

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.

Introdução

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

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

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

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

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

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

Preliminares

Elementos de Biologia

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

e capacidade de movimentação (em alguns casos). As células se dividem em dois grupos, procariotesⁱ e eucariontesⁱⁱ.

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

Nucleotídeos e Ácidos Nucleicos

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

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

A Molécula de DNA

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

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

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

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

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

ⁱⁱ 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 procariotes. Possuem membrana nuclear individualizada e vários tipos de organelas. Aqui estão incluídos os animais, vegetais, fungos e leveduras, que são seres pluricelulares.

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

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

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

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

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

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

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

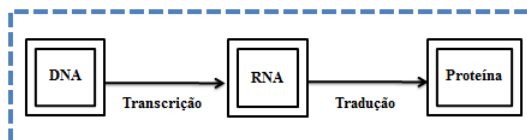


Figura 1: Esquema da síntese proteica.

ⁱⁱⁱCódon: sequência de três nucleotídeos do RNAm a serem traduzidos em um único aminoácido.

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

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

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

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

Elementos de Álgebra

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

$$\gamma: G \times G \longrightarrow G$$

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

munido das seguintes propriedades:

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

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

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

é um grupo.

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

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

$$\begin{aligned} + : A \times A &\longrightarrow A \\ (a, b) &\longmapsto a + b. \end{aligned}$$

$$\begin{aligned} * : A \times A &\longrightarrow A \\ (a, b) &\longmapsto a * b. \end{aligned}$$

e é munido das seguintes propriedades:

Para a soma:

- (i) **associatividade:** $\forall a, b, c \in A, (a + b) + c = a + (b + c)$;
- (ii) **comutatividade:** $\forall a, b \in A, a + b = b + a$;
- (iii) existe um **elemento neutro** 0_A tal que: $\forall a \in A, a + 0_A = a$;
- (iv) existe **elemento oposto:** $\forall a \in A, \exists (-a) \in A$ tal que $a + (-a) = 0_A$;

Para a multiplicação:

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

cujas notações são dadas por $(A, +, *)$, denotará um anel A com as operações $+$ e $*$.

Definição: Um anel $(A, +, *)$, onde a operação $*$ é comutativa é denominado de **anel comutativo**.

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

- (i) K é um **grupo comutativo** sob $+$. O elemento identidade é o 0 (zero).
- (ii) O conjunto dos elementos não nulos em K é um grupo comutativo em $*$. O elemento identidade é o 1 (um).
- (iii) A multiplicação é distributiva sob adição, isto é, para quaisquer a, b e c em K , $a * (b + c) = a * b + a * c$,

cujas notações são dadas por $(K, +, *)$, denotará um corpo K com as operações $+$ e $*$.

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

A partir da definição, temos as seguintes propriedades:

1. Para todo a em um corpo, $a \cdot 0 = 0 \cdot a$.
2. Para quaisquer dois elementos não nulos a e b em um corpo, temos que $a \cdot b \neq 0$.
3. Se $a \cdot b = 0$ e $a \neq 0$, então $b = 0$.
4. Para quaisquer dois elementos a e b em um corpo, temos que $-(a \cdot b) = (-a) \cdot b = a \cdot (-b)$.
5. Para $a \neq 0$, temos que, se $a \cdot b = a \cdot c$, então $b = c$.

Exemplo: Considere o conjunto $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$. Este conjunto é um Corpo de Galois, cujas 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

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

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

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

$$\begin{aligned} 0 \cdot \alpha^j &= \alpha^j \cdot 0 = 0, \\ 1 \cdot \alpha^j &= \alpha^j \cdot 1 = \alpha^j, \\ \alpha^i \cdot \alpha^j &= \alpha^j \cdot \alpha^i = \alpha^{i+j}, \end{aligned}$$

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

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

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

$$\begin{aligned} X^{2^m-1} + 1 &= q(X) \cdot p(X), \\ X^{2^m-1} + 1 &= q(X) \cdot 0, \\ X^{2^m-1} + 1 &= 0, \\ X^{2^m-1} &= 1. \end{aligned}$$

^{iv}Polinômio primitivo: é um polinômio irredutível, 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 $n = 2^m - 1$.

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

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

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

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

Elementos de Sistemas de Comunicação

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

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

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

Durante essa transmissão, temos o canal atuando como um filtro, fazendo com que o sinal 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.

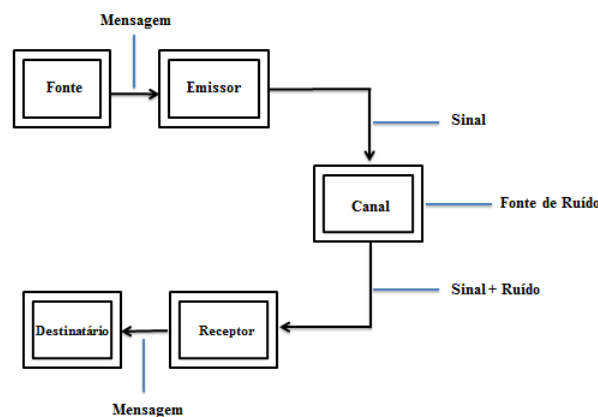


Figura 2: Elementos Básicos de um Sistema de Comunicação.

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.

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

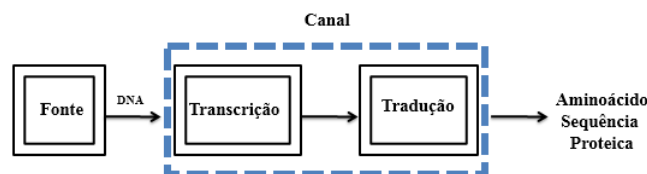


Figura 3: Modelo proposto por Gatlin.

Modelo de Comunicação de May

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

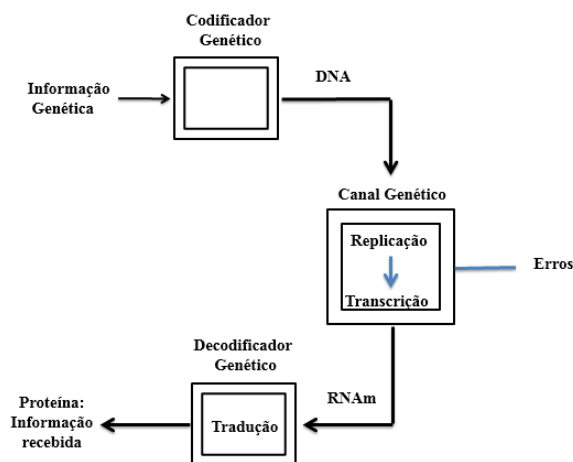


Figura 4: Modelo proposto por May.

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

Modelo de Comunicação de Rocha e Palazzo Júnior

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

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

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

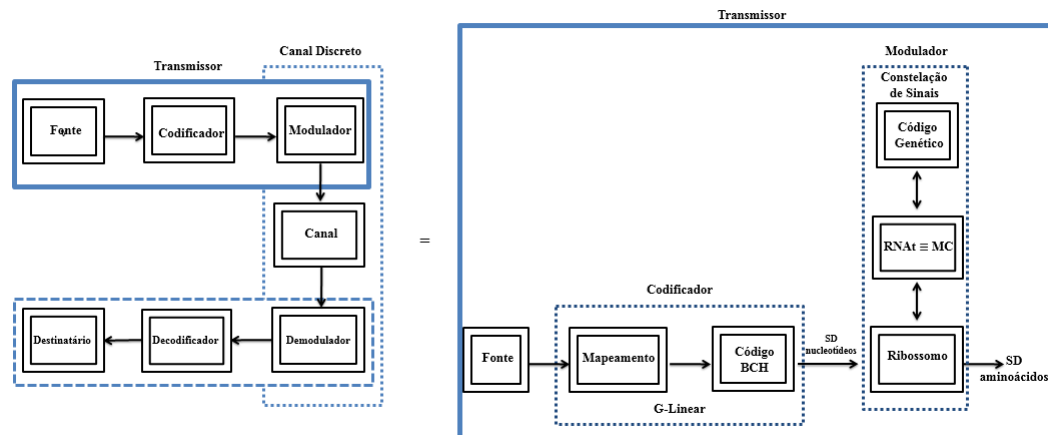


Figura 5: Modelo proposto por Rocha e Palazzo Júnior.

Modelo de Comunicação de Faria e Palazzo Júnior

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

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

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

^vPalavras-código: é uma sequência finita de dígitos. O número de dígitos de uma palavra-código é o seu comprimento.

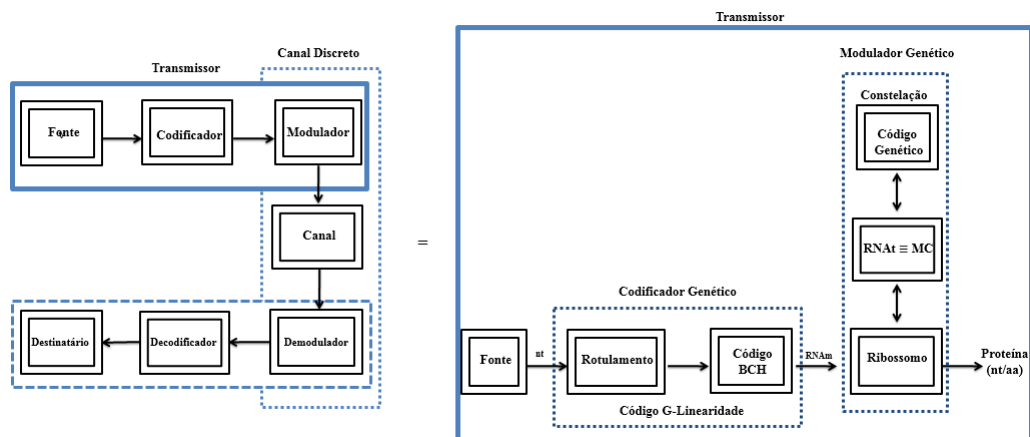


Figura 6: Modelo de um sistema de comunicação de informação genética proposto por Faria e Palazzo Júnior.

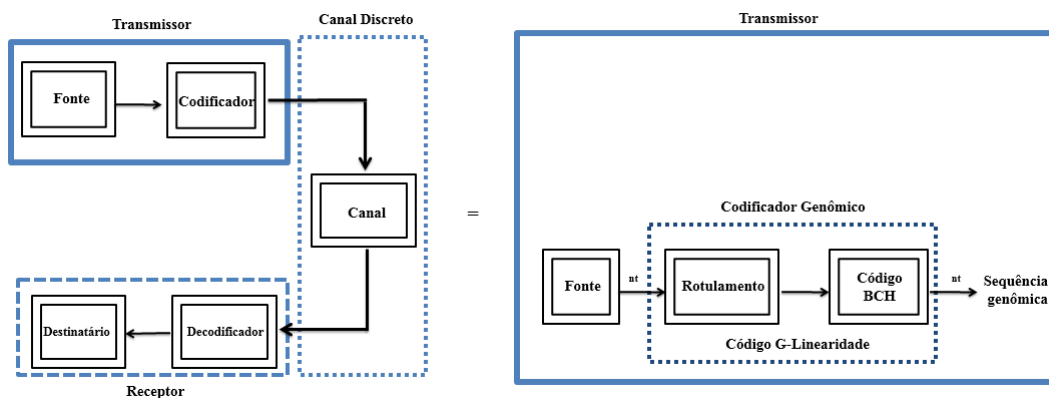


Figura 7: Modelo de um sistema de comunicação de informação genômica proposto por Faria e Palazzo Júnior.

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,

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

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

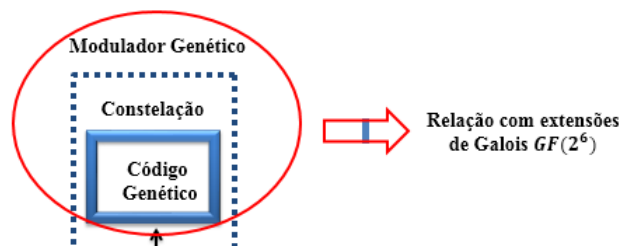


Figura 8: Associação algébrica do modelo de comunicação genética.

Conclusões

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

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

Referências

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

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