

## Aplicações de Estruturas Algébricas na Conexão entre Sistemas de Comunicação Padrão e Genética

**Resumo:** *Um dos grandes desafios para a comunidade científica é analisar a existência de uma estrutura matemática relacionada com o DNA e com um sistema de comunicação. Este artigo tem como objetivo identificar a aplicação de estruturas algébricas na conexão entre os dogmas centrais da teoria de comunicações e da biologia molecular. Tendo em vista o objetivo traçado, foi necessário um conhecimento prévio de biologia celular e molecular, bem como conceitos de álgebra abstrata e sistemas de comunicação, para a associação algébrica do código genético e um sistema de comunicação. Neste estudo, foram analisados os modelos de sistemas de comunicação genética propostos por Gatlin (1972), May et al. (2004), Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011). Por meio dessa associação, percebe-se uma interessante conexão entre elementos biológicos (código genético), engenharia (constelação de sinais em um processo de transmissão de informação) e elementos de álgebra (extensão de Galois  $GF(2^6)$ ). Os resultados apresentados neste trabalho contribuem em uma área de pesquisa em franca expansão, fazendo com que a biologia se torne uma ciência fundamentada teoricamente em conceitos matemáticos.*

**Palavras-chave:** Código genético; extensões de Galois; modelos de comunicação; analogias.

**Abstract:** *One of the big challenges for the scientific community is to analyze the existence of a mathematical structure related to DNA and a communication system. This article aims to identify the application of algebraic structures in the connection between the central dogmas of communications theory and molecular biology. By considering the outlined aim, it was necessary a previous knowledge of cellular and molecular biology, as well as concepts of abstract algebra and communication systems, for the algebraic association with the genetic code in a communication system. In this study, it was analyzed the models of genetic communication systems proposed by Gatlin (1972), May et al. (2004), Rocha and Palazzo Júnior (2010) and Faria and Palazzo Júnior (2011). Through this association, one can perceive, an interesting connection between biological elements (genetic code), engineering (signals constellation in an information transmission process) and elements of algebra (Galois extension  $GF(2^6)$ ). The results presented in this paper contribute to an area of research that is expanding, making biology become a science theoretically based on mathematical concepts.*

**Keywords:** genetic Code; Galois extensions; models of communication; analogies.

## 35 Introdução

36 Embora aparentemente não apresentem relação, tanto a teoria de comunicações quanto a  
37 genética preocupam-se com a transferência de informação. Levando em consideração o fato de  
38 que a teoria de comunicações é realizada pelo homem, e a genética por um processo natural,  
39 ambas diferem entre si, principalmente por não operarem na mesma dimensão. Sendo assim,  
40 enquanto a teoria de comunicações está programada para enviar mensagens no espaço, de um  
41 lugar para outro, a genética envia mensagens hereditárias no tempo (BATTAIL, 2008).

42 Neste estudo, buscamos identificar a aplicabilidade de estruturas algébricas na conexão entre  
43 um sistema de comunicação padrão e um sistema de comunicação genética. Um dos grandes  
44 desafios para a comunidade científica é analisar a existência de uma estrutura matemática re-  
45 lacionada com o DNA e com o sistema de comunicação. Existem várias vertentes que tratam  
46 sobre este tema, em especial, Shannon (1948), matemático americano, engenheiro eletrônico e  
47 criptógrafo que ficou conhecido como o “pai da teoria da informação”. Ele foi o precursor para  
48 a teoria de códigos e a teoria de comunicação.

49 Historicamente, na década de 1970, iniciou-se a aplicação da teoria da informação para  
50 análise de dados genéticos, porém sem muito sucesso. Após alguns anos, este estudo despertou  
51 novamente o interesse da aplicação da teoria de comunicações com o estudo do genoma. Com  
52 isso, muitos pesquisadores, assim como May et al. (2004), têm explorado o dogma central da  
53 genética a partir de um ponto de vista da transmissão de informação, cuja premissa central é  
54 a de que os genes sejam perpetuados sob a forma de sequências de ácidos nucleicos e gerados  
55 como proteínas.

56 Há um grande número de ferramentas matemáticas, principalmente algébricas, que são uti-  
57 lizadas para estudar diversos fenômenos biológicos. Os estudos de Rocha (2010), Faria (2011)  
58 e May (2004) são algumas das pesquisas recentes que têm focado nesta questão, com ênfase  
59 na proposta de modelos para o sistema de comunicação de informação genética e genômica,  
60 análogos ao sistema de comunicação digital. Assim, a aplicabilidade das estruturas algébricas  
61 tem sido alvo de diversos estudos, cujo intuito é representar matematicamente o que o mundo  
62 biológico realiza (OLIVEIRA, 2012).

63 Desta forma, podemos aplicar estruturas algébricas na conexão entre um sistema de co-  
64 munição padrão e um sistema de comunicação biológico. Esta afirmação pode ser mostrada  
65 a partir de respostas à seguinte questão: por meio das estruturas algébricas, como é possível  
66 estabelecer conexões entre um sistema de comunicação padrão e um sistema de comunicação  
67 biológico?

68 Sendo assim, o propósito deste artigo é identificar a aplicação de estruturas algébricas na  
69 conexão entre os dogmas centrais da teoria de comunicações e da biologia molecular. Além disso,  
70 pretende-se compreender conceitos de biologia celular e molecular, compreender conceitos de  
71 álgebra abstrata como: grupos, anéis, corpos, corpos de Galois e suas propriedades, relacionar  
72 o sistema de comunicação com a estrutura do DNA e o processo de geração de proteínas e  
73 relacionar as analogias entre um sistema de comunicação padrão e um sistema de comunicação  
74 genética, a partir de alguns modelos propostos na literatura.

## 75 Preliminares

### 76 Elementos de Biologia

77 A célula é a menor porção de matéria com vida do organismo, são microscópicas e só con-  
78 seguem ser visualizadas com aparelhos especiais. Foi descoberta por Robert Hooke, em 1665.  
79 Ele as descobriu quando examinava no microscópio um pedaço de cortiça e percebeu que ele  
80 era composto por compartimentos, os quais designou células. As principais funções das células  
81 são: captar nutrientes, utilizando-os na produção de energia, eliminar resíduos do metabolismo

82 e capacidade de movimentação (em alguns casos). As células se dividem em dois grupos, proca-  
83 riontes <sup>i</sup> e eucariontes <sup>ii</sup>.

84 De acordo com (ALBERTS et al., 2010) a origem das células está diretamente relacionada  
85 com o surgimento da vida em nosso planeta. Estudos com fósseis determinaram que as células  
86 procarióticas surgiram a aproximadamente  $3,5 - 3,0 \times 10^9$  anos e as eucarióticas a aproxima-  
87 damente  $1,5 \times 10^9$  anos.

## 88 Nucleotídeos e Ácidos Nucleicos

89 Os ácidos nucleicos são macromoléculas de suma importância biológica. São encontradas em  
90 todas as células vivas, atribuindo-se aos mesmos as importantes funções de conter, transmitir  
91 e traduzir as informações genéticas de um determinado organismo. Vale ressaltar que, a vida,  
92 como a conhecemos está intimamente relacionada à química dos nucleotídeos e dos ácidos nu-  
93 cleicos. Portanto, a maioria dos organismos vivos possuem ácidos nucleicos na forma de ácido  
94 desoxirribonucleico (DNA) e ácido ribonucleico (RNA).

95 O DNA é o grande responsável pelo armazenamento da informação genética, através de uma  
96 sequência de bases (adenina, timina, citosina e guanina) que estão interligadas por pontes de  
97 hidrogênio e também é composto por um grupo fosfato e uma pentose (desoxirribose), sendo  
98 que pode ser transcrita ou traduzida. Ela é transcrita em moléculas de RNA que depois são  
99 traduzidas em proteínas, por meio dos aminoácidos. As bases nitrogenadas são unidas por meio  
100 de pontes de hidrogênio, sendo estas complementares, onde sempre a adenina se liga à timina,  
101 por meio de duas pontes de hidrogênio e a citosina se liga à guanina, por meio de três pontes de  
102 hidrogênio. Esta complementariedade foi descoberta por Erwin Chargaff, em meados da década  
103 de 40.

## 104 A Molécula de DNA

105 O DNA é um ácido desoxirribonucleico, cuja estrutura mais difundida é a dupla hélice  
106 de nucleotídeos. Segundo (PAMPHILE e VICENTINI (2011)) o DNA descreve as estruturas  
107 primárias de todas as proteínas, RNAs e, por meio das enzimas, afeta diretamente todos os  
108 processos constituintes celulares, como forma, tamanho e função dos seres vivos.

109 O modelo da molécula de DNA elaborado por Watson e Crick em 1953, foi muito signi-  
110 ficativo para o desenvolvimento de pesquisas na área da Biologia Molecular e suas principais  
111 características são:

- 112 1. Duas cadeias polinucleotídicas (chamadas fitas), formando a dupla hélice.
- 113 2. As duas fitas de DNA são antiparalelas.
- 114 3. As bases nitrogenadas ocupam o centro da hélice e as cadeias de açúcar-fosfato estão na  
115 periferia.
- 116 4. Cada base está ligada a outra por meio de pontes de hidrogênio, formando um par de  
117 base planar, onde adenina se liga com timina e vice-versa e guanina se liga com citosina,  
118 e vice-versa.

119 O processo de duplicação (ou replicação) do DNA ocorre por meio da presença da enzima  
120 DNA *polimerase*. Com a presença dessa enzima, ocorrem as seguintes etapas:

---

<sup>i</sup> Células procariontes: sua principal característica é a ausência de núcleo (carioteca) e da maioria das organelas, tais como: mitocôndria, cloroplasto e complexo de Golgi. Aqui estão inclusas as bactérias e as cianobactérias (algas azuis).

<sup>ii</sup> Células eucariontes: apresentam um núcleo, um compartimento bem definido onde o material genético fica isolado do citoplasma. São células mais complexas que as procarióticas e, existem estudos que abordam que a célula eucarionte é um aperfeiçoamento de células procariontes. Possuem membrana nuclear individualizada e vários tipos de organelas. Aqui estão inclusos os animais, vegetais, fungos e leveduras, que são seres pluricelulares.

- 121 1. Primeiramente, ocorre o rompimento das pontes de hidrogênio, que ligam as bases nitro-  
122 genadas e as duas fitas se afastam.
- 123 2. Os nucleotídeos livres na célula se encaixam nas duas fitas que se afastaram, respeitando  
124 a complementariedade da Regra de Chargaff.
- 125 3. Quando as duas fitas originais tiverem sido completadas por novos nucleotídeos, obtemos  
126 duas moléculas de DNA idênticas entre si.

127 A transcrição consiste no processo em que o RNA é sintetizado a partir de um molde de  
128 DNA, ou seja, o DNA produz moléculas de RNAm e, tais moléculas migram para o citoplasma  
129 e controlam a construção das proteínas. Uma importante diferença em relação ao processo de  
130 duplicação do DNA é que uma das fitas do DNA serve de molde para produzir a molécula de  
131 RNAm. Sendo assim, o RNA produzido será, portanto, uma fita simples e não dupla. Tal  
132 processo ocorre através das seguintes etapas:

- 133 1. É necessária a presença da enzima RNA polimerase.
- 134 2. As duas fitas de DNA se afastam, quando as pontes de hidrogênio se desfazem.
- 135 3. Denominamos por fita ativa, o encaixe dos nucleotídeos livres do RNA numa das fitas.
- 136 4. A molécula de RNA destaca-se de seu molde de DNA e migra para o citoplasma.
- 137 5. Através do pareamento das duas fitas de DNA, reconstitui-se a molécula original.

138 O DNA presente no núcleo da célula controla toda a síntese proteica. Esse controle é efetuado  
139 através do RNA que o DNA fabrica durante o processo de transcrição e envia para o citoplasma.

140 A correspondência entre o RNA e o DNA ocorre base por base, respeitando a complementa-  
141 riedade dos nucleotídeos, ou seja, quando há adenina no DNA, entra a uracila no RNA; quando  
142 há timina no DNA, entra a adenina no RNA, e assim por diante.

143 Já a correspondência entre o RNA e a proteína ocorre de modo diferente. Cada três bases  
144 de RNA (trincas), codificam um aminoácido específico da proteína. A tradução é o processo de  
145 síntese ou fabricação de proteínas (construção da cadeia de aminoácidos). Denominamos por  
146 código genético, a correspondência entre trincas de bases do DNA, RNA e aminoácidos. Cada  
147 trinca de DNA ou RNA é denominada códon<sup>iii</sup>, onde as trincas representam “palavras” do código  
148 genético, cada “palavra” corresponde a um “objeto”, no caso, o aminoácido. O código genético  
149 possui 64 trincas, correspondentes a 20 aminoácidos. Assim, percebe-se que mais de um códon  
150 pode codificar o mesmo aminoácido.

151 Na Figura 1, é apresentado um esquema básico da síntese proteica. É importante destacar  
152 que o processo visto no esquema ocorre de maneira diferente entre as células eucarióticas e  
153 procarióticas. Nas eucarióticas, a transcrição ocorre no núcleo e a tradução no citoplasma, já as  
154 procarióticas tanto a transcrição quanto a tradução ocorrem bem próximas umas das outras.

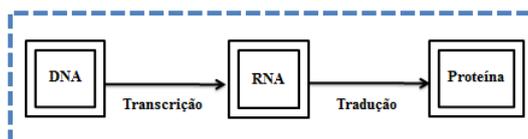


Figura 1: Esquema da síntese proteica.

<sup>iii</sup>Códon: sequência de três nucleotídeos do RNAm a serem traduzidos em um único aminoácido.

155 O DNA de cada célula possui as informações genéticas do indivíduo, essas informações são  
 156 passadas através da transcrição e tradução. No processo de transcrição todos os RNA's são  
 157 feitos a partir de um DNA.

158 A síntese proteica ocorre a partir do DNA, mas não é ele que sintetiza, pois como inter-  
 159 mediário temos o RNAm. A sequência do RNAm é lida e assim são sintetizadas muitas proteínas,  
 160 é esse o processo de tradução, que ocorre nos ribossomos.

161 As proteínas são polímeros formados através da combinação aleatória de 20 aminoácidos.  
 162 Existem 64 possíveis combinações, utilizando as quatro bases nitrogenadas do RNA ( $U, C, A, G$ )  
 163 agrupadas três-a-três, caracterizando a especificação dos 20 distintos aminoácidos, denominado  
 164 código genético.

165 A síntese proteica é feita em blocos de três nucleotídeos, que são os códon, onde cada  
 166 códon é responsável por um aminoácido específico. A sequência correspondente do códon para  
 167 um aminoácido é o que denominamos de código genético, apresentado na Tabela 1 com seus  
 168 respectivos aminoácidos, cujas abreviações são: asparigina (Asp), glutamina (Gln), arginina  
 169 (Arg), histidina (His), lisina (Lys), cisteína (Cys), glicina (Gly), serina (Ser), treonina (Thr),  
 170 ácido aspártico (Asp), ácido glutamânico (Gln), tirosina (Tyr), alanina (Ala), leucina (Leu),  
 171 valina (Val), isoleucina (Ile), prolina (Pro), fenilalanina (Phe), metiotina (Met) e triptofano  
 172 (Trp).

Tabela 1: O código genético.

Primeira Base	Segunda Base				Terceira Base
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

## 173 Elementos de Álgebra

174 **Definição:** Um conjunto  $G \neq \emptyset$  com uma operação  $*$  onde:

$$175 \quad \gamma: G \times G \longrightarrow G$$

$$176 \quad (a, b) \longmapsto a * b,$$

177 munido das seguintes propriedades:

178 1. **associativa:**  $a * (b * c) = (a * b) * c, \forall a, b, c \in G,$

179 2.  $\exists$  **elemento neutro**  $e$  para a operação  $*$ :  $e * a = a = a * e, \forall a \in G,$

180 3.  $\forall a \in G, \exists b$  (**elemento inverso**):  $a * b = b * a = e, \forall a, b \in G,$

181 é um grupo.

182

183 **Definição:** Um grupo  $(G, *)$  é chamado de **grupo comutativo** (ou grupo abeliano, em honra  
184 ao matemático norueguês N.H. Abel–(1802-1829)), se a operação binária  $*$  satisfaz a seguinte  
185 condição: para quaisquer  $a$  e  $b$  em  $G$ ,  $a * b = b * a$ .

186

187 **Definição:** Um sistema matemático constituído de um conjunto não vazio  $A$  ( $A \neq \emptyset$ ) onde este-  
188 jam definidas duas operações (adição e multiplicação), denotados por “+” e “\*” respectivamente,  
189 onde:

190

$$+ : A \times A \longrightarrow A$$

191

$$(a, b) \longmapsto a + b.$$

192

$$* : A \times A \longrightarrow A$$

193

$$(a, b) \longmapsto a * b.$$

194 e é munido das seguintes propriedades:

195

Para a soma:

196

(i) **associatividade:**  $\forall a, b, c \in A$ ,  $(a + b) + c = a + (b + c)$ ;

197

(ii) **comutatividade:**  $\forall a, b \in A$ ,  $a + b = b + a$ ;

198

(iii) existe um **elemento neutro**  $0_A$  tal que:  $\forall a \in A$ ,  $a + 0_A = a$ ;

199

(iv) existe **elemento oposto:**  $\forall a \in A$ ,  $\exists (-a) \in A$  tal que  $a + (-a) = 0_A$ ;

200

Para a multiplicação:

201

(v) **associatividade:**  $\forall a, b, c \in A$ ,  $a * (b * c) = (a * b) * c$ ;

202

(vi) **distributividade** (em relação a adição):  $\forall a, b, c \in A$ ,  $a * (b + c) = a * b + a * c$  e  
203  $(a + b) * c = a * c + b * c$ ,

204

cuja notação é dada por  $(A, +, *)$ , denotará um anel  $A$  com as operações  $+$  e  $*$ .

205

206 **Definição:** Um anel  $(A, +, *)$ , onde a operação  $*$  é comutativa é denominado de **anel co-**  
207 **mutativo**.

208

209 **Definição:** Seja  $K$  um conjunto de elementos onde estejam definidas duas operações binárias  
210 (adição e multiplicação), e denotaremos por “+” e “\*” respectivamente. Então,  $K$  é um corpo  
211 com as duas operações se as condições forem satisfeitas:

212

(i)  $K$  é um **grupo comutativo** sob  $+$ . O elemento identidade é o 0 (zero).

213

(ii) O conjunto dos elementos não nulos em  $K$  é um grupo comutativo em  $*$ . O elemento  
214 identidade é o 1 (um).

215

(iii) A multiplicação é distributiva sob adição, isto é, para quaisquer  $a, b$  e  $c$  em  $K$ ,  $a * (b + c) =$   
216  $a * b + a * c$ ,

217

cuja notação é dada por  $(K, +, *)$ , denotará um corpo  $K$  com as operações  $+$  e  $*$ .

218

219 **Definição:** Corpo de Galois é um corpo com um número finito de elementos representado  
220 por  $GF(p)$ , onde  $p$  é um número primo.

221

A partir da definição, temos as seguintes propriedades:

222

1. Para todo  $a$  em um corpo,  $a \cdot 0 = 0 \cdot a$ .

223

2. Para quaisquer dois elementos não nulos  $a$  e  $b$  em um corpo, temos que  $a \cdot b \neq 0$ .

224

3. Se  $a \cdot b = 0$  e  $a \neq 0$ , então  $b = 0$ .

225

4. Para quaisquer dois elementos  $a$  e  $b$  em um corpo, temos que  $-(a \cdot b) = (-a) \cdot b = a \cdot (-b)$ .

226

5. Para  $a \neq 0$ , temos que, se  $a \cdot b = a \cdot c$ , então  $b = c$ .

227 **Exemplo:** Considere o conjunto  $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ . Este conjunto é um Corpo de Galois, cujas  
 228 operações de adição e multiplicação módulo - 5 são apresentadas nas Tabelas 2 e 3.

Tabela 2: Operação de adição módulo - 5.

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Tabela 3: Operação de multiplicação módulo - 5.

*	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

229

230 Para qualquer inteiro  $m$ , é possível estender um corpo primo  $GF(p)$  com  $p$  elementos para  
 231 um corpo estendido  $GF(p^m)$  com  $p^m$  elementos. Portanto, a ordem de qualquer corpo finito  
 232 estendido é potência de um primo.

### 233 Construções de Corpos de Galois $GF(2^m)$

234 Considere dois elementos de  $GF(2)$ , 0 e 1 e um símbolo  $\alpha$ , com a operação de multiplicação  
 235 “.” da seguinte forma:

236

$$0 \cdot \alpha^j = \alpha^j \cdot 0 = 0,$$

237

$$1 \cdot \alpha^j = \alpha^j \cdot 1 = \alpha^j,$$

238

$$\alpha^i \cdot \alpha^j = \alpha^j \cdot \alpha^i = \alpha^{i+j},$$

239 onde  $i, j = 0, 1, 2, \dots$

240 Logo, construímos o conjunto  $F = \{0, 1, \alpha, \alpha^2, \dots, \alpha^j, \dots\}$ , sobre o qual a operação de “.” está  
 241 definida. O elemento 1 é algumas vezes denotado por  $\alpha^0$ .

242 Assumindo a condição sobre o elemento  $\alpha$ , tal que o conjunto  $F$  contenha apenas  $2^m$  elemen-  
 243 tos e seja fechado sobre a multiplicação. Seja  $p(X)$  um polinômio primitivo <sup>iv</sup> de grau  $m$  sobre  
 244  $GF(2)$ . Assim, assumimos que  $p(\alpha) = 0$ , isto é,  $\alpha$  é raiz de  $p(X)$ . Como  $p(X)$  divide  $X^{2^m-1} + 1$ ,  
 245 temos que  $X^{2^m-1} + 1 = q(X) \cdot p(X)$ . Trocando  $X$  por  $\alpha$ , obtemos:

246

$$X^{2^m-1} + 1 = q(X) \cdot p(X),$$

247

$$X^{2^m-1} + 1 = q(X) \cdot 0,$$

248

$$X^{2^m-1} + 1 = 0,$$

249

$$X^{2^m-1} = 1.$$

<sup>iv</sup>Polinômio primitivo: é um polinômio irreduzível, ou seja, se  $p(X)$  é não divisível por qualquer polinômio sobre  $GF(2)$  de grau menor que  $m$  e maior que zero, se o menor inteiro positivo  $n$  para o qual  $p(X)$  divide  $X^n + 1$  é  $n = 2^m - 1$ .

250 Logo, sob a condição que  $p(\alpha) = 0$ , o conjunto  $F$  se torna finito e contém os elementos:  
 251  $F^* = \{0, 1, \alpha, \dots, \alpha^{2^m-2}\}$ .

252 A construção de um Corpo de Galois, a partir de um polinômio primitivo, resulta em três  
 253 tipos de representação para os seus elementos: a representação por potência em  $F^*$ , a repre-  
 254 sentação polinomial e a representação vetorial.

255 **Exemplo:** Seja  $m = 4$ . O polinômio  $p(X) = 1 + X + X^4$  é um polinômio primitivo sobre  $GF(2)$ .  
 256 Estabelecendo  $p(\alpha) = 0$ , temos que  $1 + \alpha + \alpha^4 = 0$ , ou seja,  $\alpha^4 = 1 + \alpha$ . Usando esta relação,  
 257 podemos construir  $GF(2^4)$ , cujas representações são apresentadas na Tabela 4.  
 258

Tabela 4:  $GF(2^4)$  gerado por  $p(X) = 1 + X + X^4$ .

Representação por potência	Representação polinomial	Representação vetorial
0	0	(0000)
1	1	(1000)
$\alpha$	$\alpha$	(0100)
$\alpha^2$	$\alpha^2$	(0010)
$\alpha^3$	$\alpha^3$	(0001)
$\alpha^4$	$1 + \alpha$	(1100)
$\alpha^5$	$\alpha + \alpha^2$	(0110)
$\alpha^6$	$\alpha^2 + \alpha^3$	(0011)
$\alpha^7$	$1 + \alpha + \alpha^3$	(1101)
$\alpha^8$	$1 + \alpha^2$	(1010)
$\alpha^9$	$\alpha + \alpha^3$	(0101)
$\alpha^{10}$	$1 + \alpha + \alpha^2$	(1110)
$\alpha^{11}$	$\alpha + \alpha^2 + \alpha^3$	(0111)
$\alpha^{12}$	$1 + \alpha + \alpha^2 + \alpha^3$	(1111)
$\alpha^{13}$	$1 + \alpha^2 + \alpha^3$	(1011)
$\alpha^{14}$	$1 + \alpha^3$	(1001)

259

## 260 Elementos de Sistemas de Comunicação

261 A transmissão da informação é a motivação do desenvolvimento dos códigos corretores de  
 262 erros, porém, sua aplicação e realização estão totalmente vinculadas à teoria das comunicações.

263 Podemos considerar que um sistema de comunicação digital é um conjunto de meios físicos,  
 264 equipamentos, ou até mesmo de um organismo, cujo principal objetivo é transferir dados de  
 265 uma fonte de informação para um determinado destino (destinatário), de maneira confiável,  
 266 permitindo que a mensagem seja recebida de forma fidedigna à informação original, por meio  
 267 de um canal de comunicação.

268 É importante saber como são os modelos de redes de sistemas de comunicação. A **fonte**  
 269 gera símbolos a serem transmitidos, pertencentes a um alfabeto finito que na maioria dos casos  
 270 é binário, o **emissor** possui o papel de converter esses dados em sinais adequados, por meio  
 271 de processos de modulação e codificação, para o **sistema de transmissão**, presente entre a  
 272 fonte e o destinatário, que por sua vez transforma os dados em sinais, transportando-os para o  
 273 **receptor**, que transforma os sinais novamente em dados, que chegam ao **destino** que consome  
 274 os dados.

275 Durante essa transmissão, temos o canal atuando como um filtro, fazendo com que o sinal  
 276 seja transmitido de forma correta, distorcendo sua forma de onda. Todavia, pode-se ocorrer

no processo de transmissão de sinal efeitos indesejáveis que afetam o sinal, por meio de interferências, ruídos e distorções que impedem a passagem dos sinais para o receptor, fazendo com que eles cheguem de modo indesejado.

A **distorção** é a resposta imperfeita do sistema ao sinal utilizado e desaparece quando o mesmo é desligado. A **interferência** é uma contaminação do sinal que ocorre devido a fontes humanas, que podem ser linhas cruzadas que atrapalham a transmissão dos sinais, os **ruídos** são sinais aleatórios que podem afetar de tal forma a transmissão que pode ocorrer perda total dos sinais. Os principais elementos de um sistema de comunicação são apresentados na Figura 2.

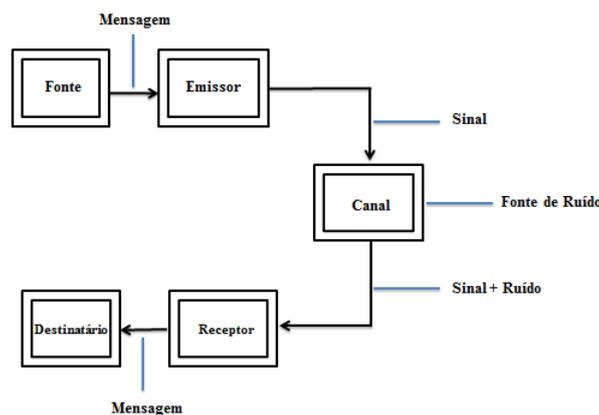


Figura 2: Elementos Básicos de um Sistema de Comunicação.

## Resultados e Discussões

A premissa central da genética é a de que os genes são perpetuados sob a forma de sequências de ácidos nucleicos e gerados como proteínas (MAY et al.(2004)). Diversos pesquisadores (GATLIN (1972); ROLMAN-ROLDAN(1996); MAY et al (2004); MAY (2004); ROCHA e PALAZZO JÚNIOR(2010); FARIA e PALAZZO JÚNIOR (2011)), desenvolveram modelos que tentam capturar várias informações sobre aspectos teóricos do sistema genético.

Nesta seção, serão apresentados alguns dos mais importantes modelos para os sistemas de comunicação biológico, evidenciando as semelhanças e diferenças entre os modelos analisados, mostrando as conexões entre biologia, sistemas de comunicação e estruturas algébricas.

Foram analisados os modelos propostos por Gatlin, May, Rocha e Palazzo Júnior e Faria e Palazzo Júnior encontrados em Gatlin (1972), May et al. (2004), Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011).

### Modelo de Comunicação de Gatlin

Gatlin apresenta um dos trabalhos pioneiros sobre as propriedades teóricas da informação em sistemas genéticos. Segundo Gatlin (1972), a vida pode ser definida operacionalmente como um sistema de processamento, que é adquirido através da evolução, aumentando a capacidade de armazenar e processar as informações necessárias para se obter sua reprodução de modo satisfatório.

A Figura 3 apresenta a visão da teoria da comunicação proposta por Gatlin do sistema genético. Em seu modelo, as sequências de bases do DNA representam a mensagem codificada gerada por uma fonte, posteriormente, o DNA codificado passa por um canal (definido pela transcrição e tradução), que Gatlin refere-se a todos os mecanismos para produzir proteínas. A sequência de aminoácidos é a mensagem recebida, que refere-se ao receptor/destinatário.

309 Em seu trabalho, não é explícito onde se encaixa a replicação ou se ela considera o processo  
 310 de replicação como parte do codificador. Apesar da existência de erros de transcrição e de  
 311 tradução, a replicação também introduz erros que se propagam além de um único evento de  
 312 replicação, sendo uma fonte significativa de ruído que deve ser abordada separadamente.

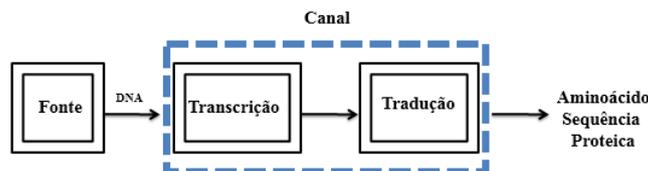


Figura 3: Modelo proposto por Gatlin.

### 313 Modelo de Comunicação de May

314 O objetivo do modelo proposto por May é interpretar o mecanismo de iniciação da tradução  
 315 em organismos procarióticos. Sendo assim, May et al. (2004), define em seu modelo o RNAm  
 316 como a saída do canal de comunicação e incorpora um decodificador que traduz o RNAm em  
 317 cadeias de aminoácidos formadora de proteína. Originalmente, o canal consiste na replicação  
 318 do DNA e no processo de transcrição durante os quais erros são introduzidos na sequência  
 319 nucleotídica. Este modelo é apresentado na Figura 4.

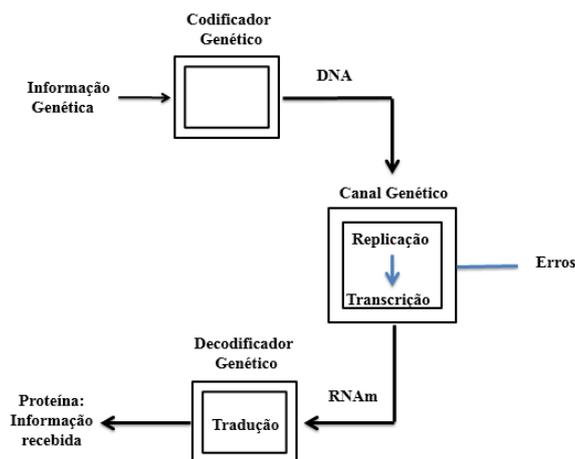


Figura 4: Modelo proposto por May.

320 O processo de decodificação genética é separado em três fases: transcrição, início da tradução  
 321 e alongamento da tradução mais terminação. Do ponto de vista da comunicação, as mutações  
 322 (transferidas do pai para a prole), podem ser vistas como os erros ocorridos na transmissão,  
 323 quando corrompem a mensagem original. Durante a replicação, os erros que podem ocorrer são:  
 324 eliminação de bases de DNA e a inserção de bases de DNA incorretas. Tais erros ou mutações no  
 325 código genético podem corromper e distorcer o sinal genético, que é vital para a sobrevivência  
 326 do organismo (MAY, 2004).

### 327 Modelo de Comunicação de Rocha e Palazzo Júnior

328 Rocha e Palazzo Júnior (2010), propõem um modelo de comunicação biológico para a im-  
 329 portação de proteínas mitocondriais, conforme a Figura 5, com o objetivo de identificar estru-

330 turas matemáticas associadas às sequências de DNA. O processo de codificação é realizado pelo  
 331 bloco transmissor, e o processo de decodificação é realizado pelo bloco receptor.

332 O codificador é o responsável pelas transformações das bases nitrogenadas  $\{A, C, G, T\}$  em  
 333 um alfabeto matemático  $\{0, 1, 2, 3\}$ , usado pelo código através da estrutura algébrica de anel,  
 334 estabelecendo o mapeamento. O código BCH refere-se as palavras-códigos<sup>v</sup> na saída do codifica-  
 335 dor que está relacionada à sequência de direcionamento em termos de nucleotídeos, construídas  
 336 através de extensões de Galois, enquanto que o modulador consiste do código genético, do RNAt  
 337 e do ribossomo e, a palavra-código na saída se refere à sequência de direcionamento em termos  
 338 de aminoácidos.

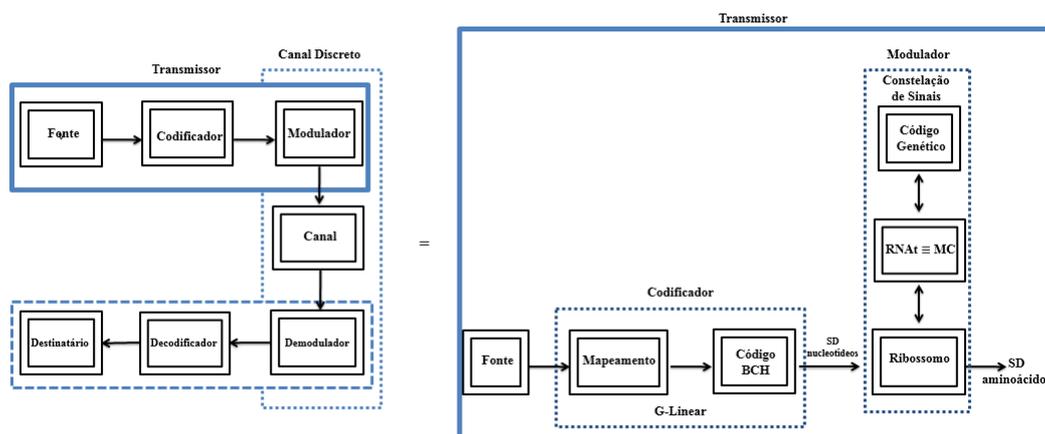


Figura 5: Modelo proposto por Rocha e Palazzo Júnior.

### 339 Modelo de Comunicação de Faria e Palazzo Júnior

340 Faria e Palazzo Júnior (2011) propõem um modelo de sistema de comunicação de informação  
 341 genética e um sistema de comunicação de informação genômica, análogos ao modelo de um  
 342 sistema de comunicação digital, Figuras 6 e 7, respectivamente.

343 O modelo de informação genética é constituído de um codificador genético e de um modulador  
 344 genético. A palavra-código na saída do codificador relaciona-se com a sequência de nucleotídeos  
 345 (RNAm) e na saída do modulador com a sequência dos aminoácidos (proteína). A função  
 346 do RNAt é realizar o mapeamento do código genético, porém é necessária uma caracterização  
 347 matemática no contexto de um sistema de comunicação digital, onde os 64 códonos representam  
 348 os sinais da constelação de sinais.

349 O modelo de informação genômica associa-se com o bloco do codificador de canal. Portanto,  
 350 o processo de codificação está relacionado com os possíveis erros nesse canal, sendo esse modelo  
 351 responsável pela identificação e reprodução na informação genômica contida na dupla hélice do  
 352 DNA.

<sup>v</sup>Palavras-código: é uma sequência finita de dígitos. O número de dígitos de uma palavra-código é o seu comprimento.

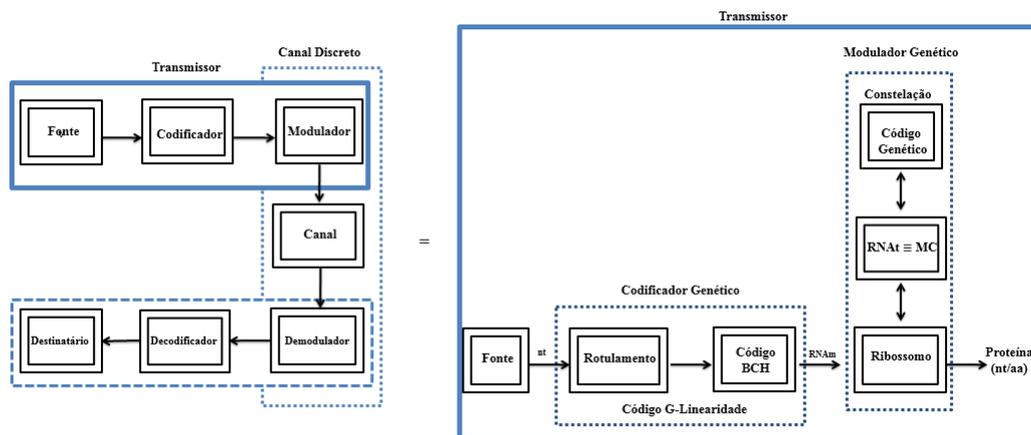


Figura 6: Modelo de um sistema de comunicação de informação genética proposto por Faria e Palazzo Júnior.

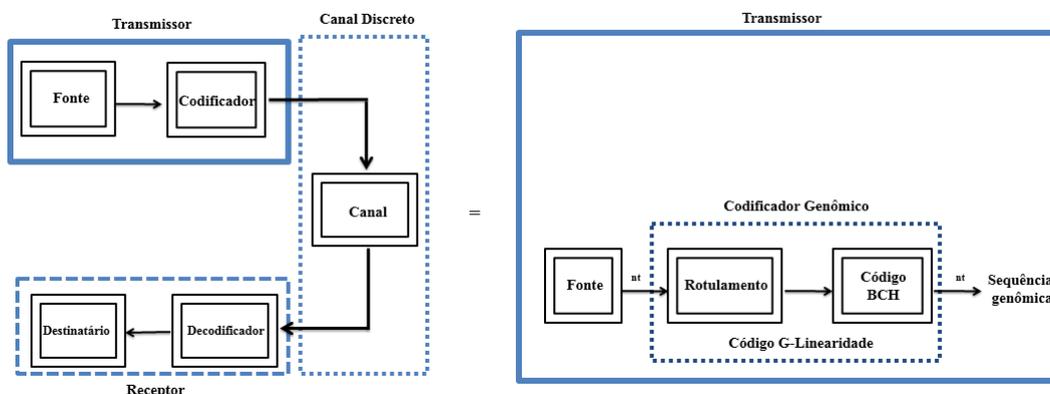


Figura 7: Modelo de um sistema de comunicação de informação genômica proposto por Faria e Palazzo Júnior.

### 353 Comparações dos Modelos Analisados

354 No modelo proposto por Gatlin, percebe-se que não se deixa claro onde ocorre o processo  
 355 de replicação, apenas destaca que este processo é uma fonte significativa de ruído, que deve ser  
 356 abordada separadamente. Considera a sequência de DNA como a entrada do canal, e a proteína  
 357 como a saída. Gatlin, não aborda o papel da replicação durante a transmissão e, o ruído são os  
 358 mecanismos genéticos responsáveis pela síntese proteica (transcrição e tradução).

359 A proposta de Gatlin, se difere da proposta de May pois, ela caracteriza o RNAm como a  
 360 saída do canal de comunicação e, seu modelo aborda o processo de replicação do DNA, como  
 361 sendo o canal que introduz os erros, o que não é explícito no outro modelo. Contudo, May  
 362 não aborda os erros de transcrição e tradução e, embora eles não se propaguem com os erros da  
 363 replicação, é necessária uma análise mais detalhada de seus impactos no sistema de comunicação  
 364 genética.

365 Rocha e Faria diferem das propostas analisadas, uma vez que elas apresentam um modelo  
 366 de sistema de comunicação para a importação de proteínas, que consiste em caracterizar os  
 367 processos de codificação e de decodificação no direcionamento dessas proteínas. Conforme men-  
 368 cionado, em seus modelos, a palavra-código na saída codificador se refere aos nucleotídeos e na  
 369 saída do modulador se refere aos aminoácidos. O mapeamento realizado no modelo de Rocha,

370 consiste da associação das bases nitrogenadas do DNA e do RNA com a estrutura algébrica de  
 371 anel. Entretanto, Faria expande o mapeamento para um rotulamento presente no codificador  
 372 genético, que consiste em 24 possibilidades (conjunto das bases nitrogenadas do DNA/RNA) e,  
 373 além de fazer a associação dessas bases nitrogenadas com a estrutura algébrica de anel, também  
 374 é feita a associação com a extensão de Galois de grau 2 de  $GF(2)$ . Os erros ocorridos no processo  
 375 de codificação destes modelos estão localizados em um canal discreto.

376 Através do modelo proposto por Rocha e Faria, podemos realizar uma associação algébrica  
 377 para fazer a relação do código genético e a constelação de sinais, através de extensões de Galois,  
 378 conforme apresenta a Figura 8.

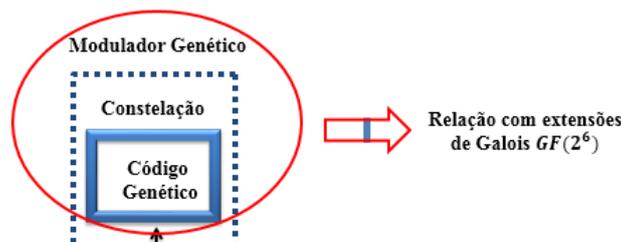


Figura 8: Associação algébrica do modelo de comunicação genética.

## 379 Conclusões

380 Ao apresentarmos as semelhanças existentes entre a teoria de comunicações com o dogma  
 381 da biologia molecular, percebe-se que os pesquisadores desenvolveram em seus modelos diversas  
 382 informações sobre aspectos teóricos do sistema genético e, alguns buscaram suprir limitações não  
 383 abrangidas em outros modelos, limitações estas que corroboram com os resultados apresentados  
 384 nos trabalhos de Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011), cujo sistema  
 385 de codificação biológico consiste de um codificador e de um modulador, construídos através de  
 386 estruturas algébricas (anel e extensão de Galois), relacionando com o sistema de comunicação  
 387 padrão.

388 Neste artigo foram apresentadas interessantes relações e conexões, mostrando um elo entre  
 389 sistema de comunicação, estruturas algébricas e elementos biológicos, buscando uma contribuição  
 390 em uma área de pesquisa em franca expansão, fazendo com que a biologia se torne uma ciência  
 391 fundamentada teoricamente em conceitos matemáticos e de engenharia.

## 392 Referências

- 393 ALBERTS, B. et al. *Biologia Molecular da Célula*. 5. ed. Porto Alegre: Artmed, 2010.  
 394  
 395 BATTAIL, G. *An Outline of Informational Genetics*. Morgan e Claypool Publishers, 2008.  
 396  
 397 DOMINGUES, H. H.; IEZZI, G. *Álgebra Moderna*. 4. ed. São Paulo: Atual, 2003.  
 398  
 399 FARIA, L. C. B.; PALAZZO JÚNIOR, R. *Existências de Códigos Corretores de Erros e*  
 400 *Protocolos de Comunicação em Sequências de DNA*. 2011. 147 f. Tese (Doutorado em  
 401 Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação Universidade  
 402 Estadual de Campinas (Unicamp), Campinas, 2011.  
 403  
 404 GARCIA, A.; LEQUAIN, Y. *Elementos de álgebra*. 5. ed. Rio de Janeiro: IMPA, 2010.  
 405

- 406 GATLIN, L. L. I. *Information Theory and the Living System*. Columbia University Press, New  
407 York, NY, 1972.
- 408
- 409 GONÇALVES, A. *Introdução à álgebra*. Projeto Euclides, 5. ed. Rio de Janeiro: IMPA, 2009.
- 410
- 411 LIN, S.; COSTELO JR, D. J. *Error Control Coding: Fundamentals and Applications*.  
412 Englewood Cliffs: Prentice Hall, 1983.
- 413
- 414 MAY, E. E. Towards a Biological Coding Theory Discipline. *Review New Thesis*, v.1, n.1, p.  
415 19-38, 2004.
- 416
- 417 MAY, et al. An error-correcting code framework for genetic sequence analysis. *Journal of The*  
418 *Franklin Institute*, v.341, n.23, p.89-109, 2004.
- 419
- 420 OLIVEIRA, A. J.; PALLAZO JÚNIOR, R. *Análise Algébrica dos Rotulamentos Associados ao*  
421 *Mapeamento do Código Genético*. 2012. 105 f. Dissertação (Mestrado em Engenharia Elétrica)  
422 - Faculdade de Engenharia Elétrica e de Computação Universidade Estadual de Campinas  
423 (Unicamp), Campinas, 2012.
- 424
- 425 PAMPHILE, J. A. (Org.), VICENTINI, V. E. P. (Org.). *Genética*. Maringá: EDUEM, 2011.
- 426
- 427 ROCHA, A. S. L., PALAZZO JÚNIOR, R., SILVA-FILHO, M.C. *Modelo de sistema de*  
428 *comunicações digital para o mecanismo de importação de proteínas mitocondriais através de*  
429 *códigos corretores de erros*. 2010. 155 f. Tese (Doutorado em Engenharia Elétrica) - Faculdade  
430 de Engenharia Elétrica e de Computação Universidade Estadual de Campinas (Unicamp),  
431 Campinas, 2010.
- 432
- 433 SHANNON, C. E.; A mathematical theory of communication. *The Bell System Technical*  
434 *Journal*, v.27, p. 379-423, 1948.
- 435