

Endomorphism Representation Matrix From Standard Genetic Code

Isah Aisah¹⁾, B. Subartini¹⁾, A. Muhaemin¹⁾

¹⁾Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Padjadjaran University, Bandung Indonesia
E-mail: isah.aisah@unpad.ac.id; betty.subartini@unpad.ac.id; amuhaemin55@gmail.com

(Article History: Received 29-01-2020; Accepted 19-04-2020; Published 20-04-2020)

ABSTRACT

Mutations are changes in genetic material that can occur at the level of genes or chromosomes. Mutations at the gene level are structural changes in the genetic code. In this paper we will investigate the necessary and sufficient conditions for an endomorphism called automorphism. The method used in this research paper is a review of literature conducted by collecting literature from previous studies in accordance with the study discussed. Mathematically, genetic mutations can be viewed with an endomorphism and automorphism f in the vector space $(C_g)^N$ which maps the standard genetic code sequence of wild-type genes into mutant genes. In the endomorphism there is a matrix called the endomorphism representation matrix.

Keywords: automorphism, endomorphism, mutation

Matriks Representasi Endomorfisma Dari Kode Genetik Standar

ABSTRAK

Mutasi merupakan perubahan pada materi genetik yang dapat terjadi pada tingkat gen maupun kromosom. Mutasi pada tingkat gen merupakan perubahan struktur dalam kode genetik. Pada penelitian ini akan dibahas syarat perlu dan syarat cukup sebuah endomorfisma disebut automorfisma. Metode yang digunakan dalam makalah penelitian ini adalah tinjauan literatur yang dilakukan dengan mengumpulkan literatur dari penelitian-penelitian sebelumnya sesuai dengan penelitian yang dibahas. Secara matematis, mutasi genetik dapat ditinjau dengan sebuah endomorfisma dan automorfisma f pada ruang vektor $(C_g)^N$ yang memetakan barisan kode genetik standar gen *wild-type* ke dalam gen mutan. Pada endomorfisma tersebut terdapat sebuah matriks yang disebut matriks representasi endomorfisma.

Kata kunci: Automorfisma, endomorfisma, mutasi Genetik

INTRODUCTION

Genetic code can be viewed as a set of rules that define how a nucleotide sequence in a DNA or RNA determines a particular amino acid sequence in protein synthesis. In the Standard genetic code, there is often a structural change at the level of genes known as mutations, which is a change in the DNA sequence of a living thing. Mutations can occur due to chemical effects, physical effects or no mistakes when replicating DNA (Akhtar, *et al.*, 2015).

Mutations also one of the factors that determine the evolution of living beings.

Mutations that run continuously can lead to the emergence of new varieties that are different from their ancestors which resulted in the evolutionary process. Evolution in the study of biology means a change in the inherited characteristics of a population of organisms from one to the next generations.

When organisms reproduce, their offspring will carry the new properties. This new trait can be obtained from genetic changes due to genetic mutations or gene transfer between populations and between species. Mutations that lead to evolution in living organisms that reproduce only going to happen if this mutation occurs in cells gametes

There have been many studies that represent genetic mutations algebra, including Sanchez *et al.*, 2004 by the journal entitled "The Genetic Code Boolean Lattice"; Jose *et al.*, 2007 in journal entitled "An Extended RNA code and its Relationship to The Standard Genetic Code: An Algebraic and Geometrical Approach".

The set of nucleotide bases can be matched with a member of the set of integers \mathbb{Z}_4 , and can be viewed as a set of structures that form a group with a binary operation. Then set the standard genetic code C_g expanded into a set of rows of standard genetic code $(C_g)^N$ as a result of a matching set of bases with integers can be reviewed \mathbb{Z}_{64} algebraic structure. The set of the standard genetic code line in genetics is a member of the DNA.

Genetic mutations can be reviewed in algebra, by satisfying certain properties. Sanchez, *et al.*, 2005 in the journal entitled "Gene Algebra from a Genetic Structure Algebraic Code" suggests that the genes, a similar situation arises in which almost all mutations in codon associated with automorfisma. On the basis of these problems, in this study the authors will represent a genetic mutation using automorfisma on space-algebra \mathbb{Z}_{64} .

RESEARCH METHOD

The method used in this research paper is a review of literature conducted by collecting literature from previous studies in accordance with the study discussed.

The initial step of research is examining the structure of the genetic code line $(C_g)^N$, which is a \mathbb{Z}_{64} -algebra. Further codon represented in the α vector for the initial codon (wild-type) and β to codon mutations (mutagens), $\alpha, \beta \in (C_g)^N$ with the order of 2^{m_α} and 2^{m_β} with $\alpha = 2^{6-m_\alpha}x_\alpha$ and $\beta = 2^{6-m_\beta}x_\beta$ with x_α, x_β is the maximum sequence element $(\mathbb{Z}_{64})^N$.

The next step is to build endomorfisme $f: (C_g)^N \rightarrow (C_g)^N$ which maps the wild-type codon into codon mutagens, which can be represented by a matrix called the matrix representation endomorfisme f . Endomorfisme matrix representation obtained by determining each element of the diagonal.

RESULTS AND DISCUSSION

R-Algebraic of Standard Genetic Code Set

Nucleotide bases can be seen as an ordered set obtained by adjusting the difference in the number of hydrogen bonds between bases and the chemical properties (purines and pyrimidines) of bases. Nucleotide bases there which has two hydrogen bonds and three bonds that has two bonds, namely adenine (A) and thymine (T) / uracil (U), whereas the nucleotide bases have three bonds, namely guanine (G) and cytosine (C) (Jose, *et al.*, 2006).

This sequence arrangement starts by selecting the nucleotide base that has the least hydrogen bonds so that you can choose A or U. If U is chosen, then G will then be chosen because it has different chemical properties than U. Furthermore, C is chosen because of the different chemical properties from G, and finally A is chosen because of differences in chemical properties with C, so that the obtained set $\{U, G, C, A\}$ is obtained. If the first base chosen is A, then with the same sorting rule we get an ordered set $\{A, C, G, U\}$ (Shánchez, *et al.*, 2005).

The set of nucleotide bases is converted to a set of standard genetic code (triplet nucleotide) denoted by (C_g) . Conversion is done by sorting three nucleotide bases as genetic code so that there is a $4 \times 4 \times 4 = 64$ possible sequence formed as a standard genetic codes. Each member of (C_g) will be matched with a member set of \mathbb{Z}_{64} .

The standard genetic code set be expanded to the standard genetic code set using the direct sum of N times denoted by $(C_g)^N = \{(u_1, u_2, \dots, u_N) | \forall u_i \in C_g, N \in \mathbb{Z}^+\}$ so the number of members of the set $(C_g)^N$ is 64^N . $((C_g)^N, +, \cdot)$ can be viewed as a commutative ring and left module over \mathbb{Z}_{64} ring, so $(C_g)^N$ is \mathbb{Z}_{64} algebra (Shánchez, *et al.*, 2005).

Algebra Structure of Standard Genetic Code

$(C_g)^N$ is vector spaces over \mathbb{Z}_{64} . The basis for this vector space is the set of the following vectors

$$\begin{aligned} e_1 &= (UUG, UUU, \dots, UUU), e_2 \\ &= (UUU, UUG, \dots, UUU), \dots, e_N \\ &= (UUU, UUU, \dots, UUG) \end{aligned}$$

so that each DNA sequence $\alpha \in (C_g)^N$ over the \mathbb{Z}_{64} ring has their respective representations as $\alpha = a_1e_1 + a_2e_2 + \dots + a_Ne_N$ where $(a_i \in \mathbb{Z}_{64})$, in other words ordered pairs (a_1, a_2, \dots, a_N) is the coordinate representation of the gene $\alpha \in (C_g)^N$ in \mathbb{Z}_{64} .

Each element $(C_g)^N$ can be represented by ring \mathbb{Z}_{64} , because $(C_g)^N$ is \mathbb{Z}_{64} -algebra (Szekeres, 2004) and $(C_g, +, \cdot)$ isomorphic with $(\mathbb{Z}_{64}, +, \cdot)$.

Each element of the group $(C_g)^N$ can be grouped into seven classes. According to the Lagrange theorem (Dummit & Foote, 2004; Fraleigh, 2003; Gallian, 2010; Heirstein, 1996), the order of the \mathbb{Z}_{64} is a divisor of 64 and the order has the form $2^m (m = 0, 1, 2, 3, 4, 5, 6)$. Elements that have the order 2^m are the form $2^{6-m}x$, where x is an odd integer between 1 and $(2^m - 1)$.

The grouping of elements from groups $(C_g, +)$ can be seen from the following table 1. Elements with order 64 are loaded into two columns, where each same row is an inverse of the multiplication operation for the \mathbb{Z}_{64} ring.

Endomorphism of Genetic Mutations

Gene mutations can be seen as linear transformations $f(\alpha) = \beta$ which is endomorphism in vector space $(C_g)^N$ with $\alpha \in (C_g)^N$ is a wild-type gene and $\beta \in (C_g)^N$ is a mutant gene of α .

Genetic mutations can be seen as endomorphism f at $(C_g)^N$. Mapping construction $f: (C_g)^N \rightarrow (C_g)^N$, which maps the standard genetic code sequence of wild-type genes into the standard genetic code sequence of mutant genes.

Table 1. Partitions from Groups $(\mathbb{Z}_{64}, +)$ into Seven Classes according to the order

Order 2^m	1	2	4	8	16	32	64	
M	0	1	2	3	4	5	6	
Elemen	0	32	16	8	4	2	1	1
			48	24	12	6	3	43
				40	20	10	5	13
				56	28	14	7	55
					36	18	9	57
					44	22	11	35
					52	26	15	47
					60	30	17	49
						34	19	27
						38	21	61
						42	23	39
						46	25	41
						50	29	53
						54	31	31
						58	33	33
						62	37	45
						51	59	
						63	63	

An endomorphism $f: (C_g)^N \rightarrow (C_g)^N$ can be called a local endomorphism if there are $k \in \{1, 2, \dots, N\}$ and $a_{ik} \in \mathbb{Z}_{64}$ with $i = 1, 2, \dots, N$ such that: $f(e_i) = (0, \dots, a_{ik}, \dots, 0) = a_{ik}e_i + e_i$, for $i \neq k$ and $f(e_k) = a_{kk}e_k$, so for each $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N) \in (C_g)^N$,

$$\begin{aligned} f(\alpha) &= f(\alpha_1, \alpha_2, \dots, \alpha_N) \\ &= \left(\alpha_1, \alpha_2, \dots, \sum_{i=1}^N \alpha_i a_{ik}, \dots, \alpha_N \right) \end{aligned}$$

Local endomorphism f is called diagonal if $f(e_k) = (0, \dots, a_{kk}, \dots, 0) = a_{kk}e_k$ and $f(e_i) = e_i$ for $i \neq k$, in other words $f(\alpha_1, \alpha_2, \dots, \alpha_N) = f(a_1e_1 + a_2e_2 + \dots + a_Ne_N) = (\alpha_1, \alpha_2, \dots, a_{kk}\alpha_k, \dots, \alpha_N)$.

If the representation matrix endomorphism f which depends on the canonical base is a diagonal matrix, then it is called the diagonal endomorphism matrix f .

There is a relationship between algebraic properties and biological properties in $(C_g)^N$ above algebra \mathbb{Z}_{64} -algebra in terms of the order of each element (Shánchez, et al., 2005). Quantitatively explained in the following theorem.

Theorem 1:

For each $\alpha, \beta \in (C_g)^N$ there is endomorphism $f: (C_g)^N \rightarrow (C_g)^N$, so $f(\alpha) = \beta$ if and only if order β divides the order α .

Proof:

1. It is known that $(C_g)^N$ is an isomorphic abelian group with abelian groups $(\mathbb{Z}_{64})^N$ and $f: (C_g)^N \rightarrow (C_g)^N$, endomorphism so $f(\alpha) = \beta$. We will proof order β divides the order α . Because f is a mapping that is surjectif, then every element in the range has a pre-map, so the range is a subgroup $(C_g)^N$, based on the Lagrange theorem, so β divides α .

2. Next, consider $f: (C_g)^N \rightarrow (C_g)^N$, is endomorphism and for each $\alpha, \beta \in (C_g)^N$, order β divides order α , will prove to be $f(\alpha) = \beta$. Let $\alpha, \beta \in (C_g)^N$ in the order $(\alpha) = 2^{m_\alpha}$ and order $(\beta) = 2^{m_\beta}$ with $\alpha = 2^{6-m_\alpha}x_\alpha$ and $\beta = 2^{6-m_\beta}x_\beta$ with x_α, x_β is the element with the maximum order at $(\mathbb{Z}_{64})^N$.

Based on the definition of basis, x_α has a basis $(x_1, x_2, \dots, x_n) \in (C_g)^N$ where $x_i = x_\alpha$ and x_β have a base $(y_1, y_2, \dots, y_n) \in (C_g)^N$ where $y_i = x_\beta$ ($i = 1, 2, \dots, n$). Because there is unqu automorphism $f: (C_g)^N \rightarrow (C_g)^N$ such that $f(x_i) = y_i$ for $i = 1, 2, \dots, N$, then endomorphism f mapping α to β , thus if order β divide α , then $f(\alpha) = \beta$.

From theorem 1, this corollary is obtained.

Corollary 1:

There is a diagonal endomorphism $f: (C_g)^N \rightarrow (C_g)^N$ which converts vectors α to vectors β , where $\alpha, \beta \in (C_g)^N$ if and only if for each $\alpha_i, \beta_i \in \mathbb{Z}_{64}$; ($i = 1, 2, \dots, N$) satisfies $m_{\beta_i} \leq m_{\alpha_i}$

Representatin Matrix of Genetic Mutation

Diagonal endomorphism f is said to be automorphism at $(C_g)^N$ if and only if the a_{kk} is an odd number (Szekeres, 2004).

For each endomorphism or automorphism $f: (C_g)^N \rightarrow (C_g)^N$, there is a matrix A called the representation matrix of endomorphism f which depends on the canonical basis $\{e_i, i = 1, 2, \dots, N\}$ ordo $N \times N$:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \dots & a_{NN} \end{pmatrix}$$

where the row entries are vector images $(e_i), i = 1, 2, \dots, N$.

Based on corrolary1, if $\alpha_i = 2^{6-m_{\alpha i}}x_i$ and $\beta_i = 2^{6-m_{\beta i}}y_i$ with $x_i, y_i \in \mathbb{Z}_{64}$ is an odd number, the diagonal a_{ii} from matrix A can be determined by the equation :

$$a_{ii} = 2^{m_{\alpha i} - m_{\beta i}} y_i x_i^{-1} \tag{1}$$

Automorphism can be used to study gene mutations in N-dimensional space (Shánchez R, et al, 2005).]. Therefore a matrix of representation of any automorphism between two genes will be determined. An endomorphism $f: (C_g)^N \rightarrow (C_g)^N$ is an automorphism if and only if the diagonal element is odd.

Suppose A is a matrix of representation of local endomorphism $f: (C_g)^N \rightarrow (C_g)^N$, where $(\alpha_k) = \beta_l$ with $\alpha_k \in \alpha, \beta_l \in \beta$ and $\alpha, \beta \in (C_g)^N$. Let $(\alpha_k) = (f(\alpha_1), f(\alpha_2), \dots, f(\alpha_N)) = (\beta_1, \beta_2, \dots, \beta_N)$ with vector elements $f(\alpha_k) = \beta_l$ is a linear combination of column elements to -1 from endomorphism matrix A, in other words $f(\alpha_k) = \sum_{k=1}^N \alpha_k a_{kl} = \beta_l$, so

$$\begin{aligned} f(\alpha_k) &= \sum_{k=1}^N \alpha_k a_{kl} \\ &= \alpha_1 a_{1l} + \alpha_2 a_{2l} + \dots + \alpha_i a_{il} + \dots + \alpha_N a_{Nl} \\ &= \left(\sum_{\substack{k=1 \\ k \neq i}}^N \alpha_k a_{kl} \right) + \alpha_i a_{il} \\ &= \delta_i + \alpha_i a_{il} = \beta_l \text{ mod } 64 \quad (1 \leq i \leq N) \end{aligned} \tag{2}$$

Next, to change the diagonal element of the endomorphism f representation matrix, the following theorem is used:

Theoreme 2:

Suppose $2^{m_{\alpha i}}$ is order of α_i and $2^{m(\beta_l - \delta_i)}$ is order $\beta_l - \sum_{\substack{k=1 \\ k \neq i}}^N \alpha_k a_{kl} = \beta_l - \delta_i$. . If the elements $\alpha_i \in \alpha$ are pivots and α_{kl} for $k \neq i$ then $m_{\alpha_i} \geq m_{(\beta_l - \delta_i)}$.

Proof:

Suppose the matrix A is a representation matrix of local endomorphism f such that :

$f(\alpha_1, \alpha_2, \dots, \alpha_N) = (\alpha_1, \alpha_2, \dots, \beta_i, \dots, \alpha_N)$
 $\alpha_i = 2^{6-m_{\alpha_i}} x_i$ and $\beta_i - \delta_i = 2^{6-m_{(\beta_i-\delta_i)}} y_{il}$
 where x_i and y_{il} are odd numbers, then equation $2^{6-m_{\alpha_i}} x_i \alpha_{il} = 2^{6-m_{(\beta_i-\delta_i)}} y_{il}$ causes $m_{\alpha_i} \geq m_{(\beta_i-\delta_i)}$. Because in the abelian group \mathbb{Z}_{64} for each $\alpha_i, \beta_i, \delta_i \in \mathbb{Z}_{64}$, the equation $\delta_i + \alpha_i \alpha_{il} = \beta_i$, has a solution if $m_{\alpha_i} \geq m_{(\beta_i-\delta_i)}$.

Specifically, according to equation (2), the $\alpha_i \in \alpha$ element can be taken as a pivot and the α_{kl} element for $k \neq i$ can be chosen arbitrarily for $m_{\alpha_i} \geq m_{(\beta_i-\delta_i)}$. So, we can specify the column element to 1.

CONCLUSIONS

There is a mapping that maps the standard genetic code sequence of wild-type genes into the standard genetic code sequence of mutant genes, namely endomorphism $f: (C_g)^N \rightarrow (C_g)^N$. Representation of genetic mutations can be done by forming an automorphism representation matrix. The necessary and sufficient conditions for an endomorphism called automorphism at $(C_g)^N$ that is the diagonal endomorphism representation matrix f is an odd number.

REFERENCES

- Akhtar, J., T. Ali & N. Gohain. 2015. Graph in Genetic Code Algebra. *Bioinformatics and Intelligent Control*, **4**: 1-4.
- Dummit, D.S., & R.M. Foote. 2004. *Abstract Algebra*, 3rd ed. John Wiley & Sons, New York.
- Fraleigh, J.B. 2003. *A First Course in Abstract Algebra*, 7th ed. Pearson Education, Illinois.
- Gallian, J.A. 2010. *Contemporary Abstract Algebra*. Richard Stratton, Belmont.
- Heirstein, I.N. 1996. *Abstract Algebra*. Prentice-Hall, New Jersey.
- Jose, M.V., et al. 2006. An Extended RNA code and its Relationship to The Standard Genetic Code: An Algebraic and Geometrical Approach. *Bull. Math. Biol.*, **69**: 215-243.
- Shánchez, R., E. Morgado & R. Grau. 2004. The Genetic Code Boolean Lattice. *MATCH Commun. Math. Comput. Chem.*, **52**: 29-45.
- Shánchez, R., E. Morgado & R. Grau. 2005. Gen Algebra from a Genetic Code Algebraic Structure. *J. Math. Biol.*, **51(4)**: 431-457.
- Szekeres, P. 2004. *A Course in Modern Mathematical Physics*. Cambridge University Press, New York.