresentem relação. Considerando o fato de que a teoria de comunicações é realizada pelo homem, e a genética por um processo natural, a principal diferença entre elas é não operarem em uma mesma dimensão, ou seja, enquanto que a teoria de comunicações envia mensagens no espaço, de um lugar para outro, a genética envia mensagens hereditárias no tempo (BATTAIL, 2008).

O sucesso considerável da tecnologia de comunicação conta com o progresso significativo na concepção de dispositivos físicos, mas também, embora muito menos perceptível, no desenvolvimento de um poderoso ferramental conceitual, consistentemente garantido pela teoria da informação. Embora esse ferramental tenha sido originalmente desenvolvido para a comunicação através do espaço, é suficientemente abrangente para aplicação na comunicação através do tempo (BATTAIL, 2008).

A questão central pode ser colocada da seguinte maneira: A estrutura teórica elaborada pelo homem pode contribuir para uma melhor compreensão dos processos naturais que envolvem a comunicação genética? A resposta é afirmativa. A perspectiva provida principalmente pela teoria da informação renova a visão que podemos ter do "mundo vivo" (BATTAIL, 2008).

Desde meados do século XX vem ocorrendo grandes avanços tanto na engenharia de comunicação quanto na engenharia genética. Em 1953, a estrutura de dupla hélice do DNA foi decifrada por James Watson, Francis Crick, Maurice Wilkins e Rosalind Franklin. Com esta descoberta ficou claro que a informação genética está armazenada na forma de duas fitas diretamente complementares compostas por letras de um alfabeto de quatro símbolos. Até a descoberta das bases moleculares da genética, os pesquisadores concentravam-se na genética clássica, baseada nas leis da hereditariedade propostas pelo monge austríaco Gregor Mendel.

Em 1940, Claude Elwood Shannon em sua tese de doutorado desenvolveu uma proposta ligando as relações matemáticas à genética Mendeliana, com o objetivo de esclarecer como diferentes combinações de características se propagaram através de várias gerações.

[†] *Autor correspondente:* catia.quilles@gmail.com.

Apesar do trabalho ser muito original naquela época, o fato de não ter sido publicado, o tornou pouco conhecido e divulgado. Após ter concluído a sua tese de doutorado, Shannon desviou seu foco para a comunicação digital e criptografia.

Em 1948, Shannon estabeleceu a teoria fundamental de um sistema de comunicação digital, introduzindo o conceito de informação baseado somente na característica estatística da fonte de informação, definindo a informação de maneira abstrata independente da semântica que não diferencia texto, vídeo ou áudio como era geralmente feito naquela época nos estudos de sistemas de comunicação. Usando tal definição de informação, Shannon provou que a mensagem gerada por uma fonte de informação pode ser compactada até o limite da entropia da fonte (teorema de codificação de fonte) e que é possível codificar a mensagem de tal maneira que possamos transmiti-la livre de erros com uma taxa máxima que o canal permite (teorema de codificação de canal) (SHANNON, 1948). Desde então, a engenharia de comunicações tem criado algoritmos e estratégias para atingir os limitantes destes dois teoremas e consequentemente realizando grandes avanços tecnológicos.

A elucidação da estrutura do DNA possibilitou descobertas fundamentais na biologia celular e molecular. Essas descobertas revolucionaram a ciência da vida e proporcionaram o desenvolvimento em tecnologias de DNA recombinante e o lançamento das indústrias biotecnológicas. Por outro lado, Shannon estabeleceu a teoria fundamental de um sistema de comunicação digital. A sua teoria ocasionou um impacto enorme em nosso cotidiano levando principalmente ao desenvolvimento dos celulares, da internet e dos computadores (SHANNON, 1948).

Susan Hockfield, presidente do Instituto Tecnológico de Massachusetts (MIT), comentou:

> Estas revoluções mostraram as sementes de uma terceira revolução que liga as ciências da vida com a engenharia e as ciências físicas de novas maneiras eficazes. Muitos dos fundadores da biologia molecular vieram das ciências físicas, trazendo para a biologia novas estratégias e tecnologias analíticas. Com a evolução dos dados e da biologia tecnológica, biólogos trabalharam cada vez mais de perto com os matemáticos, engenheiros e físicos (tradução editorial da revista Science vol.323 de 27/02/2009)

Historicamente, a aplicação da teoria da informação para análises de dados genéticos iniciou-se na década de 1970, porém esses esforços não tiveram sucesso. Após alguns anos, o aumento de dados genéticos despertou novamente o interesse na aplicação da teoria da informação ao estudo do genoma. Esse segundo período de pesquisas continua até o presente momento, porém com um número muito reduzido de pesquisadores. Os trabalhos atualmente buscam analogias entre o fluxo de informação biológica e o sistema de comunicação, dividindo-se basicamente em três linhas de pesquisas: teoria da informação genética, teoria da comunicação genética e a teoria da codificação genética (ROCHA, 2010; FARIA, 2011; FARIA, 2004; ROCHA, 2004; FARIA et al., 2010).

A aplicação das teorias da informação, comunicação e codificação em sistemas biológicos contribuem para uma melhor compreensão dos paradigmas biológicos fazendo com que a biologia, que hoje é uma ciência descritiva, se transforme em uma ciência fundamentada teoricamente.

Estruturas Algébricas e Códigos

Anéis e Corpos

Definição .1 *Um anel* $< R, +, \Delta >$ *é um conjunto não vazio* R *juntamente com duas operações binárias* $+$ *e* Δ *definidas sobre* R*, as quais chamamos de adição e multiplicação, tal que os seguintes axiomas são satisfeitos:*

1. $< R, + >$ *é um grupo abeliano;*
2. *A operação de multiplicação é associativa, isto é,* $(ab)c = a(bc), \forall a, b, c \in R$*;*
3. *Para todo* $a, b, c \in R$*, é valida a lei distributiva à esquerda,* $a(b + c) = (ab) + (ac)$*, e a lei distributiva à direita,* $(a + b)c = (ac) + (bc)$*.*

Definição .2 *Se* a *e* b *são elementos não nulos de um anel* R *tais que* $ab = 0$ *ou* $ba = 0$*, então* a *e* b *são divisores de zero.*

Definição .3 *Seja* R *um anel. Um* R*-módulo consiste de um grupo abeliano* G *e uma operação de multiplicação de cada elemento de* G *por todo elemento de* R *pela esquerda, tais que para todo* $\alpha, \beta \in G$ *e* $r, s \in R$*, as seguintes condições são satisfeitas:*

1. $(r\alpha) \in G$*;*
2. $r(\alpha + \beta) = r\alpha + r\beta$*;*
3. $(r + s)\alpha = r\alpha + s\alpha$*;*
4. $(rs)\alpha = r(s\alpha)$*.*

Definição .4 *Um corpo é um anel comutativo com unidade e tal que todo elemento não-nulo é inversível.*

Pode-se dizer então, que F é um corpo sob as operações $+$ e Δ se, e somente se, F constitui um grupo abeliano sob essas operações e, para a operação Δ, é válida a lei distributiva. Desta forma, pode-se dizer que um corpo apresenta no mínimo dois componentes: as identidades das operações $+$ e Δ. O número de elementos num corpo é a **ordem** do mesmo e um corpo onde este número é finito é chamado **corpo finito**.

Códigos Corretores de Erros

Um dos objetivos principais dos códigos corretores de erros (CCEs) é aumentar a eficiência ao se enviar uma informação, garantindo que ela chegue da mesma forma como foi mandada, e, se não chegar seja corrigida para tal. Os CCEs estão presentes no nosso cotidiano de inúmeras formas, por exemplo, quando assistimos a um programa de televisão, ouvimos um CD de música, falamos ao telefone, mandamos uma mensagem a alguém via Pager ou navegamos pela internet. Algumas outras aplicações que podem ser listadas são:

a) Uso do bit de paridade como um artifício detector de erro;

b) Armazenamento de dados em discos - CCEs estão sendo muito utilizados por aumentarem a densidade de armazenamento;

c) Transmissão de informação pelas naves espaciais;

d) Áudio digital - o aumento da popularidade do áudio digital deve-se ao desenvolvimento dos CCEs que contribuem, e muito, com o processo de digitalização. Por meio desse sistema, quando é iniciada a leitura de um CD, ele é capaz de corrigir os erros produzidos por marcas de dedos, arranhões, e outras imperfeições, para em seguida, transformar em sinais sonoros.

Códigos de bloco

Os códigos de bloco são atribuídos por serem códigos sem memória. O ponto de partida é um conjunto A, que pode ou não ser finito, chamado alfabeto.

Definição .5 *Um código C sobre um alfabeto A é qualquer subconjunto não-vazio do espaço de sequencias A^I, onde A é chamado alfabeto do código e I é o conjunto de índices das sequências $c = \{c_i : i \in I\}$. Chamamos de palavras-código os elementos, ou símbolos, no alfabeto A que compõem o código C.*

Um código de bloco é caracterizado por três critérios principais: seu comprimento, sua dimensão e sua distância mínima.

Definição .6 *Um código de bloco C de comprimento n sobre um alfabeto A é qualquer subconjunto A_n das sequências $c = \{c_i : i \in I\}$.*

Definição .7 *A dimensão de um código C é dada por $k = log_{|A|}|C|$, onde $|\Delta|$ denota a cardinalidade do conjunto.*

Definição .8 *Seja C um código de comprimento n tal que $|C| \geq 2$. A distância mínima de Hamming de C, denotada por $dmin(C)$ é dada por:*

$$dmin(C) = min_{x,y \in C, x \neq y} d(x, y).$$

Códigos lineares

Levando em consideração um código de blocos com qK palavras-código e comprimento n, se k e n são relativamente grandes, então será muito difícil um espaço para armazenar essas palavras-código. Deste modo, códigos de bloco com uma estrutura linear são mais convenientes e diminuem a complexidade do codificador. A maioria dos códigos utilizados hoje pertence à classe dos códigos lineares.

Definição .9 *Um código (n, k, d) é dito linear se, e somente se, todas as suas palavras-código formam um subespaço vetorial de dimensão K do espaço vetorial F_q^n, o conjunto das n-uplas do corpo F_q.*

Códigos cíclicos sobre $\mathbb{Z}_q$

Trabalharemos com códigos cíclicos sobre anéis do tipo $\mathbb{Z}_q$. Mais especificamente, o anel Z_4, onde é feita a associação das bases nitrogenadas (timina, citosina, guanina e adenina) que possuem quatro elementos. Os códigos cíclicos serão o alicerce para o progresso da construção do Código BCH sobre as estruturas algébricas de anéis e corpos e suas extensões de Galois.

Definição .10 *Um código linear C com parâmetros (n, k) sobre Z_q é cíclico se, para toda palavra-código $v = (v_0, v_1, v_2, ..., v_{n-1}) \in C$, todo deslocamento cíclico $v(1) = (v_{n-1}, v_0, v_1, v_2, ..., v_{n-2}) \in C$, com $v_i \in Z_q$, $0 \leq i \leq n \leq 1$.*

Códigos BCH sobre anéis e corpos

Os códigos BCH formam uma pertinente classe de códigos cíclicos principalmente pela simplicidade dos procedimentos de codificação e decodificação associados, tornando-os competentes candidatos a serem utilizados em aplicações práticas. Os códigos BCH foram descobertos por R. C. Bose, D. K. Chaudhuri e A. Hocquenghem e são uma excelente generalização dos códigos de Hamming, concedendo a múltipla correção de erros. Deste modo, formam a classe dos melhores códigos construtivos para canais onde os erros afetam os símbolos de forma independente. Contudo, esses códigos são eficientes quando o comprimento das palavras-código não é grande, caso contrário, o desempenho destes é reduzido devido às baixas taxas de transmissão.

Códigos geometricamente uniformes

Definição .11 *Um conjunto de sinais S definido sobre um espaço métrico (M, d) é um código geometricamente uniforme se, dados s_1 e s_2 em S, existe uma isometria u_{s_1,s_2} que transforma s_1 em s_2 mantendo S invariante, ou seja,*

$$u_{s_1,s_2}(s_1) = s_2.$$

Assim, S é geometricamente uniforme se a ação do grupo de simetrias $\Gamma(S)$ de S é transitiva. Se S for finito, dizemos que S é uma constelação uniforme, se S for infinito dizemos que S é um arranjo regular. Uma constelação uniforme no espaço Euclidiano é um código de grupo (código de Slepian).

Conjunto de sinais casados a grupos

A principal justificativa para considerar o codificador e o modulador como um só bloco é instituir a melhor maneira de associar uma palavra-código a um sinal a ser transmitido. Conjunto de sinais casado a um grupo compõe a forma mais apropriada de estabelecer esta associação.

Definição .12 *Seja (M, d) um espaço métrico. Dizemos que um conjunto de sinais finito S em M está casado a um grupo G se existe uma função sobrejetora $\mu : G \to S$, tal que,*

$$d(\mu(g), \mu(g')) = d(\mu(g^{-1} * g'), \mu(e)), \quad \forall g, g' \in G,$$

onde e é o elemento neutro de G. A função μ é denominada mapeamento casado. Se μ é uma função injetora, então μ^{-1} é chamada rotulamento casado.

G-linearidade

G-linearidade é uma extensão da Z_4-linearidade baseada em grupos de simetria. Esta extensão é realizada levando-se em consideração um código quaternário mais como um rotulamento do que a imagem de um código por isometria entre módulos. Foi introduzido este conceito para códigos em espaços métricos em geral. Todos os códigos binários não-lineares estudados são imagens de códigos lineares sobre Z4 por meio de um mapeamento correto.

$\mathbb{Z}_4$-linearidade

A $\mathbb{Z}_4$-linearidade é um conceito inovador e importante em teoria da codificação devido ao fato de facilitar que certas classes de códigos não-lineares de comprimento par possam ser vistas como códigos lineares sobre $\mathbb{Z}_4$. Deste modo, obtém-se uma significativa redução na complexidade do processo de decodificação dos códigos não-lineares.

Algoritmo para identificação de sequências de DNA

Apresentaremos a seguir o algoritmo para identificação de sequências de DNA através de um exemplo.

Exemplo .1 Código $BCH(n, k, d_h) = (63, k, d_h)$ sobre $GR(4, r)$.

Tomando $r = 6$, ou seja, $GR(4, 6)$ teremos o código $BCH(n, k, d_h) = (63, k, d_h)$, pois $n = 2^6 - 1 = 63$, onde n é o comprimento da sequência.

Passo 1: *Especificar a estrutura matemática e o alfabeto do código.*

O alfabeto 4-ário do código genético, $N = \{A, C, G, T/U\}$, se relaciona com o alfabeto 4-ário, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, obedecendo as operações de adição e multiplicação módulo 4, o que lhe confere uma estrutura algébrica de anel.

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

.	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Tabela 1: operações em $\mathbb{Z}_4$

Passo 2: *Determinar a extensão de Galois.*

A condição necessária para que a fatoração de $x^n - 1$ em $GR(4, r)$, grupo das unidades, seja única, é que o comprimento da sequência de DNA seja ímpar da forma $n = 2^r - 1$, onde n é o número de elementos não nulos no corpo de Galois $GF(2^r)$, ou equivalentemente, os elementos que possuem inverso, 2^r é a cardinalidade de $GF(2^r)$ e r o grau da extensão do corpo de Galois. Os valores de r serão utilizados na extensão do corpo $GF(2)$ no passo 4. Nesse exemplo, será utilizada uma sequência de DNA cujo $n = 63$ nucleotídeos. Logo, o grau r do polinômio primitivo a ser usado é $r = 6$, pois $n = 2^r - 1 = 2^6 - 1 = 63$.

Passo 3: *Determinar todos os polinômios primitivos $p(x)$ relacionados à extensão de Galois.*

Nesse passo, os $p(x)$ relacionados ao grau da extensão de Galois $r = 6$ serão encontrados. Em Peterson (1972) encontra-se a seguinte lista de polinômios irredutíveis de grau 6 sobre GF(2).

ordem	1	3	5	7	9	11	21
	103F	127B	147H	111A	015	155E	007

Tabela 2: polinômios irredutíveis de grau 6 sobre GF(2)

onde
A, B, C, D : não primitivo;
E, F, G, H : primitivo;
$0 \to 000$, $1 \to 001$, $2 \to 010$, $3 \to 011$, $4 \to 100$, $5 \to 101$, $6 \to 110$, $7 \to 111$.

Assim, pela Tabela 2 obtemos os polinômios:
$103 = 001000011 \to 1 + x + x^6$ *primitivo;*
$127 = 001010111 \to 1 + x + x^2 + x^4 + x^6$ *não primitivo;*
$147 = 001100111 \to 1 + x + x^2 + x^5 + x^6$ *primitivo;*
$111 = 001001001 \to 1 + x^3 + x^6$ *não primitivo;*
$015 = 000001101 \to 1 + x^3 + x^4$*;*
$155 = 001101101 \to 1 + x^2 + x^3 + x^5 + x^6$ *primitivo;*
$007 = 000000111 \to 1 + x + x^2$.

Dos 3 polinômios primitivos encontrados, obtemos outros três, que são os respectivos polinômios recíprocos dos polinômios dados. Assim, encontramos os 6 polinômios primitivos a seguir:

$1+x+x^6$	$1+x^5+x^6$
$1+x+x^2+x^5+x^6$	$1+x+x^4+x^5+x^6$
$1+x^2+x^3+x^5+x^6$	$1+x+x^3+x^4+x^6$

Tabela 3: polinômios primitivos de grau 6 sobre GF(2)

Passo 4: *Determinar a extensão do corpo* $GF(2)$.

O corpo $GF(2^r)$ *é obtido através do quociente do anel de todos os polinômios com coeficientes em* $GF(2)$, $GF(2)[x]$, *por um ideal gerado por qualquer um dos polinômios primitivos de grau* $r = 6$. *Nesse passo, realizamos a extensão do corpo GF(2) da seguinte maneira:*

Considere o corpo de Galois $GF(2^r) = GF(2^6) = GF(64) = F_{64}$ *dado por*

$$\frac{F_2[x]}{< p(x) >} \cong \frac{F_2[x]}{< 1+x+x^6 >} = \{a_0 + a_1x + a_2x^2 + \ldots + a_5x^5 : a_{i's} \in F_2\},$$

onde $p(x)$ *é o polinômio primitivo do passo 3.*

Seja α *um elemento primitivo em* F_{64}, *equivalentemente* α *é uma raiz de* $1+x+x^6$, *ou seja,* $1+\alpha+\alpha^6 = 0$, *logo* $\alpha^6 = -\alpha - 1$. *Como os coeficientes dos polinômios que formam o conjunto dos elementos de* F_{64} *pertencem a* F_2, *fazemos a redução módulo 2 e obtemos* $\alpha^6 = \alpha + 1$.

Os elementos de F_{64} *são então mostrados na tabela a seguir:*

F_{64}	$(\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$
0	(000000)
1	(100000)
α	(010000)
α^2	(001000)
α^3	(000100)
α^4	(000010)
α^5	(000001)
$\alpha^6 = 1+\alpha$	(110000)
$\alpha^7 = \alpha.\alpha^6 = \alpha+\alpha^2$	(011000)
$\alpha^8 = \alpha.\alpha^7 = \alpha^2+\alpha^3$	(001100)
$\alpha^9 = \alpha.\alpha^8 = \alpha^3+\alpha^4$	(000110)
$\alpha^{10} = \alpha.\alpha^9 = \alpha^4+\alpha^5$	(000011)
$\alpha^{11} = \alpha.\alpha^{10} = \alpha^5+\alpha^6 = 1+\alpha+\alpha^5$	(110001)
$\alpha^{12} = \alpha.\alpha^{11} = \alpha+\alpha^2+\alpha^6 = 1+\alpha^2$	(101000)
$\alpha^{13} = \alpha.\alpha^{12} = \alpha+\alpha^3$	(010100)
$\alpha^{14} = \alpha.\alpha^{13} = \alpha^2+\alpha^4$	(0010101)
$\alpha^{15} = \alpha.\alpha^{14} = \alpha^3+\alpha^5$	(000101)
$\alpha^{16} = \alpha.\alpha^{15} = \alpha^4+\alpha^6 = 1+\alpha+\alpha^4$	(110010)
$\vdots$	$\vdots$
$\alpha^{63} = \alpha.\alpha^{62} = 1$	(10000)

Tabela 4: elementos de F_{64}

Passo 5: *Determinar a extensão de* $\mathbb{Z}_4$.

Considere o anel $GR(p^k, r) = GR(4, 6)$ como sendo dado pelo quociente do anel $Z_(4)[x]$ pelo ideal gerado pelo mesmo $p(x)$ utilizado para realizar a extensão do corpo do Passo 4, isto é,

$$\frac{Z_(4)[x]}{< p(x) >} \simeq \frac{Z_(4)[x]}{< x^6 + x + 1 >} = b_0 + b_1x + b_2x + \ldots + b_5x^5 : b_{i's} \in Z_4$$

Sabemos que as operações em $GR(4, 6)$ são realizadas módulo $(x^6 + x + 1)$. E, como α é uma raiz do polinômio primitivo usado tanto na extensão do corpo como na do anel, então $\alpha^6 = -\alpha - 1$. Agora, como os coeficientes dos polinômios em $GR(4, 6)$ estão em Z_4, então $\alpha^6 = 3\alpha + 3$.
Considerando $\varphi = (010000) = \alpha$, todos os elementos não nulos e inversíveis do grupo cíclico do grupo $GR^*(4, 6)$ são determinados através da potenciação de φ.

Passo 6: *Determinar o grupo das unidades.*

Do passo 5, resulta que φ gera um grupo cíclico de ordem $n.d$ em $GR^*(4, 6)$, onde $d \geq 1 \in \mathbb{Z}$ e φ^d gera o subgrupo cíclico cuja ordem é 63 em $GR^*(4, 6)$. Assim, temos que $n.d = 63.d = 126$, implicando em $d = 2$.
Consequentemente, $\varphi^2 = (001000) = \alpha^2$ gera um subgrupo cíclico de ordem 63 em $GR(4, 6)$. Logo, $\beta = \alpha^2$ é o elemento primitivo que gera o subgrupo cíclico $G_n = G_{63}$, mostrado na tabela:

G_{63}	$(\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$
$\beta = \alpha^2$	(001000)
$\beta^2 = \alpha^4$	(000010)
$\beta^3 = \alpha^6 = 3 + 3\alpha$	(330000)
$\beta^4 = \alpha^8 = 3\alpha^2 + 3\alpha^3$	(003300)
$\beta^5 = \alpha^{10} = 3\alpha^4 + 3\alpha^5$	(000033)
$\beta^6 = \alpha^{12} = 1 + 2\alpha + \alpha^2$	(121000)
$\beta^7 = \alpha^{14} = \alpha^2 + 2\alpha^3 + \alpha^4$	(001210)
$\beta^8 = \alpha^{16} = 3 + 3\alpha + \alpha^4 + 2\alpha^5$	(330012)
$\beta^9 = \alpha^{18} = 3 + \alpha + \alpha^2 + 3\alpha^3$	(311300)
$\beta^{10} = \alpha^{20} = 3\alpha^2 + \alpha^3 + \alpha^4 + 3\alpha^5$	(003113)
$\beta^{11} = \alpha^{22} = 3 + \alpha^2 + 3\alpha^4 + \alpha^5$	(301031)
$\beta^{12} = \alpha^{24} = 1 + 2\alpha^2 + \alpha^4$	(102010)
$\beta^{13} = \alpha^{26} = 3 + 3\alpha + \alpha^2 + 2\alpha^4$	(331020)
$\beta^{14} = \alpha^{28} = 2 + 2\alpha + 3\alpha^2 + 3\alpha^3 + \alpha^4$	(223310)
$\beta^{15} = \alpha^{30} = 3 + 3\alpha + 2\alpha^2 + 2\alpha^3 + 3\alpha^4 + 3\alpha^5$	(332233)
$\beta^{16} = \alpha^{32} = 1 + 2\alpha + 3\alpha^3 + 2\alpha^4 + 2\alpha^5$	(120322)
$\beta^{17} = \alpha^{34} = 2 + 3\alpha^2 + 2\alpha^3 + 3\alpha^5$	(203203)
$\beta^{18} = \alpha^{36} = \alpha + 3\alpha^2 + 3\alpha^4 + 2\alpha^5$	(013032)
$\beta^{19} = \alpha^{38} = 1 + 3\alpha + 2\alpha^2 + \alpha^3 + 3\alpha^4$	(132130)
$\beta^{20} = \alpha^{40} = 1 + \alpha + \alpha^2 + 3\alpha^3 + 2\alpha^4 + \alpha^5$	(111321)
$\beta^{21} = \alpha^{42} = 2 + \alpha + 2\alpha^3 + \alpha^4 + 3\alpha^5$	(210213)
$\beta^{22} = \alpha^{44} = 3 + 3\alpha^2 + \alpha^3 + \alpha^5$	(303101)
$\beta^{23} = \alpha^{46} = 3\alpha + 2\alpha^2 + 3\alpha^4 + \alpha^5$	(032031)
$\beta^{24} = \alpha^{48} = 1 + 3\alpha^2 + 3\alpha^3 + 2\alpha^4$	(103320)
$\beta^{25} = \alpha^{50} = 2 + 2\alpha + \alpha^2 + 3\alpha^4 + 3\alpha^5$	(221033)
$\beta^{26} = \alpha^{52} = 1 + 2\alpha + 3\alpha^2 + 2\alpha^3 + \alpha^4$	(123210)
$\beta^{27} = \alpha^{54} = 3 + 3\alpha + \alpha^2 + 2\alpha^3 + 3\alpha^4 + 2\alpha^5$	(331232)
$\beta^{28} = \alpha^{56} = 1 + 3\alpha + \alpha^2 + 3\alpha^3 + \alpha^4 + 2\alpha^5$	(131312)
$\beta^{29} = \alpha^{58} = 3 + \alpha + 3\alpha^2 + 3\alpha^3 + \alpha^4 + 3\alpha^5$	(313313)
$\beta^{30} = \alpha^{60} = 3 + \alpha^3 + 3\alpha^4 + 3\alpha^5$	(300133)
$\beta^{31} = \alpha^{62} = 1 + 2\alpha + \alpha^5$	(120001)

$\beta^{32} = \alpha^{64} = 3\alpha + 2\alpha^3$	(030200)
$\beta^{33} = \alpha^{66} = 3\alpha^3 + 2\alpha^5$	(000302)
$\beta^{34} = \alpha^{68} = 2\alpha + 2\alpha^2 + 3\alpha^5$	(022003)
$\beta^{35} = \alpha^{70} = \alpha + \alpha^2 + 2\alpha^3 + 2\alpha^4$	(011220)
$\beta^{36} = \alpha^{72} = 2 + 2\alpha + \alpha^3 + \alpha^4 + 2\alpha^5$	(220112)
$\beta^{37} = \alpha^{74} = 3 + \alpha + 2\alpha^3 + \alpha^5$	(310201)
$\beta^{38} = \alpha^{76} = 3\alpha + 2\alpha^2 + \alpha^3 + 2\alpha^5$	(032102)
$\beta^{39} = \alpha^{78} = 2\alpha + 2\alpha^2 + 3\alpha^3 + 2\alpha^4 + \alpha^5$	(022321)
$\beta^{40} = \alpha^{80} = 2 + \alpha + 3\alpha^2 + 2\alpha^3 + 2\alpha^4 + 3\alpha^5$	(213223)
$\beta^{41} = \alpha^{82} = 2 + 3\alpha + 3\alpha^2 + \alpha^3 + 3\alpha^4 + 2\alpha^5$	(233132)
$\beta^{42} = \alpha^{84} = 1 + 3\alpha + 3\alpha^3 + 3\alpha^4 + \alpha^5$	(130331)
$\beta^{43} = \alpha^{86} = 1 + 3\alpha^3 + 3\alpha^5$	(100303)
$\beta^{44} = \alpha^{88} = \alpha + 2\alpha^2 + 3\alpha^5$	(012003)
$\beta^{45} = \alpha^{90} = \alpha + \alpha^2 + \alpha^3 + 2\alpha^4$	(011120)
$\beta^{46} = \alpha^{92} = 2 + 2\alpha + \alpha^3 + \alpha^4 + \alpha^5$	(220111)
$\beta^{47} = \alpha^{94} = 3 + 2\alpha + \alpha^2 + 2\alpha^3 + \alpha^5$	(321201)
$\beta^{48} = \alpha^{96} = 3\alpha + 2\alpha^2 + 2\alpha^3 + \alpha^4 + 2\alpha^5$	(032212)
$\beta^{49} = \alpha^{98} = 3 + \alpha + 2\alpha^2 + 3\alpha^3 + 2\alpha^4 + 2\alpha^5$	(312322)
$\beta^{50} = \alpha^{100} = 2 + \alpha^2 + \alpha^3 + 2\alpha^4 + 3\alpha^5$	(201123)
$\beta^{51} = \alpha^{102} = 2 + 3\alpha + 3\alpha^2 + \alpha^4 + \alpha^5$	(233011)
$\beta^{52} = \alpha^{104} = 3 + 2\alpha + \alpha^2 + 3\alpha^3 + 3\alpha^4$	(321330)
$\beta^{53} = \alpha^{106} = 1 + \alpha + 3\alpha^2 + 3\alpha^3 + \alpha^4 + 3\alpha^5$	(113313)
$\beta^{54} = \alpha^{108} = 3 + 2\alpha^2 + \alpha^3 + 3\alpha^4 + 2\alpha^5$	(302132)
$\beta^{55} = \alpha^{110} = 1 + 3\alpha + \alpha^2 + 2\alpha^4 + \alpha^5$	(131021)
$\beta^{56} = \alpha^{112} = 2 + \alpha + 3\alpha^3 + \alpha^4$	(210310)
$\beta^{57} = \alpha^{114} = 3 + 3\alpha + 2\alpha^2 + \alpha^3 + 3\alpha^5$	(332103)
$\beta^{58} = \alpha^{116} = \alpha + 3\alpha^3 + 2\alpha^4 + \alpha^5$	(010321)
$\beta^{59} = \alpha^{118} = 2 + \alpha + 3\alpha^2 + \alpha^3 + 3\alpha^5$	(213103)
$\beta^{60} = \alpha^{120} = \alpha + 3\alpha^2 + \alpha^3 + 3\alpha^4 + \alpha^5$	(013131)
$\beta^{61} = \alpha^{122} = 1 + 3\alpha^2 + \alpha^3 + 3\alpha^4 + \alpha^5$	(103131)
$\beta^{62} = \alpha^{124} = 1 + 3\alpha^4 + \alpha^5$	(100031)
$\beta^{63} = \alpha^{126} = 1$	(100000)

Tabela 5: elementos de G_{63}

O elemento primitivo $\beta = \alpha^2$ será utilizado na construção de um código BCH de comprimento $n = 63$ sobre Z_4. Quando o comprimento n da palavra-código desejada for igual a cardinalidade de G_n, faremos então a construção de um código BCH primitivo, onde φ gera um grupo cíclico de ordem $n.2$ em $GR(4, r)$.

Passo 7: *Determinar o polinômio gerador da matriz $G, g(x)$.*

Neste passo, serão calculados os polinômios geradores $g(x)$ das matrizes geradoras G dos códigos. Os polinômios geradores dos códigos de comprimento n, tem como raízes os elementos na sequência $\{(\beta^i), (\beta^i)^p, (\beta^i)^{p^2}, (\beta^i)^{p^3}, \ldots, (\beta^i)^{p^{r-1}(mod\ n)}\}$.

Estes polinômios são dados por

$$g(x) = mmc(M_1(x), M_2(x), \ldots, M_{n-1(x)}),$$

onde $M_i(x)$ é o polinômio minimal associado ao elemento β_i, $\{i = 1, 2, \ldots, n-1\}$ (β é um elemento primitivo em G_n).

No caso da palavra-código em questão, cujo comprimento é $n = 63$, os valores de $1 \leq t \leq 31$

serão analisados. Para cada valor de t, teremos uma distância equivalente e seus respectivos polinômios minimais envolvidos nos cálculos dos $g(x)$, da seguinte maneira:

1. *Cálculo das raízes dos polinômios minimais:*

As raízes R_{M_i} dos polinômios minimais $M_i(x)$, para cada $i = 1, 2, \ldots, 62$, são dadas por:

$$R_{M_1} = \{(\beta^1), (\beta^1)^2, (\beta^1)^3, \ldots, (\beta^1)^{2^{6-1}(mod\ 63)}\} = \{\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}\},$$
$$R_{M_2} = \{(\beta^2), (\beta^2)^2, (\beta^2)^3, \ldots, (\beta^2)^{2^{6-1}(mod\ 63)}\} = \{\beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta\},$$
$$\vdots$$
$$R_{M_i} = \{(\beta^i), (\beta^i)^2, (\beta^i)^3, \ldots, (\beta^i)^{2^{6-1}(mod\ 63)}\}.$$

Assim, obtemos a seguinte tabela:

i	R_{M_i}	i	R_{M_i}
1	$\{\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}\}$	32	$\{\beta^{32}, \beta, \beta^2, \beta^4, \beta^8, \beta^{16}\}$
2	$\{\beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta\}$	33	$\{\beta^{33}, \beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48}\}$
3	$\{\beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}\}$	34	$\{\beta^{34}, \beta^5, \beta^{10}, \beta^{20}, \beta^{40}, \beta^{17}\}$
4	$\{\beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta, \beta^2\}$	35	$\{\beta^{35}, \beta^7, \beta^{14}, \beta^{28}, \beta^{56}, \beta^{49}\}$
5	$\{\beta^5, \beta^{10}, \beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}\}$	36	$\{\beta^{36}, \beta^9, \beta^{18}\}$
6	$\{\beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}, \beta^3\}$	37	$\{\beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}\}$
7	$\{\beta^7, \beta^1 4, \beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}\}$	38	$\{\beta^{38}, \beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}$
8	$\{\beta^8, \beta^{16}, \beta^{32}, \beta, \beta^2, \beta^4\}$	39	$\{\beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}\}$
9	$\{\beta^9, \beta^{18}, \beta^{36}\}$	40	$\{\beta^{40}, \beta^{17}, \beta^{34}, \beta^5, \beta^{10}, \beta^{20}\}$
10	$\{\beta^1 0, \beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}, \beta^5\}$	41	$\{\beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}, \beta^{52}\}$
11	$\{\beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}\}$	42	$\{\beta^{42}, \beta^{21}\}$
12	$\{\beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}, \beta^3, \beta^6\}$	43	$\{\beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}\}$
13	$\{\beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}\}$	44	$\{\beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}\}$
14	$\{\beta^{14}, \beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}, \beta^7\}$	45	$\{\beta^{45}, \beta^{27}, \beta^{54}\}$
15	$\{\beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}\}$	46	$\{\beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}\}$
16	$\{\beta^{16}, \beta^{32}, \beta, \beta^2, \beta^4, \beta^8\}$	47	$\{\beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}\}$
17	$\{\beta^{17}, \beta^{34}, \beta^5, \beta^{10}, \beta^{20}, \beta^{40}\}$	48	$\{\beta^{48}, \beta^{33}, \beta^3, \beta^6, \beta^{12}, \beta^{24}\}$
18	$\{\beta^{18}, \beta^{36}, \beta^9\}$	49	$\{\beta^{49}, \beta^{35}, \beta^7, \beta^{14}, \beta^{28}, \beta^{56}\}$
19	$\{\beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}\}$	50	$\{\beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}\}$
20	$\{\beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}, \beta^5, \beta^{10}\}$	51	$\{\beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}\}$
21	$\{\beta^{21}, \beta^{42}\}$	52	$\{\beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}\}$
22	$\{\beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}\}$	53	$\{\beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}\}$
23	$\{\beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}\}$	54	$\{\beta^{54}, \beta^{45}, \beta^{27}\}$
24	$\{\beta^{24}, \beta^{48}, \beta^{33}, \beta^3, \beta^6, \beta^{12}\}$	55	$\{\beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}\}$
25	$\{\beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}\}$	56	$\{\beta^{56}, \beta^{49}, \beta^{35}, \beta^7, \beta^{14}, \beta^{28}\}$
26	$\{\beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}\}$	57	$\{\beta^{57}, \beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}\}$
27	$\{\beta^{27}, \beta^{54}, \beta^{45}\}$	58	$\{\beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}\}$
28	$\{\beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}, \beta^7, \beta^{14}\}$	59	$\{\beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}\}$
29	$\{\beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}\}$	60	$\{\beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}\}$
30	$\{\beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}\beta^{15}\}$	61	$\{\beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}\}$
31	$\{\beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}\}$	62	$\{\beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}\}$

Tabela 6: raízes dos polinômios minimais

Observe que alguns polinômios minimais possuem as mesmas raízes, como podemos ver na tabela a seguir:

$R_{M_1} = R_{M_2} = R_{M_4} = R_{M_8} = R_{M_{16}} = R_{M_{32}}$	$R_{M_{13}} = R_{M_{26}} = R_{M_{52}} = R_{M_{41}} = R_{M_{19}} = R_{M_{38}}$
$R_{M_3} = R_{M_6} = R_{M_{12}} = R_{M_{24}} = R_{M_{48}} = R_{M_{33}}$	$R_{M_{15}} = R_{M_{30}} = R_{M_{60}} = R_{M_{57}} = R_{M_{51}} = R_{M_{39}}$
$R_{M_5} = R_{M_{10}} = R_{M_{20}} = R_{M_{40}} = R_{M_{17}} = R_{M_{34}}$	$R_{M_{21}} = R_{M_{42}}$
$R_{M_7} = R_{M_{14}} = R_{M_{28}} = R_{M_{56}} = R_{M_{49}} = R_{M_{35}}$	$R_{M_{23}} = R_{M_{46}} = R_{M_{29}} = R_{M_{58}} = R_{M_{53}} = R_{M_{43}}$
$R_{M_9} = R_{M_{18}} = R_{M_{36}}$	$R_{M_{27}} = R_{M_{54}} = R_{M_{45}}$
$R_{M_{11}} = R_{M_{22}} = R_{M_{44}} = R_{M_{25}} = R_{M_{50}} = R_{M_{37}}$	$R_{M_{31}} = R_{M_{62}} = R_{M_{61}} = R_{M_{59}} = R_{M_{55}} = R_{M_{47}}$

Tabela 7: polinômios minimais que possuem as mesmas raízes

2. *Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, ..., 62$.*

Como alguns polinômios possuem as mesmas raízes, eles serão iguais. Assim, os polinômios minimais são calculados da seguinte maneira:

$M_1(x) = M_2(x) = M_4(x) = M_8(x) = M_{16}(x) = M_{32}(x) = (x-\beta)(x-\beta^2)(x-\beta^4)(x-\beta^8)(x-\beta^{16})(x-\beta^{32}) = x^6 + 2x^3 + 3x + 1,$

$M_3(x) = M_6(x) = M_{12}(x) = M_{24}(x) = M_{48}(x) = M_{33}(x) = (x-\beta^3)(x-\beta^6)(x-\beta^{12})(x-\beta^{24})(x-\beta^{48})(x-\beta^{33}) = x^6 + 2x^3 + 3x + 1,$

$M_5(x) = M_{10}(x) = M_{20}(x) = M_{40}(x) = M_{17}(x) = M_{34}(x) = (x-\beta^5)(x-\beta^10)(x-\beta^{20})(x-\beta^{40})(x-\beta^{17})(x-\beta^{34}) = x^6 + 2x^3 + 3x + 1,$

$M_7(x) = M_{14}(x) = M_{28}(x) = M_{56}(x) = M_{49}(x) = M_{35}(x) = (x-\beta^7)(x-\beta^14)(x-\beta^{28})(x-\beta^{56})(x-\beta^{49})(x-\beta^{35}) = x^6 + 2x^3 + 3x + 1,$

$M_9(x) = M_{18}(x) = M_{36}(x) = (x-\beta^9)(x-\beta^{18})(x-\beta^{36}) = x^6 + 2x^3 + 3x + 1,$

$M_{11}(x) = M_{22}(x) = M_{44}(x) = M_{25}(x) = M_{50}(x) = M_{37}(x) = (x-\beta^{11})(x-\beta^{22})(x-\beta^{44})(x-\beta^{25})(x-\beta^{50})(x-\beta^{37}) = x^6 + 2x^3 + 3x + 1,$

$M_{13}(x) = M_{26}(x) = M_{52}(x) = M_{41}(x) = M_{19}(x) = M_{38}(x) = (x-\beta^{13})(x-\beta^{26})(x-\beta^{52})(x-\beta^{41})(x-\beta^{19})(x-\beta^{38}) = x^6 + 2x^3 + 3x + 1,$

$M_{15}(x) = M_{30}(x) = M_{60}(x) = M_{57}(x) = M_{51}(x) = M_{39}(x) = (x-\beta^{15})(x-\beta^{30})(x-\beta^{60})(x-\beta^{57})(x-\beta^{51})(x-\beta^{39}) = x^6 + 2x^3 + 3x + 1,$

$M_{21}(x) = M_{42}(x) = (x-\beta^{21})(x-\beta^{42}) = x^6 + 2x^3 + 3x + 1,$

$M_{23}(x) = M_{46}(x) = M_{29}(x) = M_{58}(x) = M_{53}(x) = M_{43}(x) = (x-\beta^{23})(x-\beta^{46})(x-\beta^{29})(x-\beta^{58})(x-\beta^{53})(x-\beta^{43}) = x^6 + 2x^3 + 3x + 1,$

$M_{27}(x) = M_{54}(x) = M_{45}(x) = (x-\beta^{27})(x-\beta^{54})(x-\beta^{45}) = x^6 + 2x^3 + 3x + 1,$

$M_{31}(x) = M_{62}(x) = M_{61}(x) = M_{59}(x) = M_{55}(x) = M_{47}(x) = (x - \beta^{31})(x - \beta^{62})(x - \beta^{61})(x - \beta^{59})(x - \beta^{55})(x - \beta^{47}) = x^6 + 2x^3 + 3x + 1.$

3. Cálculo dos polinômios geradores para $1 \leq t \leq 31$.

O polinômio gerador para cada valor de t *é dado por*

$$g(x) = mmc\{M_1(x), M_2(x), \ldots, M_{n-1}(x)\},$$

formado pelos polinômios minimais que são diferentes entre si e possuem raízes $\beta, \ldots, \beta^{2t}$ *como mostra a tabela a seguir:*

t	$d_H \geq 2t+1$	$(\beta^1, \ldots, \beta^{2t})$	$g(x) = mmc(M_1(x), \ldots, M_{n-1}(x))$	$C(n,k,d_H)$
1	$d_H \geq 3$	(β^1, β^2)	$g(x) = mmc(M_1(x))$	$C(63,57,03)$
2	$d_H \geq 5$	$(\beta^1, \beta^2, \beta^3, \beta^4)$	$g(x) = mmc(M_1(x), M_3(x))$	$C(63,51,05)$
3	$d_H \geq 7$	$(\beta^1, \ldots, \beta^6)$	$g(x) = mmc(M_1(x), M_3(x), M_5(x))$	$C(63,45,07)$
4	$d_H \geq 9$	$(\beta^1, \ldots, \beta^8)$	$g(x) = mmc(M_1(x), M_3(x), M_5(x), M_7(x))$	$C(63,39,09)$
5	$d_H \geq 11$	$(\beta^1, \ldots, \beta^{10})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_9(x))$	$C(63,36,11)$
6	$d_H \geq 13$	$(\beta^1, \ldots, \beta^{12})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{11}(x))$	$C(63,30,13)$
7	$d_H \geq 15$	$(\beta^1, \ldots, \beta^{14})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{13}(x))$	$C(63,24,15)$
8	$d_H \geq 17$	$(\beta^1, \ldots, \beta^{16})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{15}(x))$	$C(63,18,17)$
9	$d_H \geq 19$	$(\beta^1, \ldots, \beta^{18})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{15}(x))$	$C(63,18,19)$
10	$d_H \geq 21$	$(\beta^1, \ldots, \beta^{20})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{15}(x))$	$C(63,18,21)$
11	$d_H \geq 23$	$(\beta^1, \ldots, \beta^{22})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{21}(x))$	$C(63,16,23)$
12	$d_H \geq 25$	$(\beta^1, \ldots, \beta^{24})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{23}(x))$	$C(63,10,25)$
13	$d_H \geq 27$	$(\beta^1, \ldots, \beta^{26})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{23}(x))$	$C(63,10,27)$
14	$d_H \geq 29$	$(\beta^1, \ldots, \beta^{28})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{27}(x))$	$C(63,07,29)$
15	$d_H \geq 31$	$(\beta^1, \ldots, \beta^{30})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{27}(x))$	$C(63,07,31)$
16	$d_H \geq 33$	$(\beta^1, \ldots, \beta^{32})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,33)$
17	$d_H \geq 35$	$(\beta^1, \ldots, \beta^{34})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,35)$
18	$d_H \geq 37$	$(\beta^1, \ldots, \beta^{36})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,37)$
19	$d_H \geq 39$	$(\beta^1, \ldots, \beta^{38})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,39)$
20	$d_H \geq 41$	$(\beta^1, \ldots, \beta^{40})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,41)$
21	$d_H \geq 43$	$(\beta^1, \ldots, \beta^{42})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,43)$
22	$d_H \geq 45$	$(\beta^1, \ldots, \beta^{44})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,45$
23	$d_H \geq 47$	$(\beta^1, \ldots, \beta^{46})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,47)$
24	$d_H \geq 49$	$(\beta^1, \ldots, \beta^{48})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,49)$
25	$d_H \geq 51$	$(\beta^1, \ldots, \beta^{50})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,51)$
26	$d_H \geq 53$	$(\beta^1, \ldots, \beta^{52})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,53)$
27	$d_H \geq 55$	$(\beta^1, \ldots, \beta^{54})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,55)$
28	$d_H \geq 57$	$(\beta^1, \ldots, \beta^{56})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,57)$
29	$d_H \geq 59$	$(\beta^1, \ldots, \beta^{58})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,59)$
30	$d_H \geq 61$	$(\beta^1, \ldots, \beta^{60})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,61)$
31	$d_H \geq 63$	$(\beta^1, \ldots, \beta^{62})$	$g(x) = mmc(M_1(x), M_3(x), \ldots, M_{31}(x))$	$C(63,01,63)$

Tabela 8: polinômios geradores

Considerando que a distância mínima do código seja $d_H = 3$, então o polinômio minimal será dado por $M_1(x) = g_1(x) = x^6 + 2x^3 + 3x + 1$, que gera o código desejado e está relacionado com a matriz geradora G do código BCH sobre $\mathbb{Z}_4$ com parâmetros $(n, k, d_H) = (63, 57, 3)$.

De maneira análoga, os demais polinômios geradores para outros valores de t correspondentes a outras distâncias são determinados de acordo com a tabela anterior.

Passo 8: *Determinar o polinômio gerador da matriz H, $h(x)$.*

O polinômio gerador da matriz verificação de paridade H é obtido através da relação:

$$h(x) = \frac{x^{n-1}}{g(x)} = \frac{x^{63} - 1}{x^6 + 2x^3 + 3x + 1}.$$

Assim, $h(x) = x^{57} + 2x^{54} + x^{52} + 3x^{51} + x^{47} + 2x^{46} + x^{45} + 2x^{44} + 3x^{42} + x^{41} + 3x^{40} + 3x^{39} + x^{37} + 2x^{35} + 2x^{34} + x^{33} + x^{32} + 3x^{31} + 2x^{29} + x^{28} + 2x^{26} + 3x^{25} + 2x^{24} + 3x^{23} + x^{22} + 2x^{21} + 3x^{20} + x^{19} + x^{18} + 3x^{16} + 3x^{15} + 2x^{14} + 2x^{13} + x^{12} + 3x^{11} + x^9 + 2x^8 + 3x^7 + x^5 + 3x^4 + 3x^2 + 3x + 3$, *onde os coeficientes do polinômio $h(x)$ pertencem a $\mathbb{Z}_4$.*

Passo 9: *Determinar a matriz G e a sua transposta G^T.*

O polinômio gerador $g_1(x) = 1 + 3x + 2x^3 + x^6$ está relacionado à matriz geradora G. Realizando os deslocamentos dos coeficientes do polinômio $g(x)$ da esquerda para a direita, obtemos a matriz G de dimensões 57×63:

$$G = \begin{pmatrix}
1 & 3 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 3 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 3 & 0 & 2 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 3 & 0 & 2 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & & & & & & & & & & \dots & & & & & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 3 & 0 & 2 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 3 & 0 & 2 & 0 & 0 & 1
\end{pmatrix}$$

A matriz G^T com dimensões 63×57 é determinada como sendo a troca da linha pela coluna.

Passo 10: *Determinar a matriz H e a sua transposta H^T.*

Determinando o polinômio $h(x)$ no passo 8, realizamos os deslocamentos dos coeficientes do polinômio gerador $h(x)$ da direita para a esquerda e obtemos a matriz H com dimensão 6×63:

$$H = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & \dots & 2 & 3 & 0 & 1 & 3 & 0 & 3 & 3 & 3 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 1 & \dots & 3 & 0 & 1 & 3 & 0 & 3 & 3 & 3 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 1 & 3 & \dots & 0 & 1 & 3 & 0 & 3 & 3 & 3 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 2 & 0 & 1 & 3 & 0 & \dots & 1 & 3 & 0 & 3 & 3 & 3 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 2 & 0 & 1 & 3 & 0 & 0 & \dots & 3 & 0 & 3 & 3 & 3 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 2 & 0 & 1 & 3 & 0 & 0 & 0 & \dots & 0 & 3 & 3 & 3 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

A matriz H^T com dimensão 63×6 é determinada pela troca da linha pela coluna.

Passo 11: *Rotular a sequência de DNA utilizando o passo 1.*

Neste exemplo, analisaremos se o código BCH primitivo sobre $GR(4,6)$ é capaz de reproduzir a sequência de sinal interno ($SI = seq.\ 36$) de uma proteína mitocondrial - GI número 832917 com comprimento $n = 63$ nucleotídeos. Este passo determina as 24 permutações entre o alfabeto do código genético $N = \{A, C, G, T/U\}$ e o alfabeto do código BCH, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, da sequência de DNA a ser analisada, no exemplo em questão. Uma vez que o mapeamento entre $N \to \mathbb{Z}_4$ não é conhecido, consideramos todas as permutações entre esses dois conjuntos. As 24 linhas da matriz correspondem as 24 permutações da SI.

Seja a SI do NCBI igual a

$$\{5' - GCCGTTCATGTTTACTCTGGGTTGCCTTGGTGGGGAACTATCGCGGCCACCACCATCCTCATT - 3'\}$$

então obtemos a seguinte matriz P, onde cada linha está relacionada a cada uma das 24 permutações entre $N \to \mathbb{Z}_4$.

$$P = \begin{pmatrix} 2 & 1 & 1 & 2 & 3 & 3 & 1 & 0 & 3 & 2 & \dots & 0 & 3 & 1 & 1 & 3 & 1 & 0 & 3 & 3 \\ 3 & 1 & 1 & 3 & 2 & 2 & 1 & 0 & 2 & 3 & \dots & 0 & 2 & 1 & 1 & 2 & 1 & 0 & 2 & 2 \\ 1 & 2 & 2 & 1 & 3 & 3 & 2 & 0 & 3 & 1 & \dots & 0 & 3 & 2 & 2 & 3 & 2 & 0 & 3 & 3 \\ \vdots & & & & & & & & & & \dots & & & & & & & & & \\ 0 & 2 & 2 & 0 & 1 & 1 & 2 & 3 & 1 & 0 & \dots & 3 & 1 & 2 & 2 & 1 & 2 & 3 & 1 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 & 2 & 3 & 0 & 1 & \dots & 3 & 0 & 2 & 2 & 0 & 2 & 3 & 0 & 0 \end{pmatrix}$$

Passo 12: *Verificar se a sequencia de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(a,b) = 0$, $D(a,b) = 1$ e $D(a,b) = 2$.*

O procedimento usado para determinar quais das sequencias são palavras-código dos códigos $(63, k, d_h)$, é o mesmo adotado na construção dos códigos sobre corpos, da seguinte maneira:

As linhas da matriz geradora do código (n, k, d_h) do passo 9, formam uma base do espaço vetorial identificado como o código linear C. Assim, as combinações lineares das linhas de G, são palavras-código de C. Dessa forma, o processo de codificação pode ser escrito como $v = u.G$, onde u é a informação e v é a palavra-código correspondente, no nosso caso, as sequências de DNA a serem analisadas. Para toda palavra-código v vale a relação

$$v.H^T = 0$$

A capacidade de correção de erros de um código está relacionada com o número de palavras-código: no caso em questão temos $4k$ palavras-código, onde $k = n - r$.

Nota-se que, quanto maior for o valor de k, maior será o número de palavras-código implicando assim, em uma maior complexidade computacional para gerar todas as $4K$ palavras-código. Para contornarmos esse problema, que é classificado como um problema $NP-completo$, ao invés de gerarmos todas as palavras-código para compararmos coma a sequência de DNA, consideramos que a sequência sob a aplicação de cada uma das 24 permutações do passo 12 é uma palavra-código, que corresponde ao padrão de erro denotado por $D(a,b) = 0$, ou seja, a diferença de DNA do $NCBI$ é nula. Assim, para determinarmos se cada uma dessas 24 possibilidades é de fato uma palavra-código, usamos a relação $v.H^T = 0$, onde v é a possível palavra-código e H^T é a transposta da matriz verificação de paridade determinada no passo 8.

a. $D(a,b) = 1$ *nucleotídeo de diferença.*

Para analisarmos as sequencias de DNA que apresentam o padrão de erro $D(a,b) = 1$*, consideramos as 3 outras possibilidades de nucleotídeos em cada posição na sequência de DNA para cada permutação, resultando em um total de 3 possibilidades de nucleotídeos em cada posição, multiplicados pelo comprimento da sequência* n *e pelas 24 possibilidades de permutações, neste exemplo:*

$3 \times 63 \times 24 = 4536$ *possibilidades para cada sequência de DNA, então usamos a equação:* $v.H^T = 0$.

As palavras-código encontradas são armazenadas.

Como resultado da geração de $D(a,b) = 1$ *nucleotídeos de diferença, obtemos a matriz* R *onde cada linha é uma palavra-código encontrada.*

$$R = \begin{pmatrix} 1 & 2 & 2 & 1 & 3 & 3 & 2 & 0 & 3 & 1 & \dots & 0 & 3 & 2 & 2 & 3 & 2 & 0 & 3 & 3 \\ 3 & 2 & 2 & 3 & 1 & 1 & 2 & 0 & 1 & 3 & \dots & 0 & 1 & 2 & 2 & 1 & 2 & 0 & 1 & 1 \\ 0 & 3 & 3 & 0 & 2 & 2 & 3 & 1 & 2 & 0 & \dots & 1 & 2 & 3 & 3 & 2 & 3 & 1 & 2 & 2 \\ \vdots & & & & & & & & & & \dots & & & & & & & & & \\ 0 & 1 & 1 & 0 & 2 & 2 & 1 & 3 & 2 & 0 & \dots & 3 & 1 & 2 & 2 & 1 & 2 & 3 & 1 & 1 \\ 2 & 1 & 1 & 2 & 0 & 0 & 1 & 3 & 0 & 2 & \dots & 3 & 0 & 1 & 1 & 0 & 1 & 3 & 0 & 0 \end{pmatrix}$$

b. $D(a,b) = 2$ *nucleotídeos de diferença.*

Consideramos todas as combinações simples tomados 2 a 2 dos n *nucleotídeos de comprimento da sequência, isto é*

$$C_{(n,m)} = \frac{n!}{m!(n-m)!}$$

Neste exemplo, $n = 63$ *e* $m = 2$*, temos* 1953×9 *possibilidades em* 17577 *palavras-código para cada caso de permutação de cada sequência de DNA analisada. As palavras-código também são armazenadas.*

Como resultado da geração de $D(a,b) = 2$ *nucleotídeos de diferença, obtemos várias matrizes diferentes entre si, por exemplo,* R'*, onde cada linha é uma palavra-código encontrada.*

$$R' = \begin{pmatrix} 2 & 1 & 1 & 2 & 3 & 3 & 1 & 0 & 3 & 2 & \dots & 0 & 3 & 1 & 1 & 3 & 1 & 0 & 3 & 3 \\ 3 & 1 & 2 & 3 & 2 & 2 & 1 & 0 & 2 & 3 & \dots & 0 & 2 & 1 & 1 & 2 & 1 & 0 & 2 & 2 \\ 3 & 2 & 2 & 1 & 3 & 3 & 2 & 0 & 3 & 1 & \dots & 0 & 3 & 2 & 2 & 3 & 2 & 0 & 3 & 3 \\ \vdots & & & & & & & & & & \dots & & & & & & & & & \\ 0 & 2 & 1 & 0 & 1 & 1 & 2 & 3 & 1 & 0 & \dots & 3 & 1 & 2 & 2 & 1 & 2 & 3 & 1 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 & 2 & 3 & 0 & 1 & \dots & 3 & 0 & 2 & 2 & 0 & 2 & 3 & 0 & 0 \end{pmatrix}$$

Passo 13: *Comparar todas as palavras-código armazenadas no passo 12 com a sequência de DNA original e mostrar onde os erros ocorreram.*

Neste passo, todas as palavras-código armazenadas no passo anterior estão rotuladas na forma do alfabeto do código, $\mathbb{Z}_4 = \{0,1,2,3\}$*, e serão convertidas em nucleotídeos usando*

o rotulamento do alfabeto do código genético $N = \{A, C, G, T\}$. Após o rotulamento, as palavras código são comparadas, uma-a-uma, com a sequência de DNA original mostrando onde os nucleotídeos diferem, e armazena os resultados.

Passo 14: *Voltar para o passo 7 e determinar outro $g(x)$.*

Neste passo, são determinados outros valores da distância mínima d_H, e utiliza-se o mesmo procedimento, apresentado no passo 7, para calcular o polinômio gerador relativo a esta distância.

Passo 15: *Repetir os passos 8 ao 12 para o $g(x)$ obtido no passo 14, até que se esgotem todas as possibilidades de $g(x)$.*

Passo 16: *Voltar ao passo 3 e escolher outro $p(x)$, repetir os passos de 4 ao 14 até se esgotarem todas as possibilidades de $p(x)$ do passo 3.*

Passo 17: *Fim.*

Análise Resultados

Como resultado do algoritmo de geração da sequência de DNA para $D(a, b) = 1$ nucleotídeo de diferença da sequência do NCBI, foram obtidas 8 palavras-código, sendo 1 palavra código para cada caso de permutação que corresponde aos 8 casos de permutações, representados através da matriz R do passo 12. Essas palavras-cóodigo são diferentes em termos do alfabeto do código, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, porém são iguais quando rotuladas usando o alfabeto do código genético $N = \{A, C, G, T\}$, resultando em uma única sequência de DNA. Como as sequências que foram reproduzidas são iguais, é suficiente considerarmos a sequência correspondente a um dos casos, mostrada na figura a seguir.

Seq.36 | *S. cerevisiae* **- OXA 1 – Sinal interno – GI número 832917**

Código klein-linearidade((63,57,3) BCH primitivo sobre *GR*(4,6), rotulamento C)

$p_1(x)=x^6+x+1$ - $g_1(x)=x^6+2x^3+3x+1$ - Caso 3-(A,C,G,T)=(0,2,1,3)

```
aaO:  A   V   H   V   Y   S   G   L   P   W   W   G   T   I   A   A   T   T   I   L   I
ntO: GCC GTT CAT GTT TAC TCT GGG TTG CCT TGG TGG GGA ACT ATC GCG GCC ACC ACC ATC CTC ATT
RtO: 122 133 203 133 302 323 111 331 223 311 311 110 023 032 121 122 022 022 032 232 033
RtG: 122 133 203 133 302 323 111 331 223 111 311 110 023 032 121 122 022 022 032 232 033
ntG: GCC GTT CAT GTT TAC TCT GGG TTG CCT GGG TGG GGA ACT ATC GCG GCC ACC ACC ATC CTC ATT
aaG:  A   V   H   V   Y   S   G   L   P   G   W   G   T   I   A   A   T   T   I   L   I
```

Figura 1: Sequência de DNA de sinal interno com 63 nucleotídeos e $D(a, b) = 1$.

A Figura mostra a sequência de SI (Seq.36) reproduzida pelo código BCH primitivo ((63,57,3) sobre GR(4, 6), com o rotulamento dado através do polinômio primitivo $1+x+x^6$ e do polinômio gerador $1 + 3x + 2x^3 + x^6$.

Observe que, na posição da décima trinca, ocorreu uma troca de nucletídeo ocasionando a troca do aminoácido nesta posição, sendo o Triptofano (W) substituído pela Glicina (G), indicado por (W10G). Essa alteração de aminoácidos implica na mudança de classe, da classe de aminoácido não-polar para classe de aminoácido polar. Dessa forma, observamos que os resultados dos rotulamentos apresentados em Rocha (2010) para sequências de direcionamento com $D(a, b) = 1$ nucleotídeo de diferença da sequência do NCBI foram confirmados para outras sequências de DNA.

Conclusões

Pode-se observar que através da teoria de códigos corretores de erros é possível localizar onde ocorreu uma mutação negativa em uma molécula de DNA, mas, diferentemente dos códigos, ainda não conseguimos alterá-la. Contudo, estudos mais aprofundados sobre esse tema estão sendo feitos, a fim de tornar possível reestabelecer a estrutura molecular normal frente a um DNA mutado.

References

BATTAIL, G. *An Outline of Informational Genetics*. Morgan e Claypool Publishers, 2008.

SHANNON, C.E. *A Mathematical Theory of Communications*. BSTJ 27, 1948.

ROCHA, A. S. L. *Modelo de Sistema de Comunicações Digital para o Mecanismo de Importação de Proteínas Mitocondriais Através de Códigos Corretores de Erros*, 2010. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas-SP.

FARIA, L. C. B. *Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA*, 2011. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas-SP.

FARIA, L. C. B. *Caracterização Topológica, Geométrica e Algébrica dos Produtos da Recombinação do DNA através dos Modelos Tangle e Frações Contínuas*, 2004. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas-SP.

ROCHA, A. S. L. *Modelos Matemáticos para a Previsão de Recombinações Sítio-Específica do DNA*, 2004. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas-SP.

FARIA, L.C.B.; ROCHA, A.S.L.; KLEINSCMIDT, J.H.; PALAZZO Jr., R.; Silva FILHO, M.C. DNA sequences generated by BCH codes over GF(4). *Electronics Letters*, v.46, n. 3, p. 203-204, fevereiro de 2010.

PETERSON, W.W.; WELDON JR, E.J. *Error-Correcting Codes*, 2nd ed. MIT Press, 1972.

Agradecimentos

Agradecemos ao apoio concedido pela FAPEMIG.