

Un modelo matemático para el estudio de las recombinaciones genéticas en ciliados

Lidia García García¹, Ma. Guadalupe Rodríguez Sánchez²

^{1,2}AFPI, Ciencias Básicas, Universidad Autónoma Metropolitana-A, Cd. de México, México
Teléfono (55) 53189014 E-mail: stokes_1560@hotmail.com¹, rsmg@azc.uam.mx²

^{1, 2}

Resumen — El descubrimiento del papel del ADN en el proceso de replicación del material genético y las recombinaciones de dicho material, que permiten la reproducción y evolución de la vida, constituyen un campo de conocimiento de enorme trascendencia para el entendimiento de la vida. El código genético almacenado en el ADN está compuesto por una larga cadena de combinaciones de las letras A, T, G, C, que corresponden a los elementos Adenina, Timina, Guanina y Citosina, respectivamente. Se conocen tres recombinaciones genéticas que ocurren en el código genético. Estas son de gran importancia pues permiten la variabilidad del código transmitido a nuevas generaciones de individuos de una especie, lo que incide en la evolución y sobrevivencia de dicha especie. En este artículo se explicarán una serie de conceptos que conforman un modelo matemático para la explicación e investigación sobre el ADN. El material aquí presentado está centrado en el estudio de los ciliados, seres unicelulares cuyas células tienen la particularidad de tener dos núcleos. Los conceptos matemáticos que se utilizan son las multigráficas 4-regulares, las gráficas circulares y las palabras de doble ocurrencia, todos ellos forman parte de la teoría de delta-matroides.

Palabras Clave — ADN, recombinación genética, multigráficas 4-regulares, gráficas circulares, delta-matroides

Abstract — The discovery of the role of DNA in the process of replication of genetic material and the recombinations of said material that allow the reproduction and evolution of life, constitute a field of knowledge of enormous importance for the understanding of life. Genetic information can be represented as a long chain formed by the letters: A, T, G, C, which correspond to the elements Adenine, Thymine, Guanine and Cytosine, respectively. Three genetic recombinations are known to occur in the genetic code. These are of great importance because they allow the variability of the code transmitted to new generations of individuals of a species, which affects the evolution and survival of said species. This article will explain a series of concepts that make up a mathematical model for the explanation and investigation of DNA. The material presented here is an extension of previous works, and is focused on the study of ciliates, unicellular beings whose cells have the particularity of having two nuclei. The mathematical concepts used are 4-regular multigraphs, circular graphs and delta-matroids.

Keywords — DNA, genetic recombination, 4-regular multigraphs, circular graphs, delta-matroids

I. INTRODUCCIÓN

En 1953 Watson y Crick [1] presentaron su modelo de la doble hélice de la molécula del ADN, abriendo un campo de conocimiento de enorme trascendencia. A partir de entonces y con base en el esfuerzo de muchos científicos se pudo comprender como estaba formado el código genético que permite la replicación de la vida. El código genético se encuentra almacenado en el ADN y está compuesto por cuatro clases de subunidades llamadas nucleótidos. Cada nucleótido consta de un azúcar de desoxirribosa, un grupo de fosfato y una base. Existen cuatro bases Adenina (A), Citosina (C), Guanina (G) y Timina (T), así el código puede verse como una cadena de elementos tomados de las bases y expresados por la letra inicial de su nombre. Los dos filamentos en la doble hélice están unidos por puentes de hidrógeno entre bases complementarias, A solamente se aparea con T y G se aparea con C. Cada hebra de la doble hélice se orienta de acuerdo al enlace que se forma entre el grupo fosfato en la posición 3' de un nucleótido y el grupo hidroxilo en la posición 5' en el siguiente nucleótido. Los dos filamentos se orientan con polaridad opuesta, una en la dirección 5'-3' y la otra en la dirección 3'-5'. Un gen es una secuencia orientada de pares de bases que empieza en una *cola* (t) y termina en una *cabeza* (h), las cuales son conocidas como las extremidades del gen. A lo largo del ADN, cada gen está separado de otro por tramos de secuencias de pares de bases no genéticas, cuya función conocida es la separación de genes sucesivos.

La variabilidad de las características que hereda una generación a sus descendientes depende de las recombinaciones de la información contenida en su código genético, es decir, de operaciones que se efectúan en sus genes. Estas operaciones son: la escisión de lazos, la recombinación de horquilla y la doble recombinación de lazo. Con base en material matemático del campo de la Combinatoria, se elaboró un modelo matemático que puede modelar las recombinaciones y explicar el proceso que ocurre en los genes de una cadena de ADN, ver [2],[3]. Los conceptos matemáticos que se emplean son tomados de los resultados en los campos de las multigráficas 4-regulares [4] [5], gráficas circulares [6] y palabras de doble ocurrencia [4], enmarcados en la teoría de delta-matroides [6][5].

Lo dicho anteriormente se puede aplicar, en general, a la investigación sobre el ADN y sus recombinaciones genéticas. Sin embargo, en este trabajo, se sigue el estudio de [3] sobre

la aplicación de estos conceptos a los ciliados. Los ciliados son organismos unicelulares con la característica de poseer dos núcleos, un macronúcleo (MAC) y un micronúcleo (MIC), que son estructural y funcionalmente diferentes. El MAC tiene muchas copias de cromosomas cortos, los cuales tienen uno o pocos genes. El MIC contienen un número menor de cromosomas cada uno con numerosos genes. El código del MIC es usado para la reproducción y el MAC para la regulación general de la célula. Durante la reproducción sexual el recién formado MIC es transformado en un MAC. Este proceso es llamado *recombinación genética* y se obtiene mediante las operaciones mencionadas de recombinación.

En este trabajo se explicarán los conceptos matemáticos que intervienen en el modelo y su aplicación a la parte biológica. Para ejemplificar los procesos en el ADN y su incidencia en el modelo, se utilizarán partes del código genético de la proteína Actina I en el ciliado *Sterkiella nova*.

II. EL MODELO GENÉTICO

A. Codificación de los genes

Los genes del ADN del MIC están desordenados y posiblemente invertidos respecto al MAC correspondiente, estos están separados por segmentos cortos de ADN no genético, cada uno denominado IES (internally eliminates segment). Los IES's dividen al gen en fragmentos, cada uno llamado MDS (macromolecular destined segments). Así el MIC del gen de la Actina I del ciliado *Sterkiella nova* se puede ver en (1) y puede describirse con la siguiente cadena

$$I_0 M_3 I_1 M_4 I_2 M_6 I_3 M_5 I_4 M_7 I_5 M_9 I_6 \bar{M}_2 I_7 M_1 I_8 M_8 I_9, \quad (1)$$

donde los MDS son los M_i 's y los IES son los I_i 's de la cadena. La inversión de M_2 se representa con una barra superior, como \bar{M}_2 . En el MAC la información genética se encuentra ordenada. Para la Actina I de la *Sterkiella nova*, se tiene la cadena (2):

$$M_1 M_2 M_3 M_4 M_5 M_6 M_7 M_8. \quad (2)$$



Fig. 1. Forma MAC del gen Actina I, en el ciliado *Sterkiella Nova*.

Entre cada pareja de MDS's consecutivos en el MAC hay una pequeña intersección, estos segmentos se llaman *punteros*. En la Fig. 2, puede observarse la colocación de los punteros en un MAC con k MDS's. En la Fig. 3, se muestra un esquema de (1), en el que aparecen también los punteros.

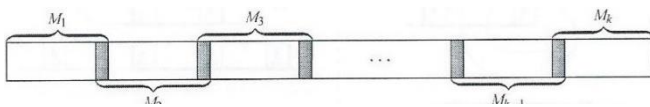


Fig. 2. Estructura MAC con k MDS's.

B. Operaciones de recombinación

Las operaciones de recombinación son formalizadas, de la siguiente manera: Sea $k \in \mathbb{Z}^+$. Considérese un alfabeto

finito Γ cuyas letras son $\{1, 2, \dots, k\} \cup \{\bar{1}, \bar{2}, \dots, \bar{k}\}$ y w una palabra sobre Γ , donde cada letra aparece exactamente dos veces. Se dice que w es una *palabra de doble ocurrencia* sobre Γ [6]. Sean k MDS's en la cadena del MIC y sea $w = w_1 w_2 \dots w_t$, con $w_i \in \Gamma$, las letras w_i representan a los punteros, que se distribuyen de la siguiente manera:

A M_i le corresponde $i(i+1)$, a su inversión \bar{M}_i le corresponde a $(i+1)\bar{i}$, para $i \in \{2, \dots, k-1\}$. A M_1 y M_k les corresponde solo un puntero, de esta manera la forma de Actina I de *Sterkiella nova* tiene la cadena de punteros: 34 45 67 56 78 9 $\bar{3}\bar{2}$ 2 89.

Se procede a explicar las tres operaciones de recombinación consideradas en el *modelo intramolecular de ensamblamiento genético* [7], el cual postula las operaciones basadas en la recombinación homóloga para ejecutar la desfragmentación del gen en su forma MIC.

Escisión de lazo

Esta operación elimina un IES entre un par de punteros consecutivos que tienen la misma orientación. Se explica con una fracción de la cadena del MIC, en la Fig. 3. La molécula se pliega alineando el par de punteros (Fig. 3a) y se lleva a cabo la recombinación análoga (Fig. 3b). Esta recombinación une los fragmentos MDS_2 y MDS_3 y elimina el IES_2 , se produce una molécula circular.

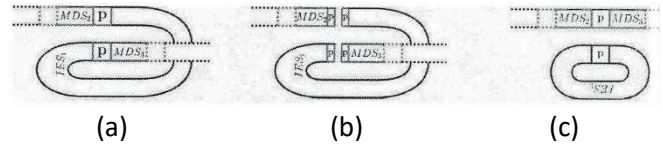


Fig. 3. Escisión de lazo

Recombinación de doble lazo

Esta operación de recombinación se aplican a una molécula de ADN con dos pares de punteros cuyos segmentos delimitados por cada par están superpuestos. La molécula se pliega en dos lazos alineando a los dos pares de punteros. Este plegamiento produce el intercambio de los segmentos superpuestos. Se ejemplifica en la Fig. 4.

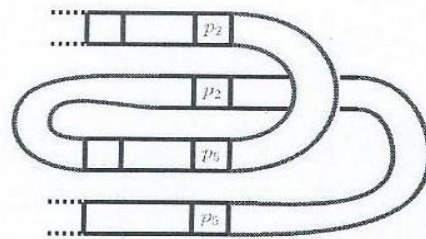


Fig. 4. Recombinación de doble lazo

Recombinación de horquilla

Se aplica a una molécula en la que el par de secuencias repetidas tienen orientación opuesta. En este caso la molécula forma un pliegue en forma de horquilla, el cual alinea el par de punteros invirtiendo el segmento entre ellos, después de lo

cual tienen lugar la recombinación homóloga. Se ejemplifica en la Fig. 5, con un fragmento de una cadena en la aparece cuatro punteros: $p_3 p_4 \bar{p}_5 \bar{p}_4$. Al formarse la horquilla de AND, el puntero p_4 presenta direcciones opuestas en la cadena, como puede apreciarse en el primer renglón de la Fig. 5, al formarse la horquilla de ADN, como se aprecia en el segundo renglón de la Fig. 5, el puntero de salida p_4 se alinea en la dirección con el puntero de entrada p_4 . Posteriormente se lleva a cabo la recombinación homóloga. En la Fig. 5, se usan direcciones en lugar de barras, sin menoscabo de la claridad.

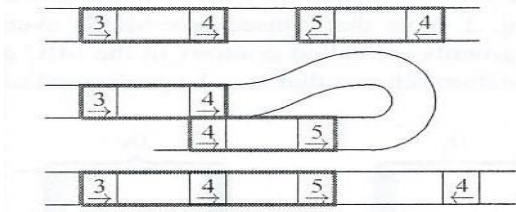


Fig. 5. Recombinación de horquilla

C. Gráficas asociadas al modelo

Se puede expandir la expresión (1) si se agregan los punteros con la regla que se explicó anteriormente, así la cadena expandida para la *Sterkiella nova* quedaría como en (3). Puede verse un esquema de la cadena (3) en la Fig. 6.

$$w' = I_0 p_3 M_3 p_4 I_1 p_4 M_4 p_5 I_2 p_6 M_6 p_7 I_3 p_5 M_5 p_6 I_4 p_7 M_7 p_8 I_5 p_9 M_9 I_6 \bar{p}_3 \bar{M}_2 \bar{p}_2 I_7 M_1 p_2 I_8 p_8 M_8 p_9 I_9 \quad (3)$$

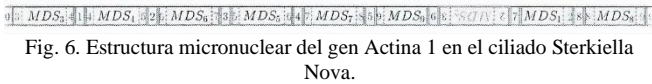


Fig. 6. Estructura micronuclear del gen Actina 1 en el ciliado *Sterkiella Nova*.

La cadena representada por w' se puede colocar sobre una circunferencia siguiendo el orden de los elementos de (3) para construir una gráfica denominada $D_{w'}$, se identifican los punteros con los vértices de una gráfica y los MDS's y IES's como las aristas en la gráfica. Si a $D_{w'}$ se le da una orientación de acuerdo con la dirección de izquierda a derecha en la cadena (3), sin importa cual es el vértice de inicio, se obtiene D_w como digráfica, ver la Fig. 7, sin las aristas azules. Es natural pensar en unir los punteros que tienen el mismo índice, mediante una arista, independientemente de que tengan barra o no, al hacer esto se obtendrá una gráfica que se denotará como $H_{w'}$, que puede verse en la Fig. 7.

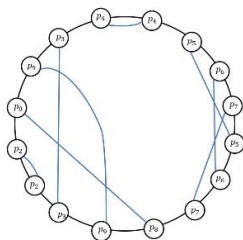


Fig. 7. Gráfica $H_{w'}$.

Si a continuación en $H_{w'}$ se contraen las aristas que unen los vértices con los punteros que tienen el mismo número, se obtendrá una gráfica tal que todos sus vértices tienen grado 4, se dice que esta es una multigráfica 4-regular, es multigráfica porque se permite que contenga aristas múltiples y lazos, dicha multigráfica se denotará como $G_{w'}$. La gráfica $G_{w'}$ se muestra en la Fig. 8, en los vértices se ha colocado solo el índice correspondiente al puntero que le corresponde.

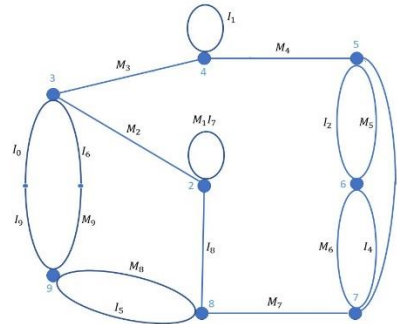


Fig. 8. Gráfica $G_{w'}$.

En la gráfica $G_{w'}$ todos los vértices tienen grado par, luego $G_{w'}$ tiene un circuito Euleriano. Por otro lado, cada vértice v de $G_{w'}$ puede dar origen a dos vértices v_1 y v_2 cada uno de grado 2, se considera una partición de las aristas incidentes a v , teniendo dos aristas cada clase obtenida, se asigna a v_i diferentes clases de la partición, $i=1,2$; hay tres maneras de realizar la partición, como puede verse en la Fig. 9.

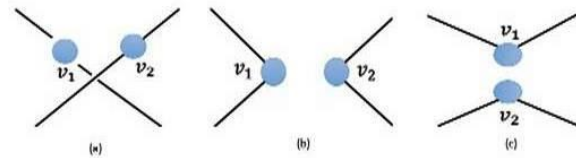


Fig. 9. Tres formas de conectar pares de aristas en un vértice de $G_{w'}$.

Se pueden seguir las aristas y vértices de la cadena w' en la multigráfica $G_{w'}$, se obtiene un circuito Euleriano que corresponde al MIC del ciliado en estudio, esto es equivalente a haber elegido en cada vértice una de las tres formas para conectar dos de sus aristas, como se explicó anteriormente. Si se eliminan en $G_{w'}$ las aristas etiquetadas con M_i para $i=1, \dots, 9$; se obtiene una gráfica con tres componentes conexas, que producen una partición de las aristas etiquetadas con I_j , $j=0, 1, \dots, 9$; estas son $\{I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9, I_0\}$. Lo mencionado que concierne a la teoría de las multigráficas 4- regulares [4][5], se podría pensar que solo es otra manera de leer la información del MIC y el MAC, pero tiene otras vertientes relacionadas con el estudio del ADN, como en el estudio de la distancia de inversión entre los cromosomas de dos especies diferentes [8],[9],[10]; que no es objeto de este artículo.

Considérese nuevamente la gráfica $H_{w'}$, para $H_{w'}$ se puede construir su gráfica de intersección que se denotará como $G_{w'}^c$, de la siguiente manera: por cada arista de $H_{w'}$ se pone un vértice de $G_{w'}^c$ y dos vértices de $G_{w'}^c$ son adyacentes si las

aristas correspondientes en H_w se intersectan. La gráfica G_w^c se denomina gráfica circular. En la Fig. 10, puede observarse la gráfica circular de G_w^c .

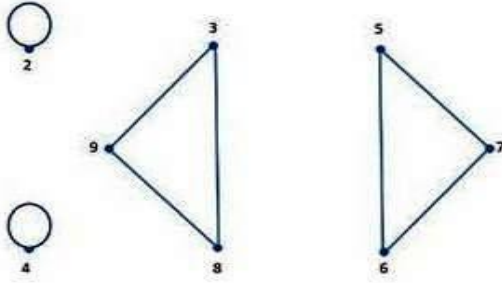


Fig. 10. Gráfica G_w^c .

D. Gráficas circulares

Las gráficas circulares se pueden ver como estructuras combinatorias que tienen una teoría propia. Se hablará de ellas en general, para después relacionarlas con el modelo de ciliados que se está estudiando en este trabajo. La teoría de las gráficas circulares fue desarrollada por Bouchet y caracterizadas en [6]. Sea G una gráfica $G=(V,E)$, donde V es un conjunto finito de vértices y $E=\{(u,v) \mid u, v \in V\}$ un conjunto de aristas, en G se permiten lazos, pero no aristas múltiples, así se considera $\{v\}$ si y solo si v es un vértice lazo. Para $X \subset V$, $G[X]$ denota a la subgráfica de G inducida por X .

Se definen operaciones básicas de complementación local y complementación de aristas sobre G . Para $v \in V$, se considera $N(v)$ la vecindad de v , definida como $N(v)=\{u \in V \mid (v,u) \in E, u \neq v\}$. La primera operación se denota por $G * v$ y consiste en reemplazar la subgráfica inducida $G[N(v)]$ por su complemento, las aristas en $V \setminus N(v)$ son las mismas en G y en $G * v$. La segunda operación, se realiza sobre una arista $e=(u,v)$ de G y se denota por $G * \{(u,v)\}$, se realiza como sigue. Para un vértice u se define su vecindad cerrada como $N(u) \cup \{u\}$ y se denota por $N(u)$. Las vecindades de u y v se pueden partir en tres conjuntos $N(u) \setminus N(v)$, $N(v) \setminus N(u)$ y $N(u) \cap N(v)$. La gráfica $G * \{(u,v)\}$ se obtiene complementando todos los pares (x,y) tales que x y y son cada uno vecinos de u o v , pero que no están en la misma clase de la partición. Así, $G * \{(u,v)\}$ no cambiará las adyacencias entre vértices que no son adyacentes a u y v , ni los lazos. Un ejemplo de estas operaciones puede verse en la Fig. 11.

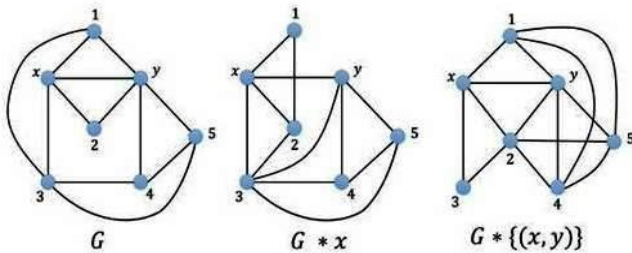


Fig. 11. a) Gráfica G , b) $G * v$, c) $G * \{(x,y)\}$.

En la gráfica circular se pueden considerar vértices con lazos, para modelar la siguiente situación: en una cadena formada

por los punteros de una cadena, se codifica su orientación relativa, para cada par de punteros que tienen diferente orientación p y \bar{p} , se agrega un lazo en el vértice que corresponde a p , en la gráfica circular correspondiente. Sea G la gráfica circular correspondiente a una cadena de código genético, en los conceptos anteriores se llamó a este concepto una palabra de doble ocurrencia. Las operaciones de recombinación vistas en la gráfica G consisten en: de doble lazo y recombinación de horquilla sobre las cadenas del código genético coinciden con las operaciones $G * v$ y $G * \{(u,v)\}$ reciprocamente, si G es la gráfica circular correspondiente. En la gráfica circular se pueden considerar vértices con lazos, para modelar la siguiente situación: en una cadena formada por los punteros, se codifica su orientación relativa, para cada par de punteros que tienen diferente orientación p y \bar{p} , se agrega un lazo en el vértice que corresponde a p , en la gráfica circular correspondiente. En este contexto, la escisión de lazo corresponde en la gráfica circular a remover un vértice aislado que no tiene lazo. La recombinación de horquilla corresponde en la gráfica G a realizar $G * v$ seguida del borrado del vértice involucrado v que corresponde al puntero en el que se realiza la operación. Respecto a la recombinación de doble lazo que involucre dos punteros diferentes p_1 y p_2 , en la Fig. 5 y que corresponden en G a los vértices u y v , esta operación en G coincide con la operación $G * \{(u,v)\}$, seguida del borrado de los vértices u y v .

IV. RESULTADOS Y DISCUSIÓN

Es importante comentar que este trabajo se basa en una serie de investigaciones plasmadas en [2][3][13], por ello está pensado como un artículo de divulgación.

Las cadenas de información del ADN, pueden verse como combinaciones de las cuatro letras de las bases presentes en el ADN, como consecuencia de este hecho, se puede pensar en construir un modelo matemático para manejar las cadenas y los cambios que ocurren en los procesos de recombinación genética. La teoría de gráficas circulares, que es parte de la teoría de delta-matroides, desarrollada por Bouchet en múltiples artículos, algunos de ellos [5][6][11], permite construir un modelo fiel para ciliados, pues las operaciones de recombinación que ocurren en su micronúcleo, durante los procesos de replicación, coinciden con las operaciones sobre la gráfica circular que se construye mediante el proceso explicado en este artículo. Se puede ampliar el modelo, construyendo la matriz de adyacencia de la gráfica circular G_w^c e introduciendo el concepto de delta-matroide para interpretar las operaciones de recombinación en este contexto.

Es importante comentar que el modelo aquí expuesto puede utilizarse para calcular la distancia entre cromosomas de diferentes especies de seres vivos, a esa distancia se le conoce como la distancia de inversión, puede verse [10][16][17][18].

Las ramas de matemáticas involucradas en este modelo son las palabras de doble ocurrencia, las multigráficas 4-

regulares y las gráficas circulares que forman parte de una teoría más extensa, los delta-matroides. Las multigráficas 4-regulares también pueden verse como una parte de los sistemas de isotropía que introdujo Bouchet [12] y que es parte del estudio, por una parte de ciertos matroides binarios y por otra parte de los 3-matroides también debidos a Bouchet. Otros materiales que involucran las multigráficas 4-regulares, pueden verse en [14] y [15].

V. CONCLUSIONES

El entendimiento del código genético que se encuentra en el ADN de toda célula es fundamental para conocer el fenómeno de la vida y de su evolución en la Tierra. El modelo matemático basado en los tópicos mencionados como parte de la teoría de delta-matroides proveen de una herramienta profunda para su comprensión y estudio.

AGRADECIMIENTOS

Las autoras agradecen el apoyo de la Universidad Autónoma Metropolitana-A (México) para la realización del presente trabajo.

REFERENCIAS

- [1] D. Quammen. El árbol enmarañado. Penguin Random House Grupo Editorial, S.A.U. (2019)
- [2] R. Brijder. Sorting by Reversals and the Theory of 4-Regular Graphs. arXiv preprint arXiv:1701.07463. (2017). 11
- [3] R. Brijder. Recombination faults in gene assembly in ciliates modeled using multimatroids. *Theoretical Computer Science*, 608. (2015). pp 27-35.
- [4] A. Kotzig. Eulerian lines in finite 4-valent graphs and their transformations. *Theory of graphs, Proceedings of the Colloquium, Tihany, Hungary*. (1966) pp 219-230. Academic Press, New York (1968)
- [5] A. Bouchet. Compatible Euler tours and supplementary Eulerian vectors. *European journal of combinatorics*, vol. 14, no 6, (1993). pp 513-520.
- [6] A. Bouchet. Circle graph obstructions. *Journal of Combinatorial Theory, Series B*, 60(1), (1994). pp 107-144.
- [7] L. Traldi. Binary matroids and local complementation. *European Journal of Combinatorics*, 45, (2015) pp 21-40
- [8] A. Caprara. Sorting permutations by reversals and Eulerian cycle decompositions. *SIAM journal on discrete mathematics*, 12(1), (1999). pp 91-110.
- [9] L. A. García García. Un modelo combinatorio en ensamblamiento genético. Tesis Maestría en Optimización. UAM-A, (2018)
- [10] L. A. García García, M. G. Rodríguez. Cálculo de la distancia de inversión cromosómica usando estructuras matroidales. *Memorias de la Sociedad Matemática Mexicana*, Vol. 15, (2019). ISSN 1870-2112.
- [11] A. Bouchet. A. Bouchet. k-transformations, local complementations and switching. In *Cycles and Rays* Springer Netherlands. (1990). pp. 41-50.
- [12] A. Bouchet. Graphic presentations of isotropic systems. *Journal of Combinatorial Theory, Series B*, 45(1), (1988). pp 58-76.
- [13] S. Hannenhalli, P. A. Pevzner. Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. *Journal of the ACM (JACM)*, 46(1), (1999) pp 1-27.
- [14] B. Jackson. A characterisation of graphs having three pairwise compatible Euler tours. *Journal of Combinatorial Theory, Series B*, 53(1), (1991). pp 80-92.
- [15] L. Traldi. Interlacement in 4-regular graphs: a new approach using nonsymmetric matrices. arXiv preprint arXiv:1204.0482. (2012)
- [16] V. Bafna, P. A. Pevzner. Genome rearrangements and sorting by reversals. *SIAM Journal on Computing*, (1996). 25(2):272–289.
- [17] A. Bergeron. A very elementary presentation of the Hannenhalli–Pevzner theory. *Discrete Applied Mathematics*, 146(2), (2005). pp 134-145.
- [18] S. Grusea, A. Labarre. The distribution of cycles in breakpoint graphs of signed permutations. *Discrete Applied Mathematics*, 161(10). (2013) pp 1448-1466.