



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

ANÁLISIS Y DISEÑO DE OPERADORES EPIGENÉTICOS PARA
ALGORITMOS EVOLUTIVOS

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS E INGENIERÍA DE LA COMPUTACIÓN

PRESENTA:
DIEGO DE JESÚS ISLA LÓPEZ

TUTOR PRINCIPAL
DRA. KATYA RODRÍGUEZ VÁZQUEZ
IIMAS

CIUDAD UNIVERSITARIA, CD. MX., 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: Dr. Gibran Fuentes Pineda
Secretario: Dr. José Antonio Neme Castillo
Vocal: Dra. Katya Rodríguez Vázquez
1er. Suplente: Dr. Augusto César Poot Hernández
2o. Suplente: Dra. Yuriria Cortés Poza

La tesis se realizó en Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas.

TUTOR DE TESIS:

Dra. Katya Rodríguez Vázquez

A la memoria de:

Beatriz Ramos Rodríguez por su amor.

Dr. Héctor Adolfo Andrade Gómez por sus enseñanzas e inspiración.

Reconocimientos

A la UNAM y al IIMAS por la oportunidad de cursar este posgrado y por la educación y enseñanzas que obtuve en él.

A la Dra. Katya Rodríguez Vázquez por su guianza, consejos y comprensión durante la realización de este proyecto.

A mi familia por el amor y apoyo incondicional.

A mis amigos: los de toda la vida, que no dejaron de darme ánimos, y los nuevos que hice en esta parte del camino, en especial a Víctor y a los *Minimum spanning teens*: Yadira, Saúl y Eduardo.

Al CONACYT por el apoyo de la beca del PNPC para poder lograr mis estudios de maestría.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea el resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Diego de Jesús Isla López. Ciudad Universitaria, Cd. Mx., 2022

Resumen

El cómputo evolutivo es un área muy activa dentro de la inteligencia artificial e importante para la resolución de diversos problemas de optimización. Por su parte, el estudio de los fenómenos epigenéticos han significado un parteaguas para la biología en la comprensión de la dinámica de la evolución de las especies. La integración de los conocimientos actuales en epigenética es un esfuerzo reciente en el diseño de algoritmos evolutivos y abre las puertas a la utilización de modelos relacionados con esta nueva área de investigación biológica. En esta tesis se propone utilizar la hipótesis de asignación diferencial como mecanismo de selección y remplazo dentro del marco de referencia propuesto por Stolfi y Alba, buscando evaluar el desempeño de un algoritmo evolutivo que incluya ideas relacionadas con fenómenos epigenéticos en todas sus etapas.

Abstract

Evolutionary computing is a very active area in artificial intelligence, its importance being mostly directed in solving various optimization problems. At the same time, the study of epigenetic phenomena has been a turning point for biology for the comprehension of the evolutionary dynamics of species. The integration of current knowledge in the area of epigenetics into evolutionary algorithms is a recent effort and opens the door to new insights on the design of such algorithms, utilizing models related to such area of research in biology. This thesis focuses on the proposal of using the differential allocation hypothesis as a method of selection and replacement in the framework proposed by Stolfi and Alba, looking forward to evaluate the performance of an evolutionary algorithm that includes mechanisms inspired by epigenetic phenomena in all its stages.

Índice general

Índice de figuras	xv
Índice de tablas	xvii
1. Introducción	3
1.1. Objetivo	4
1.2. Motivación	4
1.3. Contribuciones	5
1.4. Estructura de la tesis	5
2. Motivación biológica	7
2.1. Dogma central de la biología molecular	8
2.2. Epigenética	9
2.2.1. Regulación genética y mecanismos epigenéticos	11
2.2.2. Epigenética y la influencia del entorno	11
2.3. Hipótesis de asignación diferencial	12
2.3.1. Modelo de Haaland <i>et al.</i>	14
2.4. Epigenética, cuidados parentales y su rol en la evolución	16
2.5. Comentarios finales	17

3. Estado del arte en algoritmos genéticos	19
3.1. Estructura de la población	20
3.2. Operadores de selección	21
3.3. Operadores de cruza	23
3.4. Operadores de mutación	24
3.5. Estrategias de remplazo	24
3.6. Comentarios finales	25
4. Cómputo y epigenética	27
4.1. Introducción	27
4.2. Propuestas existentes de algoritmos inspirados en epigenética	29
4.3. Propuesta de algoritmo epigenético de Stolfi y Alba	31
4.3.1. Estructura de la población	32
4.3.2. Reproducción basada en nucleosomas	33
4.3.3. Silenciamiento de genes	35
4.4. Comentarios finales	35
5. Propuesta de operadores epigenéticos	37
5.1. Características de la población	38
5.2. Selección	38
5.3. Ajuste en reproducción nucleosómica	39
5.4. Remplazo	39
5.5. Comentarios finales	42
6. Experimentación y resultados	43
6.1. Características del entorno de evaluación	43

6.2. Problemas de prueba	43
6.3. Calibración de parámetros y métodos de comparación de resultados . .	44
6.4. Pruebas preliminares	45
6.5. Resultados de epiGA con modificaciones	50
6.6. Análisis	61
7. Conclusiones y discusión	65

Índice de figuras

2.1. Esquema del dogma central de la biología molecular	9
2.2. Esquema del paisaje epigenético de Waddington. Tomada de [27]	10
2.3. Cromosomas y factores epigenéticos	10
2.4. Diamante cebrá o mandarín (<i>Taeniopygia guttata</i>)	13
4.1. Línea del tiempo del avance en descubrimientos sobre epigenética	28
4.2. Estructura de los individuos en epiGA. Tomada de [48].	33
6.1. Opciones de métodos de comparación de desempeño para algoritmos de optimización. Tomada de [12].	45
6.2. Resultados preliminares para el problema 1	46
6.3. Resultados preliminares para el problema 2	46
6.4. Resultados preliminares para el problema 3	47
6.5. Resultados preliminares para el problema 4	47
6.6. Resultados preliminares para el problema 5	47
6.7. Resultados preliminares para el problema 6	48
6.8. Resultados preliminares para el problema 7	48
6.9. Resultados preliminares para el problema 8	48
6.10. Resultados preliminares para el problema 9	49

ÍNDICE DE FIGURAS

6.11. Resultados preliminares para el problema 10	49
6.12. Comparación de resultados para el problema 1	51
6.13. Comparación de resultados para el problema 2	52
6.14. Comparación de resultados para el problema 3	53
6.15. Comparación de resultados para el problema 4	54
6.16. Comparación de resultados para el problema 5	55
6.17. Comparación de resultados para el problema 6	56
6.18. Comparación de resultados para el problema 7	57
6.19. Comparación de resultados para el problema 8	58
6.20. Comparación de resultados para el problema 9	59
6.21. Comparación de resultados para el problema 10	60

Índice de tablas

2.1. Parámetros del modelo de Haaland <i>et al.</i>	15
4.1. Parámetros de entrada de epiGA	32
6.1. Características de los problemas de prueba	44
6.2. Configuración de parámetros utilizada en las pruebas preliminares en epiGA	45
6.3. Configuración de parámetros utilizada en las pruebas preliminares en epiDA	50
6.4. Mejores soluciones encontradas por cada algoritmo	61
6.5. p -valores preliminares	61
6.6. Resultados de posicionamiento para cada algoritmo	62
6.7. p -valores ajustados para prueba de Friedman	62
6.8. p -valores ajustados para prueba de posiciones alineadas de Friedman	62
6.9. p -valores ajustados para prueba de Quade	62

Índice de algoritmos

1.	Algoritmo genético simple (SGA)	20
2.	EGA	30
3.	Cruza con metilación	30
4.	ReGen GA	31
5.	epiGA	32
6.	Reproducción nucleosómica	34
7.	Silenciamiento de genes	35
8.	Selección inspirada en DA	39
9.	Reemplazo inspirado en DA	41

Celebro, al fin,
a esa primera célula organizada,
a la primera huérfana
y la última, a ella, inmaculada madre unicelular,
sin pecado concebida, bendita
entre toda la materia estéril.

Elisa Díaz Castelo - Oda a los ancestros (fragmento)

His own parents, he that had father'd him and she that had conceiv'd him
in her womb and birth'd him,
They gave this child more of themselves than that,
They gave him afterward every day, they became part of him.

Walt Whitman - There was a child went forth (fragmento)

Why not engineer all the mistakes away?
We could evolve into machines.
Then there wouldn't be cells
To proliferate
Uncontrollably.
There wouldn't be thoughts
Only wires.
I wouldn't end up at your door.
I wouldn't care
About the valleys,
Mountain ranges,
That your white cotton shirt stretches over.
We could be ones and zeroes-
A code for no heartbreak.

Ann Beaver - Engineer to Evolve

Capítulo 1

Introducción

La investigación y desarrollo de algoritmos evolutivos es un área muy activa tanto en el diseño de nuevas propuestas de algoritmos como en la aplicación de los mismos en distintos problemas. A su vez, en años recientes la investigación en epigenética ha cobrado mayor interés dentro de la comunidad de biología pues han surgido ideas relacionadas con el papel que tiene la epigenética en distintos aspectos de la vida, especialmente para los humanos.

La bioinformática no es del todo ajena a las ideas de la epigenética, pues existe un área llamada “epigenética computacional” la cual se encarga de desarrollar e implementar herramientas computacionales para el estudio y análisis de datos e información relacionados con epigenética. Sin embargo, en el ámbito de los algoritmos evolutivos la exploración de alternativas que introduzcan las ideas de la epigenética es escasa y muy reciente. El marco de referencia para el diseño de algoritmos epigenéticos propuesto por Stolfi y Alba en 2018 es un punto de partida interesante para proponer nuevos algoritmos evolutivos desde la perspectiva de la epigenética. Este marco de referencia deja a elección libre los métodos de selección y de remplazo en su propuesta original. Por esta razón en este trabajo se proponen dos métodos inspirados en la hipótesis de asignación diferencial (DA), un fenómeno observado principalmente en aves y relacionado con el cuidado parental, teniendo así un esquema que incorpore la idea de la existencia de factores epigenéticos influidos por el entorno (en este caso, los cuidados parentales) en el algoritmo.

1.1. Objetivo

Este trabajo tiene por objetivo presentar el diseño e implementación de un método de selección y otro de mutación inspirados en el fenómeno de asignación diferencial, el cual se elige por su relación con la idea de la influencia del entorno, especialmente de los cuidados parentales, en los procesos epigenéticos de los individuos de diferentes especies de organismos. Se realizarán análisis de los resultados obtenidos por el algoritmo y se harán comparaciones de los mismos con los obtenidos por otros algoritmos.

1.2. Motivación

La motivación de este trabajo viene de dos enfoques. El primero, el enfoque biológico, está dado por el reciente crecimiento en interés por el estudio de la epigenética y sus efectos y consecuencias en diferentes ámbitos de la vida, tanto en humanos como otros animales y plantas. La comprensión de estos efectos es un objetivo importante pues se ha observado el papel que ellos juegan en la propensidad (o resistencia) a enfermedades [47], en la calidad de vida y en la prevalencia y desaparición de diferentes características a lo largo de las generaciones [55, 50], por nombrar algunos de los ámbitos antes mencionados. Un factor importante en la dinámica de los efectos epigenéticos es la influencia del entorno en el que viven los organismos. Un ejemplo importante de este tipo de factores es el cuidado parental sobre su descendencia. Es por esta razón que se elige el fenómeno de DA como inspiración para diseñar los operadores propuestos en este trabajo, ya que implica la asignación de atributos atractivos en la descendencia producida en cada generación, teniendo esto consecuencias en la dinámica de cortejo y en el nivel de cuidado hacia la progenie por parte de los individuos de las especies en las que se ha observado este fenómeno.

El segundo enfoque, el computacional, tiene que ver con el panorama actual de la investigación en algoritmos evolutivos, en especial de los algoritmos genéticos. En la actualidad los algoritmos genéticos son herramientas de optimización muy utilizadas en distintos tipos de problemas. Sin embargo, las modificaciones propuestas para el algoritmo han sido mínimas debido a la solidez de los diferentes operadores y métodos ya conocidos, además de que dichas modificaciones suelen ser dependientes del problema al que se aplique el algoritmo. De este modo, al mismo tiempo que la investigación en epigenética toma cada vez más fuerza en biología, la integración de este nuevo conocimiento puede significar una nueva dirección en el estudio de los algoritmos evolutivos. Así, el conocimiento generado de este trabajo puede poner en evidencia la necesidad de mayor interacción entre las ciencias biológicas y las computacionales.

1.3. Contribuciones

Este trabajo busca contribuir con nuevos métodos y perspectivas que ayuden a avivar el interés por la inclusión de nuevas técnicas inspiradas en procesos epigenéticos dentro de algoritmos genéticos y evolutivos. Como ya se ha mencionado, la investigación del paradigma epigenético en algoritmos evolutivos es muy reciente, por lo que es interesante explorar los diferentes caminos y posibilidades que ofrece el integrar estas ideas dentro de los algoritmos.

1.4. Estructura de la tesis

Este trabajo de tesis se encuentra dividido en seis capítulos:

- El capítulo 1 explora la motivación biológica detrás del proyecto, dando una introducción a la epigenética y a la hipótesis de asignación diferencial.
- El capítulo 2 discute el estado del arte de los algoritmos genéticos enfocándose en los diferentes operadores propuestos recientemente.
- En el capítulo 3 se habla en detalle del marco de trabajo para algoritmos epigenéticos propuesto por Stolfi y Alba.
- El capítulo 4 presenta la propuesta de los operadores en los que se enfoca este trabajo de tesis.
- El capítulo 5 muestra los detalles de la experimentación y los resultados obtenidos, así como la comparación con el desempeño de otros algoritmos existentes en la literatura.
- Finalmente, en el capítulo 6 se presentan las conclusiones a las que se llega en este trabajo.

Capítulo 2

Motivación biológica

En la década de 1920 se acuñó el término de “síntesis moderna” de la biología, que actualmente se conoce como *neodarwinismo*. La finalidad de crear dicho término fue integrar las diferentes teorías existentes en ese entonces sobre la evolución de las especies. Desde entonces todos los descubrimientos y áreas de la biología que han surgido han hecho que se considere necesario el crear una nueva teoría integrativa. Una de las propuestas es la que actualmente se conoce como *síntesis evolutiva extendida* (EES por sus siglas en inglés) que, entre otras adiciones, incluye a la teoría epigenética como parte fundamental para tener un mejor entendimiento de los procesos evolutivos. Laland *et al.* [32] mencionan la importancia de integrar técnicas de las matemáticas y las ciencias de la computación para la EES, ya que reconoce su utilidad en la elaboración de modelos que ayuden a la comprensión de la dinámica de diferentes procesos evolutivos, desde el nivel genético hasta la dinámica de poblaciones.

El punto de vista neodarwinista es la base de los algoritmos evolutivos. Con el mismo espíritu de la EES, el presente proyecto de tesis busca integrar ideas relacionadas a la epigenética y al cuidado parental en el diseño de algoritmos evolutivos, buscando así generar nuevas ideas y perspectivas sobre las capacidades de dicho tipo de algoritmos.

En este capítulo se hablará de los fundamentos biológicos del proyecto descrito en este trabajo de tesis. En primera instancia se hablará de la estructura básica del material genético como puede encontrarse en los organismos eucariontes, pasando después a explicar de manera más detallada el campo de estudio de la epigenética. Posteriormente, se centrará el enfoque hacia la influencia del medio ambiente en los efectos epigenéticos, en especial en los cuidados parentales, describiendo el fenómeno de Asignación diferencial en el resto de la extensión del capítulo.

2.1. Dogma central de la biología molecular

El ácido desoxiribonucleico (ADN), la molécula que contiene el material genético, se encuentra alojado en el núcleo de las células eucariotas y se compone de dos cadenas conformadas por azúcares y grupos fosfato conectadas mediante bases nitrogenadas. Estas bases son guanina, timina, citosina y adenina, representadas por las letras G, T, C y A, respectivamente. A su vez, el ADN se apoya de otra molécula llamada ácido ribonucleico (ARN), la cual consta de una sola cadena de las mismas bases nitrogenadas que el ADN a excepción de la timina, la cual es sustituida por uracilo (U).

A las secciones del ADN que codifican una instrucción en particular se les conoce como **genes**. Estas instrucciones tienen como resultado la creación de proteínas, las cuales determinan las funciones celulares. Cabe mencionar la existencia de genes *no codificantes*, los cuales en lugar de dar lugar a proteínas, contienen instrucciones para la regulación genética; a su vez, existen otros genes no codificantes que producen diferentes tipos de ARN. Los genes pueden variar en longitud y el número total de ellos también varía entre diferentes especies de organismos. Dado de que la molécula de ADN tiene una longitud muy grande (alrededor de 2 metros en el caso de los humanos), es necesario compactarla para que pueda almacenarse en el núcleo de cada célula. En las células eucariotas, la cadena de ADN se enrolla a un octámero (grupo de ocho moléculas) de proteínas llamadas **histonas** para generar la estructura conocida como **nucleosoma** [34]. A la cadena de nucleosomas, en combinación con otras proteínas no histonas y ARN, se le conoce como **cromatina**, la cual es la base para crear **cromosomas**. Los cromosomas, entonces, son la estructura final de compactación del material genético.

La idea de *dogma central de la biología molecular* es indispensable ya que introduce el concepto de *expresión génica*, es decir, la producción de proteínas de acuerdo a la instrucción codificada en cada gen. Francis Crick, co-descubridor del ADN, fue quien propuso esta teoría en 1958, profundizando en ella en 1970 [15]. Crick postula en esta teoría los posibles escenarios de transferencia de información genética, siendo los tres principales la **replicación** (ADN a ADN), la **transcripción** (ADN a ARN) y la **traducción** (ARN a proteína). A su vez, Crick menciona los casos especiales, los cuales ocurren bajo condiciones específicas, en especial como parte del ciclo de producción de ciertos tipos de virus. Estos casos son la **transcripción inversa** (ARN a ADN) y la replicación de ARN (ARN a ARN). Finalmente, se estipula que una vez que ocurre el proceso de traducción (es decir, se genera una proteína), es imposible revertir la estructura de la proteína creada hacia ácidos nucleicos por métodos naturales. Sin embargo, el proceso de traducción inversa se puede realizar por medios sintéticos en un laboratorio. En la figura 2.1) se ilustra mediante un diagrama el orden en el que ocurren los diferentes procesos descritos en el dogma central, siendo las flechas de color naranja los casos principales y las flechas punteadas de color verde los casos especiales.

⁰Créditos: Dhorspool/Wikipedia (CC BY-SA 3.0)

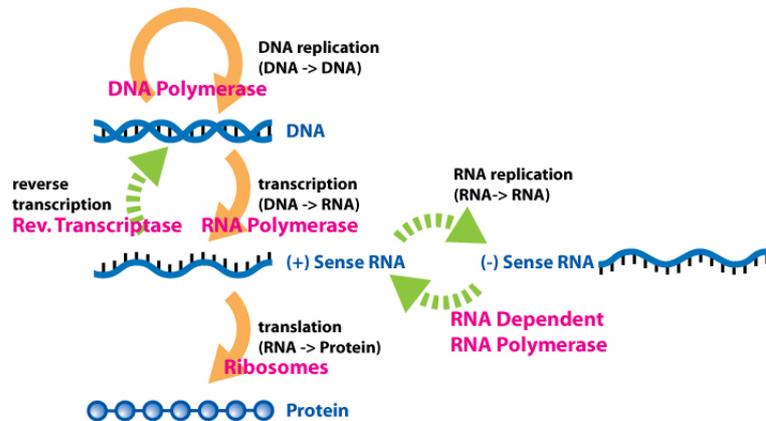


Figura 2.1: Esquema del dogma central de la biología molecular)¹

2.2. Epigenética

La epigenética (del prefijo *epi-*: sobre, más allá) se refiere al estudio tanto de los cambios heredables de la actividad y expresión genética como de las alteraciones a largo plazo en el potencial de transcripción de la célula, las cuales no son necesariamente heredables [43]. Furrow *et al.* [17] mencionan que la creación de métodos de análisis de la variación epigenética a través del tiempo de vida de un organismo como un reto a superar en esta década para impulsar el avance del campo de estudio de la epigenética.

El término epigenética fue utilizado por primera vez por Conrad Waddington, en primera instancia como lo que llamó “paisaje epigenético” (figura 2.2) para después acuñar el término que utilizamos en la actualidad, definiendo la epigenética como “la rama de la biología que estudia las interacciones causales entre los genes y sus productos, lo que hace que surja el fenotipo” Waddington [54]. La idea del paisaje epigenético sirvió como una metáfora para los diferentes caminos que puede tomar una célula para terminar formando parte de un órgano en específico, llamándolos *creodes*. Por su parte, Barbara McClintock previó los fenómenos que ahora identificamos como metilación de ADN y acetilación de histonas durante sus estudios sobre la adaptabilidad del maíz, intuyendo la existencia de elementos externos a la cromatina cubriendo secciones de ésta. Cabe mencionar que McClintock también previó la existencia de lo que llamó “genes saltarines”, ahora conocidos como transposones, por lo cual fue acreedora del premio Nobel. Vineis [53]

2. MOTIVACIÓN BIOLÓGICA

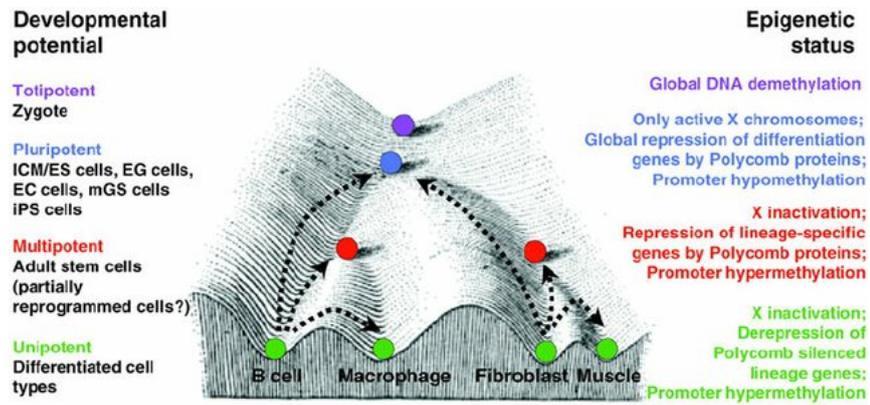


Figura 2.2: Esquema del paisaje epigenético de Waddington. Tomada de [27]

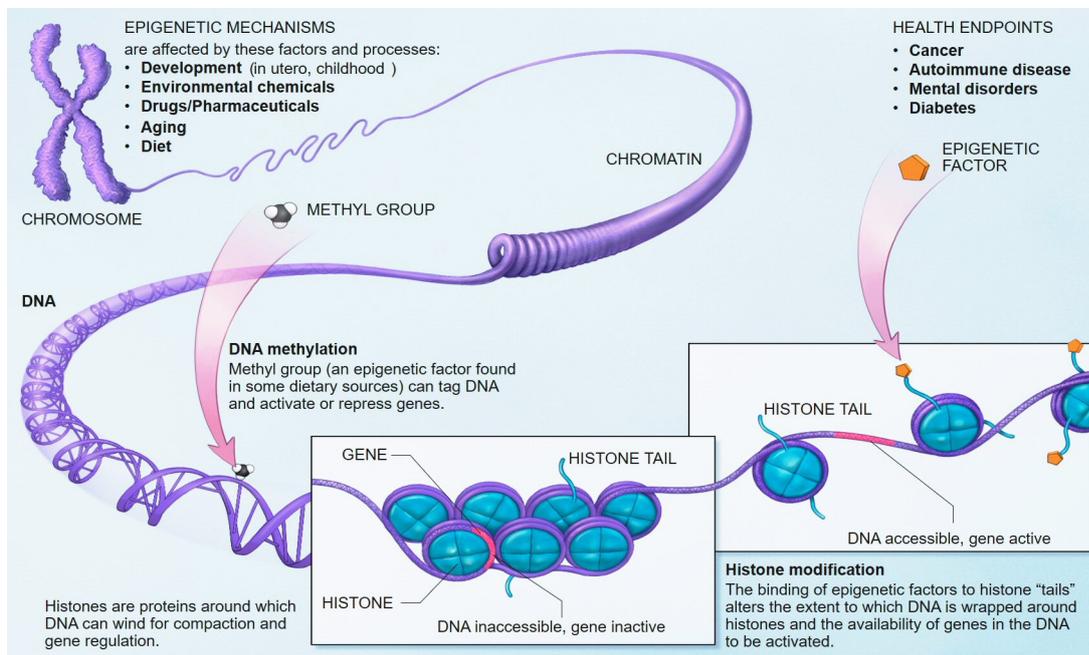


Figura 2.3: Cromosomas y factores epigenéticos¹

2.2.1. Regulación genética y mecanismos epigenéticos

Vineis [53] menciona tres de los mecanismos epigenéticos más estudiados, a pesar de no ser los únicos identificados (mostrados en la figura 2.3). El primero es el proceso de **metilación de ADN**, el cual consiste en la adición de un grupo metilo en una secuencia CpG (citosina seguida de guanina), provocando que dicha secuencia quede inactiva. El segundo proceso se conoce como **modificación de histonas**, el cual engloba diferentes escenarios. Los dos escenarios más estudiados son la acetilación y deacetilación. Si existe acetilación de las histonas, la distancia entre nucleosomas se extiende; es decir, la cromatina se relaja y deja expuestas regiones de ADN que pueden ser leídas y eventualmente expresadas. Por el contrario, la deacetilación provoca que los nucleosomas se aglomeren, haciendo imposible que se puedan expresar las regiones de ADN entre ellos. Por último, el tercer proceso es el **silenciamiento por miRNA** (micro ARN). Las moléculas de miRNA se acoplan a las moléculas de mRNA (ARN mensajero) producidas durante la etapa de transcripción haciendo que se inactiven ya sea cortándolas o bloqueando alguna sección específica que no pueda ser leída por la enzima encargada de la traducción. Por lo tanto, la etapa de traducción se cancela y no se produce una nueva proteína.

2.2.2. Epigenética y la influencia del entorno

Los mecanismos mencionados en el apartado anterior son generalmente provocados por estímulos externos, tanto en la etapa de desarrollo embrionario como después del nacimiento de un individuo. La dieta y el metabolismo son dos de los factores con consecuencias epigenéticas más significativas. De este modo, se ha determinado que el entorno y las experiencias de vida de los organismos tienen efectos sobre la salud y bienestar de los mismos, fenómeno al que se le llama “incrustación biológica” (*biological embedding*) [5]. Guerrero-Bosagna [21] menciona que existe evidencia sustancial de fenómenos como la falta de correlación entre cambios genéticos y el aumento de aptitud en los individuos de una especie. Esto quiere decir

Selección y efectos parentales

Debido a la variedad de dinámicas de cortejo presentes en las diferentes especies animales, es imposible ignorar el papel de la selección sexual en la dinámica epigenética. Durante mucho tiempo se ha observado la importancia que tienen diversos atributos físicos en la selección de parejas sexuales en diferentes especies animales. Estas carac-

¹Crédito: *National Institutes of Health* (Dominio público)

2. MOTIVACIÓN BIOLÓGICA

terísticas son producto del dimorfismo sexual presente en la mayoría de estas especies, dando atributos específicos a miembros de los diferentes sexos que como parte del proceso de evolución, cambian según las preferencias de cada grupo de individuos. Ejemplos muy conocidos de esto son los diferentes patrones de colores, la preferencia con respecto al tamaño o fuerza del macho por parte de la hembra, entre otros. La idea predominante de la explicación de los mecanismos detrás de la aparición y asignación de estos atributos es la hipótesis de los “genes buenos”; es decir, los atributos antes mencionados se encuentran codificados en el código genético de aquellos individuos con mayor aptitud. De este modo, son más fáciles de reconocer por posibles parejas sexuales y la probabilidad de participar en un ciclo reproductivo aumenta. Un ejemplo de un modelo que aborda estas ideas son los procesos Fisher-Lande-Kirkpatrick (FLK). Sin embargo, Achorn y Rosenthal [1] mencionan evidencia existente de que la idea de los genes buenos representa una fracción más bien modesta de la aptitud sexual total de los individuos.

Un aspecto que es posible pasar por alto es la habilidad de un individuo, de cualquiera de los dos grupos sexuales, de no solo producir descendencia con buena aptitud, sino de proveer los cuidados necesarios para aumentar la probabilidad de que esta descendencia tenga un desarrollo adecuado y logre sobrevivir. A estos comportamientos se les llama **efectos parentales**, los cuales se pueden dividir en **paternos** si vienen del padre y **maternos** si vienen de la madre. Qvarnström y Price [37] mencionan varios ejemplos en los que los efectos maternos y paternos influyen en la selección de parejas. Como mencionan Reyes y Klug [40], los procesos de selección sexual y de efectos parentales están fuertemente ligados, además de que hacen notar la importancia de los factores que pueden hacer que alguno de los dos sexos tenga preferencias más específicas; al final de sus conclusiones exaltan la importancia de profundizar en el estudio y comprensión de la relación entre inversión parental y las estrategias de apareamiento.

2.3. Hipótesis de asignación diferencial

La hipótesis de asignación diferencial (DA) fue presentada por primera vez por Burley en [10, 9]. En palabras de Burley, la hipótesis formula el favorecimiento de la selección natural hacia los individuos que asignan recursos reproductivos más costosos en proporción al atractivo de apareamiento de sus parejas. El término *atractivo de apareamiento* se refiere a fenotipos específicos preferidos por los miembros de un sexo en una población. Burley define dos tipos de beneficios derivados de la asignación diferencial: los beneficios directos son aquéllos que influyen en el número de individuos gestados en cada generación; por su parte, los beneficios indirectos son aquéllos que afectan la aptitud de dichos nuevos individuos.

Se describe el efecto del cuidado paterno sobre la aptitud de la descendencia en térmi-



Figura 2.4: Diamante cebra o mandarín (*Taeniopygia guttata*)¹

nos de atributos atractivos y su supervivencia. La hipótesis además aborda la selección de parejas sexuales basada en estos atributos, principalmente por parte de las hembras. Burley reportó el fenómeno al observarlo durante su trabajo con aves de la especie *Taeniopygia guttata* llamada coloquialmente como diamante cebra o mandarín (figura 2.4). La hipótesis ganó fuerza al observarse fenómenos similares e incluso opuestos en otras especies, como el identificado por Gowaty [20] al que llamó teoría de compensación reproductiva (RC). En RC, Gowaty [20] describe la disminución de inversión parental por parte de individuos que se aparean con sus parejas predilectas. Esto lleva a que la aptitud de la descendencia de individuos que se cruzan con parejas no preferidas sea menor a la de los individuos que se aparean con sus parejas preferidas. En [19], Gowaty enumera las principales diferencias entre RC y DA:

1. En DA se considera que los rasgos atractivos regulan las preferencias de los individuos, mientras que en RC este no siempre es el caso.
2. RC se enfoca en mejorar la resistencia a patógenos por parte de la descendencia, mientras que DA toma en cuenta otros tipos más comunes de cuidado parental.
3. RC aplica tanto para especies que exhiben conductas de cuidado parental post-cigóticas como para las que no. En el caso de DA, es necesario que exista un cuidado aún después del nacimiento de la descendencia.
4. En RC se especifica que los miembros de ambos sexos son propensos a ejercer compensación, independientemente de si el macho o la hembra tengan la limitante de falta de atracción por su pareja. Eso indica que RC predice la cooperación en términos de compensación por parte de ambos padres, mientras que en DA no se puede hacer esta predicción.

¹Créditos: Jim Bendon/Flickr ([CC BY-SA 2.0](#))

2. MOTIVACIÓN BIOLÓGICA

Una observación importante que hace Gowaty es que RC no es aplicable si las preferencias sexuales de los individuos no se traducen en una mejora de la aptitud de la descendencia.

En el año 2000 Sheldon [46] hace un recuento del conocimiento que se tenía del fenómeno hasta la fecha. Este artículo es importante en la cronología del estudio de DA por diferentes razones:

- Enlista diferentes artículos de la época cuyas conclusiones implicarían que diferentes tipos de DA pueden ocurrir de manera simultánea.
- Menciona la importancia que tiene la DA en el estudio del posible efecto de los cuidados maternos en la evolución y variación en las especies.
- Indica una posible falla en los experimentos que se apoyan en la hipótesis de los buenos genes, al no tomar en cuenta DA como un posible escenario.

Harris y Uller [24] presentan el primer modelo matemático del fenómeno, comparando los resultados con los reportados para RC, ya que este fenómeno se había empezado a considerar como contrario a DA. Ratikainen y Kokko [39] consideran que el modelo de Harris y Uller está incompleto pues ignora algunos escenarios en los que puede suceder DA, además de que proponen integrar DA y RC como un mismo modelo por diversos factores, de esta manera cambiando el término de RC por “DA negativa”. En 2015, supervisado por Ratikainen, Haaland realiza esta integración de los modelos como su tesis de maestría y publica un artículo en 2017 [22] donde muestra con detalle los modelos diseñados y los diferentes escenarios en los que se puede presentar el fenómeno de DA.

2.3.1. Modelo de Haaland *et al.*

En su artículo, Haaland *et al.* [22] hablan con detalle de los escenarios posible para DA, haciendo la distinción entre DA positiva y DA negativa. A su vez propone un modelo estocástico (ecuación 2.1):

$$W(u_f, x, q_m, t) = b_t(u_f, q_m) + \alpha \left[\sum_{i=1}^{N_m} p_i \{ \lambda W^*(x - c(u_f, q_m) + d, q_i, t + 1) + (1 - \lambda) W^*(x - c(u_f, q_m), q_i, t + 1) \} \right] \quad (2.1)$$

que considera más apegado a la realidad utilizando diversos parámetros (tabla 2.1). El modelo es la función de aptitud para las hembras, donde los costos reproductivos dependen de la calidad del macho (asignada de acuerdo a la distribución definida en la tabla 2.1), los cuales tienen el efecto de disminuir la energía de la hembra. En los experimentos de Haaland se hicieron corridas inversas (iniciando desde el valor deseado) para obtener una configuración de parámetros, la cual después fue verificada mediante modelos de Markov.

Para entender el modelo es necesario descomponer algunas de sus partes. En primer lugar $b_t(u_f, q_m)$ se refiere a los “beneficios” obtenidos por una hembra, entendiendo por “beneficios” a la aptitud de su progenie. A $b_t(u_f, q_m)$ se le agrega la suma de las inversiones por parte de la hembra para cada tipo de calidad de los machos, siendo esta suma devaluada por la probabilidad de supervivencia (α).

Variable	Descripción
b	Función de aptitud de la descendencia
c	Función de costo de las hembras
u_f	Inversión reproductiva de las hembras
q_m	Calidad de los machos
x	Estado energético de la hembra
α	Probabilidad de supervivencia
λ	Probabilidad de encontrar comida
d	Valor energético de la comida
T	Tiempo de vida máximo
m	Calidades del macho (1 = bajo, 2 = medio, 3 = alto)
N_m	No. de categorías de calidad para los machos
P	Distribución de probabilidad de encontrarse con machos de diferentes calidades

Tabla 2.1: Parámetros del modelo de Haaland *et al.*

2.4. Epigenética, cuidados parentales y su rol en la evolución

Richards *et al.* [42] hacen un análisis extensivo del papel de las modificaciones epigenéticas sobre la evolución. Mencionan que la regulación de la expresión mediante mecanismos epigenéticos puede representar un mecanismo de generación de variaciones, junto con los procesos de mutación y recombinación. Otro punto importante que mencionan es cómo la herencia epigenética facilita la exploración del panorama de adaptabilidad. Acerca de esto, apuntan que el hecho de que un mismo genotipo pueda expresarse de diferentes maneras mediante diferentes configuraciones epigenéticas permite una capacidad de exploración mayor, llevando a la posibilidad de que individuos con genotipos de aptitud reducida puedan presentar fenotipos asociados a una aptitud alta. Además, el hecho de que las variaciones epigenéticas sean heredables puede hacer que los fenotipos deseables se mantengan por múltiples generaciones, para finalmente ser asimilados dentro del código genético.

Kappeler y Meaney [29] reconocen el papel de los efectos parentales en la dinámica de los mecanismos epigenéticos y sus posibles consecuencias evolutivas. En este artículo se hace referencia a diferentes fenómenos observados y bien documentados en la literatura, como el caso del efecto de los lamidos por parte de las madres hacia sus crías en las ratas. Asimismo, tratan otros aspectos que influyen en la configuración de los marcadores epigenéticos en cada generación, en especial el aspecto nutricional. Un punto igualmente importante es la observación de la perpetuación de conductas a las que fueron expuestas los individuos en cada generación. Es decir, tomando como ejemplo el caso de las ratas, aquellas ratas que recibieron lamidos por parte de sus madres son más propensas a repetir esta conducta con sus propias crías, que las ratas que no recibieron este tipo de cuidado.

Una de las conclusiones importantes de Haaland *et al.* [22] es que en los escenarios que utilizaron para sus experimentos encuentran una gran evidencia de que DA tiene consecuencias evolutivas al inducir variaciones en la dinámica de selección sexual, tomando en cuenta que DA le otorga especial importancia a los esfuerzos de cuidado parental al momento de elegir una pareja. A su vez, consideran que sigue siendo necesaria la creación de un modelo que describa a detalle la evolución de las decisiones tomadas por los individuos, pues concluyen que DA y la selección son procesos con un alto grado de interdependencia.

De este modo, podemos concluir que además de la transferencia de información epigenética por medio de la línea germinal, la inducción de modificaciones epigenéticas por medio de los cuidados parentales en la etapa de desarrollo de los nuevos individuos en cada generación provocan la aparición y prevalencia de rasgos fenotípicos que se-

rán determinantes en el proceso de selección sexual en cada ciclo reproductivo. Por lo tanto, el proceso evolutivo de las especies tendrá una fuerte influencia de las decisiones tomadas en cada ciclo reproductivo tomando en cuenta el resultado de los efectos parentales. Así, los individuos con mayor probabilidad de transmitir su información genética son aquellos que presentan un fenotipo más atractivo como consecuencia de los efectos epigenéticos de los efectos parentales.

2.5. Comentarios finales

En este capítulo se ha presentado una explicación breve acerca de la epigenética y su relación con los cuidados parentales y las consecuencias que estos dos fenómenos en conjunto tienen en la evolución de las especies. Ambos fenómenos y sus interacciones son temas que son estudiados con mucho interés y aunque al día de hoy no se tiene una comprensión completa de su funcionamiento, el conocimiento con el que contamos nos lleva a plantear hipótesis y a desarrollar ideas que muchas veces trascienden los límites de la biología para convertirse en proyectos interdisciplinarios. Desde el punto de vista computacional, es posible ver una relación bidireccional: el cómputo como herramienta para analizar datos y obtener conclusiones y utilizar la descripción de los fenómenos epigenéticos como inspiración para pensar en nuevos enfoques para la resolución de problemas. Esto último compone la parte principal de este trabajo y se discutirá en los capítulos siguientes.

Capítulo 3

Estado del arte en algoritmos genéticos

El algoritmo genético original, también referido en ocasiones como algoritmo genético simple (SGA) fue propuesto por Holland *et al.* en la década de 1970 en su libro *Adaptation in Natural and Artificial Systems*. El algoritmo es una **metaheurística** que desde su origen fue reconocida por su desempeño y su versatilidad por el rango de problemas al que se puede aplicar. SGA (Algoritmo 1) funciona creando una población de soluciones candidatas, a la cual se le aplican los diferentes operadores que conforman el algoritmo: selección, cruza y mutación, los cuales guían la exploración del algoritmo sobre el espacio de búsqueda del problema a abordar. Además de estos operadores es posible utilizar diferentes estrategias para el remplazo de la población generada por el algoritmo, especialmente técnicas elitistas que conserven las mejores soluciones encontradas en cada generación. Una de las bondades de SGA es el número reducido de parámetros que necesita para funcionar: el tamaño de población, la probabilidad de cruza y la probabilidad de mutación.

Como cualquier heurística, SGA no está exento de desventajas. Como menciona Vie [52], el algoritmo suele requerir un gran número de iteraciones para lograr alcanzar resultados deseables, además de que el desempeño del mismo depende fuertemente de la configuración inicial de sus parámetros. Este punto está relacionado con los llamados teoremas *no free lunch* de Wolpert y Macready [56], los cuales conciernen a todos los algoritmos de optimización que puedan encontrarse.

A través de los años se han hecho propuestas de modificaciones y técnicas alternativas para cada uno de los operadores de SGA, existiendo en la literatura diversos análisis de ellos, tanto en funciones de prueba ya conocidas como en problemas aplicados de diferentes áreas. En este capítulo se hará un recuento de estas propuestas.

Algoritmo 1: Algoritmo genético simple (SGA)

Datos: Tamaño de población n , dimensiones d , número de generaciones max , tasa de mutación m , tasa de cruza c

Resultado: Mejor solución encontrada

```
1 Generar población  $P$  de  $n$  individuos de  $d$  dimensiones
2 Obtener mejor solución y asignarla a  $best_{global}$ 
3  $gen = 0$ 
4 mientras  $gen < max$  hacer
5   | Aplicar mecanismo de selección
6   | Aplicar mecanismo de cruza utilizando  $c$ 
7   | Aplicar mecanismo de mutación utilizando  $m$ 
8   | Reemplazar la población actual por la nueva
9   | Obtener mejor solución  $best_i$  de la iteración
10  si  $best_i$  mejor que  $best_{global}$  entonces
11    |  $best_{global} = best_i$ 
12    |  $gen = gen + 1$ 
13 fin
14 Devolver  $best_{global}$ 
```

3.1. Estructura de la población

Aunque la mayor parte de los esfuerzos en aumentar las capacidades de GA recaen en el diseño de nuevas propuestas de operadores genéticos, la estructura y tamaño de la población juega un papel importante en el desempeño general del algoritmo. El diseño original de SGA propone una población de individuos que se cruzan sin distinción de sexo u otras formas de agrupar la población. Sin embargo, en la literatura es posible encontrar diferentes propuestas de modelado de la población en el algoritmo.

Una de las modificaciones más naturales a la estructura de la población es distinguir a los individuos por sexo, lo cual tiene un impacto directo en la dinámica de los operadores de SGA. El estudio de los efectos de modelar reproducción sexual en cómputo evolutivo se pueden rastrear hasta el trabajo de Todd y Miller [51] donde muestran los beneficios de implementar la impronta genética en SGA (expresión de genes dependiendo del sexo del progenitor que lo haya transmitido), en especial en la habilidad del algoritmo para escapar de óptimos locales al fomentarse la diversidad de la población además de otras consecuencias observadas como la aparente formación de especies diferentes y la aparición de novedad en los comportamientos de apareamiento. Estos últimos puntos

son temas de reciente interés en el área de aprendizaje profundo donde también se han aplicado técnicas evolutivas para mejorar las estructuras de las redes neuronales, área de estudio que lleva por nombre **neuroevolución** [18].

De este modo, la finalidad de crear poblaciones sexuadas va más allá de restringir de manera directa los subconjuntos de la población en la etapa de selección, como mencionan Sánchez-Velazco y Bullinaria [45]. En dicho artículo se proponen diferentes parámetros de selección dependiendo del sexo, inspirados tanto en la competitividad entre miembros del mismo sexo como la diferencia en aportes y medidas de aptitud, una idea que se retoma para las propuestas de este trabajo de tesis y que se explica con mayor detalle más adelante. Otra propuesta de diferenciación sexual es la presentada por Raghuwanshi y Kakde [38], donde se determina el sexo de los individuos de acuerdo a su capacidad de producir mejores hijos (los individuos con un rendimiento superior al promedio se clasifican como hembras). Ellos a su vez dividen la población en especies con base en la proximidad entre individuos. Zhu *et al.* [59] aprovechan la distinción sexual para aplicar tasas de mutación diferentes para ambos sexos, las cuales se actualizarán conforme a la edad de cada individuo. De manera similar a la propuesta de Sánchez-Velazco y Bullinaria [45], Ansótegui *et al.* [4] utilizan un método de selección diferente para cada sexo. Un aspecto interesante de su propuesta es que aplican su algoritmo para la calibración de parámetros de resolvers SAT.

3.2. Operadores de selección

En la etapa de selección se escogen a los individuos que participarán en la etapa de cruce dependiendo de las condiciones o características del método elegido para llevar a cabo este proceso. En algunos métodos existe la posibilidad de que ciertos individuos queden descartados dependiendo de las condiciones impuestas, en especial con relación a la aptitud de los individuos evaluados. Los métodos más utilizados son:

- **Selección por ruleta:** En este método, utilizando como analogía el juego de la ruleta, se le asigna a cada individuo un escaque de mayor o menor tamaño según su valor de aptitud, asignando una probabilidad p de ser elegido con base en esta aptitud mediante

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \quad (3.1)$$

donde i es el índice de cada individuo y f_i es el valor de aptitud del individuo.

- **Selección por torneo:** Este método es muy popular por su fácil implementación y su sencillez. En él, para cada individuo se eligen dos posibles parejas de manera aleatoria y se elige la de mayor aptitud.

- **Muestreo estocástico universal:** Es una variante del método de ruleta. Se elige un individuo como punto de partida y se escoge el siguiente según un intervalo determinado, lo que hace que cada individuo tenga oportunidad de participar en la etapa de cruza.
- **Selección por *ranking*:** Esta es otra extensión del método de ruleta. Se ordenan los individuos según su aptitud, asignando el lugar 1 al de menor aptitud y el lugar N al de mejor aptitud. Esta técnica ha demostrado solucionar el problema de convergencia rápida del método de ruleta.

En la literatura existen propuestas de modificaciones a los métodos antes mencionados; sin embargo, recientemente han surgido nuevos enfoques. Ismkhan [28] presenta un método que evalúa la similitud entre los vectores de solución de los individuos seleccionados, enfoque que a su vez se aplica en la etapa del remplazo. Haq *et al.* [23] proponen un método que llaman “selección escalonada”, el cual busca encontrar un balance entre los enfoques de explotación y exploración de los métodos de selección tradicionales.

Dentro de las propuestas recientes de métodos de selección existen algunas que dependen de una estructura de población dividida en sexos, como las mencionadas en la sección anterior. Como se mencionó anteriormente, una propuesta muy interesante, en especial para este trabajo de tesis, es la presentada por Sánchez-Velazco y Bullinaria [45]. En este artículo los autores presentan un modelo en el que se toman en cuenta diferentes factores para la selección de individuos de cada sexo; a su vez, la mutación también depende del sexo de los individuos. En el caso de las hembras, la selección se realiza de manera directa comparando el valor de la función de aptitud de cada par seleccionado (similar a la selección por torneo). Por otro lado, la función de selección para los machos toma en cuenta una serie de factores de ponderación. El método queda descrito como

$$Sel_x = \frac{w_1 f(x) + w_2 \Delta f(y) + w_3 (Edad(x))}{w_1 + w_2 + w_3} \quad (3.2)$$

donde:

- $f(x)$ es la función de aptitud del macho
- $f(y)$ es la función de aptitud de la hembra
- w_1 es el factor de aptitud competitiva
- w_2 es el factor de edad y fertilidad
- w_3 es el factor de aptitud cooperativa

El desempeño de este método de selección prueba ser significativamente superior a SGA en la resolución de problemas como TSP y descomposición de gráficas, tanto en calidad de las soluciones como en el número de generaciones para llegar a resultados deseables Vie [52].

3.3. Operadores de cruce

En la etapa de cruce cada uno de los individuos elegidos en la etapa de selección es emparejado con otro individuo para crear una solución nueva, la cual será añadida a la población nueva que terminará remplazando a la actual. La idea básica detrás de la mayoría de los métodos de cruce existentes en la literatura es elegir uno o más puntos de corte en el vector de genes de cada individuo y crear uno nuevo con la información de ambos padres ordenada de una manera diferente. En este proceso también entra en juego un valor de probabilidad de cruce, el cual ayuda a controlar la frecuencia con la que se lleva a cabo este proceso.

Los métodos de cruce tradicionales son:

- **Cruce en un punto:** Es el método más común de cruce. Se elige un gen en una posición de referencia para “cortar” el vector de genes de cada individuo. Habiendo hecho esto, se procede a mezclar los fragmentos creados para crear dos nuevas soluciones.
- **Cruce en N puntos:** Es una generalización del método anterior, eligiendo dos o más puntos de cruce para hacer los fragmentos y mezclarlos.
- **Cruce uniforme:** En este método se recorre gen por gen de ambos padres y de manera aleatoria se elige el valor del padre seleccionado en el gen indicado.

Hassanat y Alkafaween [25] proponen una serie de métodos que buscan competir con los ya existentes:

- **Corte en el peor gen:** Es una variación de la cruce en un punto. Este método identifica el gen que incrementa la función de costo, el cual se define dependiendo del problema a resolver. Al identificar el peor gen en cada padre, se elige el punto de corte en el índice del peor de los dos y se hace el intercambio.
- **Método de colisión:** Utilizando las ecuaciones de colisión entre objetos provenientes de la física, se “colisionan” los cromosomas de cada padre, tomando como velocidad la función de costo de cada uno. Así, se calcula el efecto de la colisión

para cada gen. Se consideran como buenos los genes que cambian su dirección de movimiento o que se vuelven estáticos. Para los demás, se hace un intercambio directo de los genes de ambos padres en las posiciones correspondientes.

- Cruza múltiple: Esto se refiere al uso de más de un método cruza de manera simultánea utilizando dos tipos de enfoques:
 - Seleccionar al mejor: en este enfoque se realizan todos los métodos de cruza elegidos de manera paralela, escogiendo mejores individuos resultado de cada operación
 - Selección indistinta: En cada ciclo de cruza se elige un método de cruza de manera aleatoria

Hassanat y Alkafaween [25] llegan a la conclusión de que utilizar los enfoques de múltiples métodos de cruza durante la ejecución del algoritmo es una manera efectiva de mejorar el rendimiento de los algoritmos genéticos, además de ayudar a mantener la diversidad en la población.

3.4. Operadores de mutación

El proceso de mutación, al igual que el de cruza, es un componente clave para la evolución de los individuos y para que siga existiendo un nivel de diversidad ideal a lo largo de las generaciones, con el fin de evitar la convergencia prematura del algoritmo. A su vez, esta etapa también está regulada por un valor de probabilidad, el cual es común configurar con valores pequeños. El proceso básico de mutación consiste en elegir un elemento o más del vector de genes de manera aleatoria, para elegir un valor aleatorio dentro del rango permitido que sustituya a cada elemento.

3.5. Estrategias de remplazo

En SGA el remplazo de los individuos se hace de una manera simple: la descendencia creada en cada generación sustituye a la población anterior. Sin embargo, el concepto de elitismo da oportunidad para proponer métodos que ayuden a mejorar el desempeño general del algoritmo. El elitismo se refiere a dar prioridad y conservar a las mejores soluciones de cada generación.

Un enfoque innovador es el mencionado en la sección 3.2 en la propuesta presentada por Ismkhan [28], en donde para cada nuevo individuo se busca la solución previa con mayor

similitud; en caso de haber más de una opción, se elige aquella con menor aptitud. De este modo, se compara la solución nueva con la previa y se conserva la de mejor aptitud.

3.6. Comentarios finales

La historia de GA muestra que el algoritmo está lejos de ser una técnica olvidada o fuera del interés de la comunidad científica. Tanto por la intuición que nos brinda al entender los fenómenos biológicos que sirven como inspiración como por los resultados que se observan al aplicarlo a diferentes problemas, GA y sus diferentes versiones siguen siendo un punto de partida para el estudio y desarrollo de técnicas de cómputo inspiradas en la naturaleza. Sin embargo, es posible notar cómo en los diferentes operadores del algoritmo las modificaciones cada vez distan más de inspirarse en fenómenos biológicos y terminan siendo adaptaciones de métodos ya existentes, principalmente a la hora de utilizar el algoritmo para la resolución de problemas específicos. En este sentido se considera que optar por desarrollar métodos más cercanos a las dinámicas de interacción entre individuos de diferentes sexos es un camino viable para la innovación en algoritmos evolutivos puesto que en diferentes ocasiones este enfoque ha mostrado un desempeño favorable. A expensas de perder una de las cualidades de GA, la sencillez de su implementación, la ganancia en las posibilidades de integrar nuevos descubrimientos de genética y biología evolutiva es mucho mayor, proporcionando la oportunidad de experimentar con la modelación de dichos fenómenos, como es el caso de la epigenética, cuyas contribuciones en el campo del cómputo evolutivo se exploran en el siguiente capítulo.

Capítulo 4

Cómputo y epigenética

All biology is computational biology.

Florian Markowetz

4.1. Introducción

Al igual que con el estudio de fenómenos genéticos, el cómputo ha sido una herramienta muy útil al realizar investigación de los efectos epigenéticos en los organismos. En el caso de la bioinformática se ha acuñado el término “epigenética computacional” para referirse a la intersección y colaboración entre estas dos áreas.

Angarica y Sol [3], Ruskin y Barat [44] y He y Song [26] mencionan técnicas de procesamiento computacional de información epigenética, además de enumerar algunas de las bases de datos disponibles con información de metilación del epigenoma y modificaciones de histonas. Yen [57] presenta en su tesis doctoral algunos casos de estudio utilizando algunas de estas bases de datos y herramientas computacionales existentes. Del mismo modo, se mencionan diferentes proyectos internacionales que buscan crear un mapa epigenético de los humanos, de manera similar al proyecto que identificó el genoma humano a finales del siglo XX. Entre estas iniciativas se encuentran el *NIH Roadmap Epigenomics Mapping Consortium*, el *International Human Epigenome Consortium*, la *Encyclopedia of DNA Elements* (ENCODE), entre otras, que han puesto a disposición de la comunidad científica diferentes herramientas y fuentes de información

4. CÓMPUTO Y EPIGENÉTICA

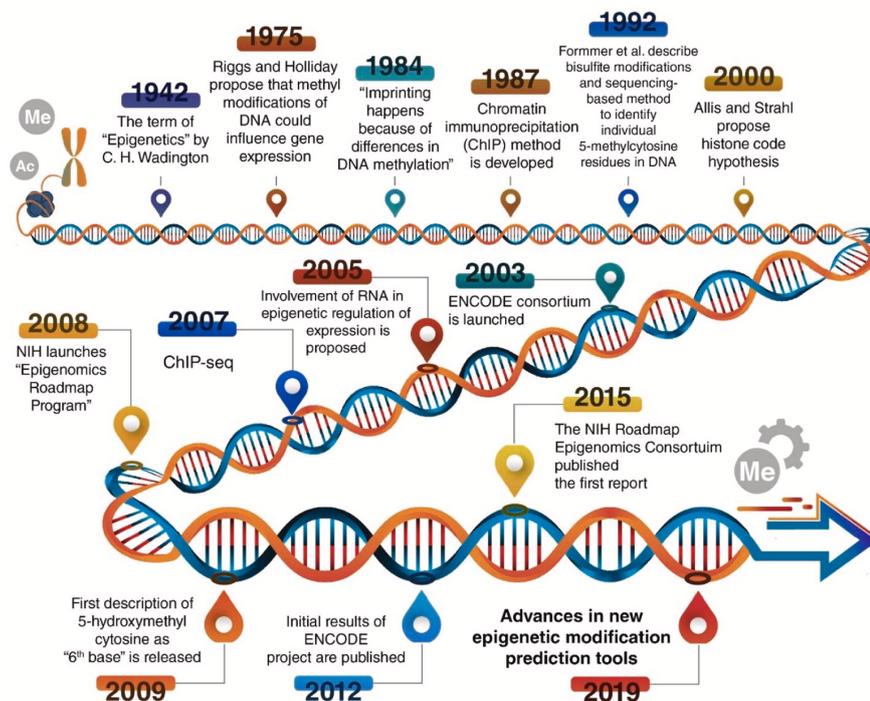


Figura 4.1: Línea del tiempo del avance en descubrimientos sobre epigenética

relacionadas con el epigenoma humano. Algunos de estos repositorios destacados son el *Gene Expression Omnibus*¹, *WashU Epigenome Browser*² y *Roadmap Epigenomics Visualization Hub*, conocido también como *VizHub*³. En la figura 4.1 (tomada de Chenarani *et al.* [13]) se muestra una línea del tiempo de los descubrimientos hechos en epigenética desde las primeras ideas de Waddington hasta la creación de los proyectos antes mencionados.

Ruskin y Barat [44] hacen hincapié en que “el rol de la epigenética computacional en el desarrollo de teorías, modelos y métodos requeridos para la comprensión de datos médicos y biológicos no puede ser desestimado”. La mayoría de la investigación hecha con datos epigenéticos tiene el fin de encontrar patrones relacionados a diferentes enfermedades, así como a descubrir medicamentos para tratar dichas afecciones. Muchos de estos estudios están enfocados en descubrir patrones para diferentes tipos de cáncer, VIH, enfermedades cardiovasculares, entre otras afecciones de difícil tratamiento. Brasil *et al.* [6] De este modo, no se puede más que reafirmar lo mencionado por Ruskin y Barat [44]: la relación entre cómputo y epigenética cobra cada vez más importancia

¹<https://www.ncbi.nlm.nih.gov/geo/>

²<http://epigenomegateway.wustl.edu/>

³<https://vizhub.wustl.edu/>

y esta interacción refuerza a su vez la importancia de la bioinformática, un área del conocimiento que ha tenido un crecimiento enorme en los últimos años.

4.2. Propuestas existentes de algoritmos inspirados en epigenética

En el ámbito del cómputo evolutivo, el desarrollo de propuestas inspiradas en epigenética es una iniciativa reciente, pudiendo rastrear la aparición de algoritmos con inspiración epigenética en la década de los 2000. Como es posible apreciar en la exploración del estado del arte de los algoritmos genéticos hecha en los capítulos anteriores, es evidente que las múltiples propuestas de estrategias existentes en la literatura siguen dependiendo de las diferentes combinaciones que se puedan realizar entre los genomas de los individuos. Desde este punto de vista, la posibilidad de empezar a integrar las ideas de la epigenética en algoritmos genéticos representa un panorama prometedor, pues es posible que nos abra las puertas hacia nuevas formas de entender los algoritmos genéticos y su funcionamiento.

Aunque hasta ahora son limitadas, en la literatura es posible encontrar algunas propuestas de integración de ideas de epigenética en algoritmos evolutivos. En 2003, Tanev y Yuta [49] presentan una variante de GP a la que llaman Programación Epigenética (EP). La Cava *et al.* [31] presentan una propuesta de “búsqueda local epigenética” para GP, la cual agrega un vector de activación/inactivación para cada símbolo de los programas generados, expresándose solamente aquellos símbolos que aparezcan activados. A su vez, en Ricalde [41] se presenta un mecanismo epigenético para GP aplicado a control de semáforos.

Tal vez el primer intento de hacer una variante de SGA directamente inspirada en epigenética sea EGA (algoritmo 2), presentado por Periyasamy *et al.* [36]. En esta propuesta los autores presentan un enfoque mixto utilizando ideas de algoritmos genéticos y de inteligencia colectiva, enfocándose en mecanismos epigenéticos intra-generacionales utilizados por biomoléculas, siendo el objetivo optimizar la organización interna de una célula para mantener la **homeostasis**, la condición de estabilidad dentro de los procesos internos de la célula.

Recientemente Chrominski *et al.* [14] propusieron un nuevo método de cruce inspirado en el proceso de metilación de citosina. Al momento de la cruce (descrita en el algoritmo 3) se bloquean fragmentos aleatorios del cromosoma, los cuales no se toman en cuenta

4. CÓMPUTO Y EPIGENÉTICA

Algoritmo 2: EGA

- 1 Generar población heterogénea con interacciones y tiempo de vida
 - 2 Evaluar cada individuo
 - 3 **mientras** *no se cumpla condición de paro* **hacer**
 - 4 | Seleccionar organismos supervivientes y eliminar a los muertos
 - 5 | Añadir individuos por autocatálisis
 - 6 | Evaluar cada individuo
 - 7 **fin**
-

al llevar a cabo el método de cruce elegido. El proceso de bloqueo se lleva a cabo solo en el individuo de menor aptitud de los que participan en el proceso de cruce. De este modo, los nuevos individuos tendrán una herencia genética conformada en mayor parte por los genes del padre con mayor aptitud.

Algoritmo 3: Cruza con metilación

- Entrada:** Población, probabilidad P_m , función de aptitud F_p
- 1 Seleccionar un par de individuos según P_m
 - 2 **para cada** *par seleccionado* **hacer**
 - 3 | **si** $F_{p_1} > F_{p_2}$ **entonces**
 - 4 | | Bloquear parte del genotipo del individuo 2
 - 5 | **en otro caso**
 - 6 | | Bloquear parte del genotipo del individuo 1
 - 7 | **fin**
 - 8 | Realizar cruce sin tomar en cuenta las partes bloqueadas
 - 9 **fin**
-

Por su parte, Álvarez Camacho [2] presenta ReGen GA (GA con genes regulados), el cual es similar en estructura a epiGA (descrito más adelante). La contribución de dicho trabajo es la asignación de marcadores epigenéticos a alelos. Los marcadores, implementados como una cadena binaria de 8 bits, representan operaciones que se aplicarán directamente en el alelo asignado. A su vez, el algoritmo evalúa a los individuos generados en la etapa de cruce con base en los fenotipos generados en la fase de marcado.

Algoritmo 4: ReGen GA

- 1 Iniciar población con soluciones aleatorias
 - 2 Evaluar soluciones
 - 3 **repetir**
 - 4 Seleccionar padres
 - 5 Recombinar padres seleccionados
 - 6 Mutar progenie
 - 7 **si** *es periodo de marcado* **entonces**
 - 8 Aplicar marcado en la progenie
 - 9 **fin**
 - 10 Obtener fenotipos decodificando la función de crecimiento de la progenie
 - 11 Evaluar fenotipos
 - 12 Seleccionar individuos para la siguiente generación
 - 13 **hasta que** *se cumpla la condición de término*
-

Yuen *et al.* [58] hacen un análisis de las áreas de oportunidad para añadir mecanismos epigenéticos en algoritmos evolutivos. En dicho artículo los autores mencionan algunas de las propuestas de algoritmos epigenéticos nombradas anteriormente y de acuerdo a análisis que hacen de ellas, hacen propuestas a considerar para la implementación de características inspiradas en epigenética dentro de algoritmos evolutivos y que consideran que hacen falta en los algoritmos inspirados en epigenética existentes en la literatura. Los autores destacan que un punto crucial es realizar una implementación robusta del epigenotipo dentro de los algoritmos. Sin una buena implementación de los marcadores epigenéticos no es posible explotar de manera adecuada los diferentes mecanismos epigenéticos disponibles para ser implementados.

4.3. Propuesta de algoritmo epigenético de Stolfi y Alba

Stolfi y Alba proponen epiGA (algoritmo 5) en Stolfi y Alba [48] como un marco de referencia para desarrollar algoritmos inspirados en fenómenos epigenéticos. Su propuesta, a diferencia de la presentada en Periyasamy *et al.* [36], se aleja de un escenario que toma en cuenta una sola célula para utilizar individuos que constan de una serie de células diferentes, las cuales son las que entrarán en juego para la etapa de reproducción. Otro punto interesante es que descarta el operador de mutación para sustituirlo por los mecanismos epigenéticos que el usuario desee utilizar. Algo a destacar de epiGA es que se puede considerar estructuralmente idéntico a SGA, incluso compartiendo varios de los parámetros de inicio (tabla 4.1). Sin embargo, la diferencia yace en la mecánica y en las estrategias utilizadas en los operadores de cruce y en la inclusión de los mecanismos epigenéticos antes de la etapa de remplazo. Los parámetros de entrada nuevos se explicarán en las secciones siguientes.

4. CÓMPUTO Y EPIGENÉTICA

Parámetro	Descripción
N_i	No. de individuos
N_c	No. de células por individuo
P_e	Probabilidad epigenética
P_n	Probabilidad nucleosómica
M	Mecanismos epigenéticos
E	Probabilidad de entorno
R	Radio de nucleosoma

Tabla 4.1: Parámetros de entrada de epiGA

Algoritmo 5: epiGA

Datos: $N_i, N_c, P_e, P_n, R, M, E$

Resultado: Mejor solución encontrada

- 1 Generar población P de tamaño N_i individuos con N_c células
 - 2 Generar población auxiliar vacía Q
 - 3 $t = 0$
 - 4 **mientras** *no se cumpla la condición de paro* **hacer**
 - 5 Seleccionar individuos de P y asignarlos a Q
 - 6 Generar nucleosomas para individuos de Q usando P_n y R
 - 7 Aplicar reproducción nucleosómica en Q
 - 8 Aplicar mecanismos epigenéticos en Q usando P_e, M y E
 - 9 Sustituir P por Q mediante operador de remplazo
 - 10 $t = t + 1$
 - 11 **fin**
 - 12 Devolver mejor solución encontrada
-

4.3.1. Estructura de la población

Como se mencionó anteriormente, en epiGA los individuos consisten en conjuntos de células, las cuales representan una solución candidata del problema a resolver. En la figura 4.2 se observa de manera gráfica la estructura de los individuos, en donde f representa el cromosoma del padre, m el cromosoma de la madre, x el vector de solución y n el vector nucleosoma. En dicha figura se utiliza una codificación binaria como ejemplo. Sin embargo, epiGA también funciona con codificación real, la cual se usa para los fines de este trabajo de tesis.

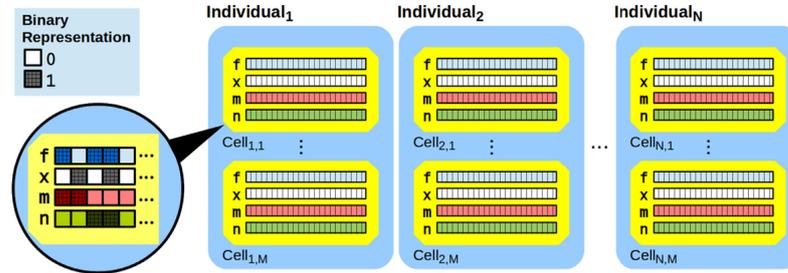


Figura 4.2: Estructura de los individuos en epiGA. Tomada de [48].

4.3.2. Reproducción basada en nucleosomas

En lugar de usar un método de cruce convencional, epiGA utiliza un método de “reproducción nucleosómica” (algoritmo 6), el cual tiene como producto dos células nuevas, en lugar de dos individuos. Este método usa el vector nucleosoma presente en la mejor célula de los individuos ind_1 e ind_2 elegidos para la cruce. Se realiza una operación lógica OR entre el nucleosoma n_1 y el nucleosoma n_2 de cada individuo respectivamente para generar un nuevo vector que será asignado a las nuevas células generadas. Basándose en este nuevo vector, se hará el intercambio de genes entre las células seleccionadas utilizando solo aquellos que correspondan a una posición en la que el vector nucleosoma tenga el valor 0 (representando una región no colapsada del genoma). Las dos células c_1 y c_2 que resultan como producto del proceso reemplazarán a las peores células de ind_1 e ind_2 respectivamente.

Algoritmo 6: Reproducción nucleosómica

Datos: P **Resultado:** Población con individuos actualizados

```
1 Iniciar población auxiliar  $P' = \emptyset$ 
2 para cada  $i_1, i_2 \in P$  hacer
3   Obtener células  $c_1, c_2$  de  $i_1, i_2$  respectivamente
4   Obtener soluciones  $x_1, x_2$  de  $c_1, c_2$  respectivamente
5   Obtener nucleosomas  $n_1, n_2$  de  $c_1, c_2$  respectivamente
6    $N = n_1 \vee n_2$  // nuevo nucleosoma
7   mientras  $j < |N|$  hacer
8     si  $N(j) = 1$  entonces
9       // las zonas no colapsadas no cambian
10       $x'_1(j) = x_1(j)$ 
11       $x'_2(j) = x_2(j)$ 
12     en otro caso
13       // las zonas colapsadas cambian
14       $x'_1(j) = x_2(j)$ 
15       $x'_2(j) = x_1(j)$ 
16     fin
17      $j = j + 1$ 
18   fin
19   Crear  $c'_1$  usando  $x'_1$  y  $N$ 
20   Crear  $c'_2$  usando  $x'_2$  y  $N$ 
21   Reemplazar la peor célula de  $i_1$  con  $c'_1$ 
22   Reemplazar la peor célula de  $i_2$  con  $c'_2$ 
23   Añadir individuos modificados a  $P'$ 
24 fin
25 Devolver  $P'$ 
```

Es posible observar que este proceso representa un método elitista interno para cada individuo. De acuerdo a Stolfi y Alba [48], al solo reemplazar la peor célula se mantiene la diversidad de soluciones dentro de cada miembro de la población. Además, las posibles configuraciones de la máscara nucleosómica hacen que el proceso de cruce, de manera ideal, sea distinto para cada par de individuos, aumentando la capacidad de exploración del algoritmo.

4.3.3. Silenciamiento de genes

La última etapa de epiGA antes del remplazo es la aplicación de mecanismos epigenéticos. En la propuesta original solo se implementa el mecanismo de silenciamiento de genes aunque, como se describe en el artículo, existe toda una gama de mecanismos que pueden ser tomados como inspiración para esta etapa del algoritmo, habiendo la posibilidad de aplicar más de un mecanismo dentro de esta etapa. Dentro de este proceso se usa los parámetros de probabilidad epigenética (P_e) y de entorno (E , *environment*) el cual asigna una probabilidad de silenciamiento para cada gen del vector genoma.

Algoritmo 7: Silenciamiento de genes

Datos: c, P_e, E

- 1 Obtener solución x de c
- 2 Obtener nucleosoma n de c
- 3 **mientras** $j < |x|$ **hacer**
- 4 **si** $n(j) = 1$ **entonces**
- 5 **si** $rand < P_e$ **entonces**
- 6 $x(j) = rand < E(j)$
- 7 **fin**
- 8 **fin**
- 9 $j = j + 1$
- 10 **fin**

4.4. Comentarios finales

En la revisión de la literatura hecha en este capítulo sobre la integración de las ideas de epigenética en el cómputo evolutivo, queda claro que esta línea de estudio es aún muy joven y su potencial significatividad como un enfoque innovador en el desarrollo de algoritmos genéticos es grande. Si bien algunos de los métodos revisados se asemejan a los operadores más tradicionales, se considera que los métodos propuestos en epiGA representan una alternativa novedosa debido a las diferentes capas de variación de la capa nucleosómica y de las variables relacionadas con el entorno. epiGA ofrece la flexibilidad de elegir el o los mecanismos epigenéticos que se aplicarán a las células de los individuos, además de la frecuencia con la que se realizará este proceso. Este trabajo de tesis se enfoca en la posibilidad de elegir de manera libre los métodos de selección y remplazo. Tomando en cuenta que epiGA integra ideas de influencia del entorno en la expresión genética se proponen métodos inspirados en la influencia de los cuidados parentales sobre la aptitud de los individuos de diversas especies. Esta propuesta se revisará a detalle en el capítulo siguiente.

Capítulo 5

Propuesta de operadores epigenéticos

En este capítulo se describirá el diseño de los operadores epigenéticos que conforman la propuesta presentada en este trabajo.

Tomando en cuenta lo mencionado por Yuen *et al.* [58], la implementación hecha en epiGA de la transmisión y modificación de marcadores epigenéticos no dista mucho de los métodos de mutación tradicionales ya que se disparan de manera aleatoria con probabilidades fijas. Asimismo, los autores reconocen a los efectos parentales como uno de los fenómenos medioambientales mayormente ligados a los mecanismos epigenéticos. De este modo, la propuesta de este trabajo consiste en tomar al fenómeno de DA como inspiración para el diseño de estos operadores, debido a que es un fenómeno observado con frecuencia y bien documentado, del cuál se han hecho modelos matemáticos en años recientes, como se mencionó en el capítulo 2. Además, DA no solo se centra en los cuidados parentales, sino que también trata las consecuencias de la asignación de rasgos atractivos en la dinámica de selección de parejas por parte de los individuos de ambos sexos de una población y en la probabilidad de supervivencia de los individuos, siendo esto análogo a las etapas de selección y remplazo de la estructura de los algoritmos genéticos.

Debido a lo anterior, se considera que utilizar el fenómeno de DA como referencia para la creación de nuevos operadores para algoritmos inspirados en epigenética representa un paso en la dirección marcada por Yuen *et al.* [58]. Utilizando epiGA como base, se integrarán los operadores propuestos dentro de su estructura, ya que son las etapas cuya implementación que quedan a elección del usuario en la propuesta original de Stolfi y Alba [48]. Sin embargo, serán necesarias otras modificaciones a epiGA, la cuales se describirán a continuación.

5.1. Características de la población

La población se dividirá en individuos machos y hembras. Aunque ambos grupos de población comparten la mayoría de sus atributos, existen algunas diferencias. En el caso de los machos, el atributo de calidad se asigna según lo definido en el apéndice de Haaland *et al.* [22]: clases numeradas del 1 al 3 donde la más alta es considerada la mejor. La clase se asigna de acuerdo a una distribución de probabilidad $P = \{.2, .5, .3\}$. Para las hembras, la calidad se calcula de acuerdo a la fórmula de Sánchez-Velazco y Bullinaria [45]. Se consideró que el número de parámetros original de dicha fórmula es suficiente para los fines del diseño de la propuesta de esta tesis, ya que se ajusta a los parámetros utilizados en DA. Por lo tanto, se utilizan $w_1 = 0.5$, $w_2 = 0.3$ y $w_3 = 0.2$ para los parámetros de la aptitud de la mejor célula del macho, la aptitud de la mejor célula de la hembra y el factor de edad, respectivamente. De este modo, se le da mayor importancia al aporte del individuo macho.

5.2. Selección

El objetivo de la selección por DA es crear una estrategia que tome en cuenta la inversión parental de los individuos sobre su descendencia. Como ya se ha mencionado, se tomarán como base las ideas de Sánchez-Velazco y Bullinaria [45] donde se usan diferentes funciones de aptitud para individuos de diferente sexo, además de que se harán modificaciones en la estructura de epiGA en el caso del vector de entorno. Será necesario agregar atributos a los individuos además de los que ya se consideran en epiGA. En el caso de los machos la función de aptitud será la representación en el algoritmo de los atributos atractivos mencionados en los estudios de Burley, usando las tres clases de calidad del modelo de Haaland *et al.* [22]. Por su parte, la aptitud de las hembras se calculará mediante una función basada en el modelo de Haaland (ecuación 2.1). En ambos casos, en la comparación entre pares se tomará en cuenta tanto la función de aptitud como el valor de la función objetivo.

Algoritmo 8: Selección inspirada en DA

Datos: c, P_e, E

- 1 Obtener solución x de c
- 2 Obtener nucleosoma n de c
- 3 **mientras** $j < |x|$ **hacer**
- 4 **si** $n(j) = 1$ **entonces**
- 5 **si** $rand < P_e$ **entonces**
- 6 $x(j) = rand < E(j)$
- 7 **fin**
- 8 **fin**
- 9 $j = j + 1$
- 10 **fin**

5.3. Ajuste en reproducción nucleosómica

Debido a la utilización de una población dividida en sexos, se realizaron ajustes mínimos al proceso de reproducción nucleosómica. El resultado de la cruce son dos individuos de sexos diferentes, de manera que se crean dos poblaciones de hijos del mismo tamaño de la generación anterior. Sin embargo, el cambio más importante en esta etapa del algoritmo es la reducción del atributo de energía en el individuo hembra. El cálculo se realiza en relación a la aptitud del hijo macho con respecto al padre. Según el porcentaje de mejora, se calcula la cantidad de energía que pierde la hembra, siendo esta pérdida directamente proporcional a la mejora.

5.4. Remplazo

El remplazo (algoritmo 9) se hará de diferente manera de acuerdo al sexo del individuo. Existen consideraciones en común: para ambos sexos se tomará en cuenta la edad, para la cual se tiene como límite 10 generaciones en ambos casos. Además, se toma en cuenta el parámetro de probabilidad de supervivencia (α) del modelo de Haaland y se utiliza el mismo valor propuesto (0.8). La diferencia central es en el uso del atributo de energía como determinante en la supervivencia de las hembras. Como se mencionó anteriormente, las hembras pierden puntos de energía de acuerdo a la mejora relativa entre los hijos y los padres (machos). De este modo, es posible que la hembra se quede sin energía antes de llegar al límite de tiempo de vida (10 generaciones).

Este enfoque de método de remplazo es similar al llamado “vuelo inteligente” propuesto

por Mezura-Montes y Cetina-Domínguez [35] en la modificación al algoritmo de colonia de abejas artificiales (M-ABC) presentada en su artículo. La idea detrás del vuelo inteligente es reemplazar las soluciones que no han sido mejoradas en un número determinado de generaciones utilizando información de la mejor solución de la generación actual y otra solución elegida de manera aleatoria. En el caso del remplazo inspirado en DA, se utiliza el criterio de energía además del límite de generaciones de supervivencia para un individuo, como se describió anteriormente. Sin embargo, en epiDA se evita la utilización de información del mejor individuo de la población, limitándose a utilizar la información de las células de los padres elegidos para cada iteración del proceso de cruce. De este modo se busca mantener diversidad dentro de la población.

Algoritmo 9: Reemplazo inspirado en DA**Datos:** Población anterior P_0 , hijos H , sexo

```

1 si sexo = "f" entonces
2   pob ← PoblacionFem
3   para  $0 \leq i < |P_0|$  hacer
4     madre ←  $P_0[i]$ 
5     hija ←  $H[i]$ 
6     edad ← madre.edad
7     energia ← madre.energia
8     si edad < 10 entonces
9       si energia > 0 entonces
10        si rand < 0.8 entonces
11          si edad < 10 entonces
12            | Añadir madre a pop
13          en otro caso Muere por menor aptitud
14            | Añadir hija a pop
15          fin
16        en otro caso Muere por mala suerte
17          | Añadir hija a pop
18        fin
19      en otro caso Muere por debilidad
20        | Añadir hija a pop
21      fin
22    en otro caso Muere de vejez
23      | Añadir hija a pop
24    fin
25  fin
26 si no, si sexo = "m" entonces
27   pob ← PoblacionMasc
28   para  $0 \leq i < |P_0|$  hacer
29     padre ←  $P_0[i]$ 
30     hijo ←  $H[i]$ 
31     edad ← padre.edad
32     si edad < 10 entonces
33       si rand < 0.8 entonces
34         si edad < 10 entonces
35           | Añadir padre a pop
36         en otro caso Muere por menor aptitud
37           | Añadir hijo a pop
38         fin
39       en otro caso Muere por mala suerte
40         | Añadir hijo a pop
41       fin
42     en otro caso Muere de vejez
43       | Añadir hijo a pop
44     fin
45  fin
46 fin
47 Devolver pop

```

5.5. Comentarios finales

Las modificaciones mostradas en este capítulo toman en cuenta las ideas de DA para integrarlas en el algoritmo epiGA. De este modo, a partir de ahora llamaremos epiDA a la propuesta de esta tesis. Al unir estas ideas dentro del marco de referencia de epiGA se busca encontrar posibles diferencias en el desempeño de ambos enfoques. Se espera que el cambio en la estructura de población, así como las mecánicas de selección y remplazo diferenciadas por sexo tengan un impacto significativo en el funcionamiento de epiDA. Además, es posible que pueda apreciarse el efecto de “guía de la evolución” por parte de las hembras como se ha comentado en Burley [11], pues son éstas las que, al ser el grupo de población con mayor papel en el cuidado parental, tienen el papel principal en la dinámica de selección. De este modo, las hembras son capaces de regular la herencia de rasgos ornamentales además del número de individuos de cada sexo para cada generación. Entonces, uno de los objetivos de epiDA es poder apreciar un ritmo diferente de evolución en ambos sexos.

En un espíritu similar al de epiGA y las otras propuestas de algoritmos epigenéticos, se buscó que la implementación de las modificaciones fuera estructuralmente similar a otros métodos existentes en la literatura. Sin embargo, se considera que la función de aptitud y los parámetros tomados de DA así como las modificaciones pertinentes a epiGA conforman un caso de estudio interesante tomando como base un fenómeno de consecuencias en la evolución no utilizado hasta ahora en el diseño de algoritmos evolutivos.

Capítulo 6

Experimentación y resultados

6.1. Características del entorno de evaluación

Todos los experimentos fueron realizados en una Macbook Pro 13 (2017), con procesador Intel Core i5 de doble núcleo a 2.3 GHz (Turbo Boost de hasta 3.6 GHz) y 64 MB de eDRAM y 8 GB de memoria LPDDR3 integrada a 2,133 MHz. El sistema operativo instalado fue macOS 12.2.1 (Monterey).

El lenguaje de programación utilizado para la implementación fue Python en su versión 3.8.7.

6.2. Problemas de prueba

Se eligió el conjunto de problemas del CEC 2019 para realizar la evaluación de desempeño del algoritmo, apegándose lo más posible a los requerimientos de dicho concurso. La comparación de resultados se hará entre SGA, epiGA y epiDA con las modificaciones propuestas. Se utilizó la interfaz en Python para ejecutar el código provisto en el concurso de los problemas de prueba (implementado en C++) disponible en <https://github.com/dmolina/cec2019comp100digit>. En la tabla 6.1 se detalla la lista de problemas y sus características. El concurso consistió en premiar al algoritmo que encontrara la solución óptima para todos los problemas con la precisión impuesta (1.000000000) sin límite de tiempo ni llamadas a la función objetivo. El algoritmo ganador, jDE100, es una propuesta de Brest *et al.* [7] en la que implementan una adaptación

6. EXPERIMENTACIÓN Y RESULTADOS

especial de jDE Brest *et al.* [8] para trabajar específicamente con los problemas de este concurso. La idea original de jDE es una modificación auto-adaptativa de evolución diferencial, implementando ajustes como el uso de dos poblaciones de diferente tamaño, las cuales refuerzan las cualidades de la propuesta original de jDE.

Problema	Valor óptimo	Dimensiones	Espacio de búsqueda
P1 (Storn's Chebyshev Polynomial Fitting Problem)	1	9	[-8192, 8192]
P2 (Inverse Hilbert Matrix Problem)	1	16	[-16384, 16384]
P3 (Lennard-Jones Minimum Energy Cluster)	1	18	[-4, 4]
P4 (Rastrigin's Function)	1	10	[-100, 100]
P5 (Griewangk's Function)	1	10	[-100, 100]
P6 (Weierstrass Function)	1	10	[-100, 100]
P7 (Modified Schwefel's Function)	1	10	[-100, 100]
P8 (Expanded Schaffer's F6 Function)	1	10	[-100, 100]
P9 (Happy Cat Function)	1	10	[-100, 100]
P10 (Ackley Function)	1	10	[-100, 100]

Tabla 6.1: Características de los problemas de prueba

6.3. Calibración de parámetros y métodos de comparación de resultados

Con el propósito de hacer pruebas de manera uniforme, se utilizará el paquete `irace` López-Ibáñez *et al.* [33] para calibrar los parámetros de los algoritmos. Se generará una configuración para las pruebas preliminares y otra diferente para la fase de experimentación. `irace` genera un conjunto de configuraciones de parámetros de acuerdo a las condiciones previstas y al número de parámetros a ajustar.

En cuanto a la metodología de comparación de los resultados obtenidos se utilizarán los métodos Friedman, Friedman alineado y Quade para la comparación de resultados entre pares de algoritmos y la prueba *post hoc* de Bergmann-Hommel para ajustar la comparación a la totalidad de los algoritmos. Estas pruebas están descritas en Derrac *et al.* [16] y se utilizará la biblioteca `scmpamp` Calvo y Santafe [12] disponible en R, la cual ofrece una implementación de dichas pruebas. A su vez, se elige esta batería de pruebas ya que se compararán los resultados de más de dos algoritmos a la vez, siendo el método de Wilcoxon el recomendado para comparar solo un par de algoritmos.

Los métodos de Friedman, Friedman alineado y Quade pertenecen a un tipo de pruebas

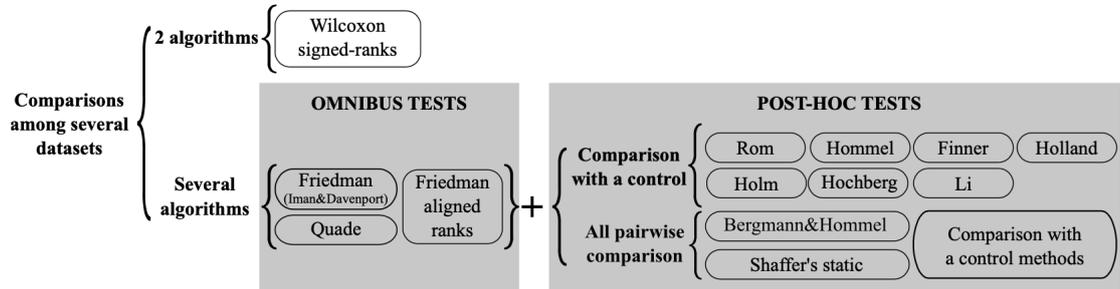


Figura 6.1: Opciones de métodos de comparación de desempeño para algoritmos de optimización. Tomada de [12].

llamadas “pruebas omnibus”, lo que significa que son capaces de rechazar hipótesis nulas para un grupo numeroso de variables. Cada prueba rechazará la hipótesis nula en el caso de que al menos un algoritmo presente una diferencia significativa de desempeño con algún otro de los evaluados. En la documentación de `scmpamp` se recomienda utilizar estas pruebas cuando el grupo de algoritmos a evaluar es menor a 5. Por su parte, la prueba de Bergmann-Hommel toma la información generada por la prueba omnibus realiza comparaciones entre todos los pares posibles de algoritmos para ajustar los valores obtenidos. En Calvo y Santafe [12] se menciona que esta prueba se considera la más poderosa para el enfoque de pruebas entre pares, sin embargo no se recomienda para grupos de más de 9 algoritmos debido a su complejidad computacional. En la figura 6.1 (tomada de Calvo y Santafe [12]) se muestran las diferentes opciones de metodologías para la comparación de desempeño entre algoritmos.

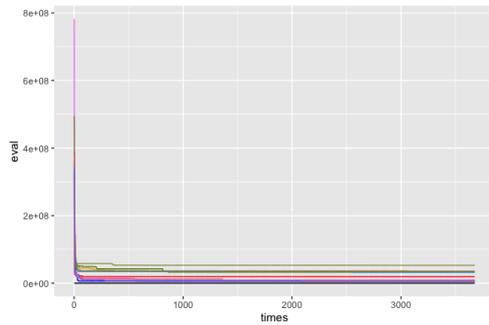
6.4. Pruebas preliminares

Como preámbulo al análisis del desempeño del algoritmo con las modificaciones propuestas, se realiza un análisis de los resultados obtenidos por SGA y epiGA en su presentación original. Para estas pruebas se realizaron 10 ejecuciones para cada problema, utilizando un límite de $1e6$ evaluaciones de la función objetivo.

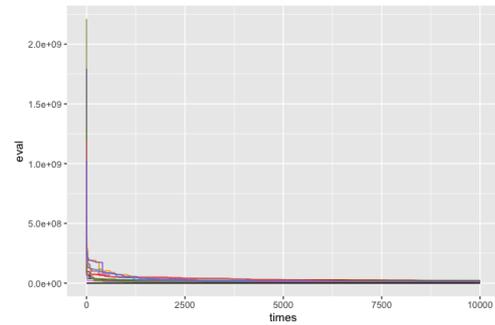
No. individuos	No. células	Prob. metilación	Prob. entorno
68	4	0.393444349	0.146421783

Tabla 6.2: Configuración de parámetros utilizada en las pruebas preliminares en epiGA

6. EXPERIMENTACIÓN Y RESULTADOS

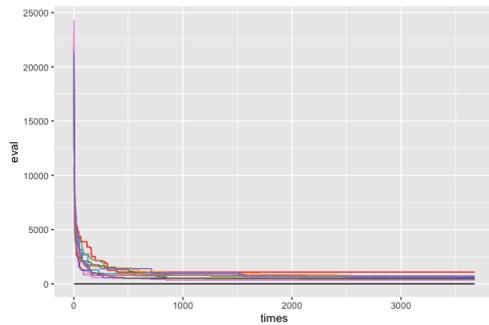


(a) epiGA

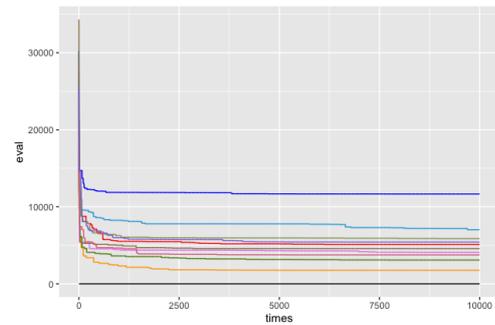


(b) SGA

Figura 6.2: Resultados preliminares para el problema 1



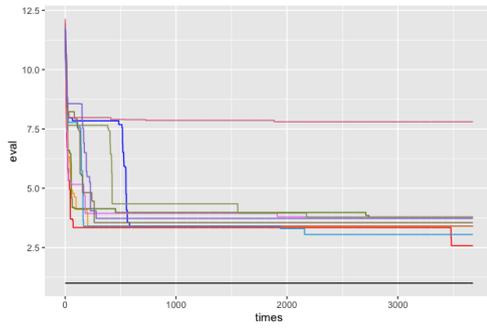
(a) epiGA



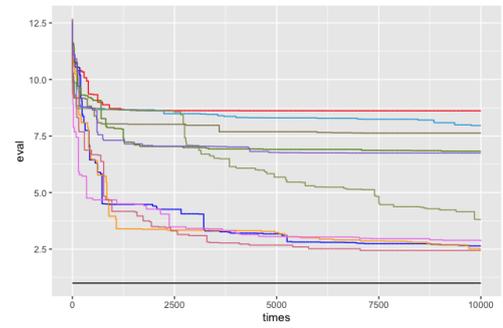
(b) SGA

Figura 6.3: Resultados preliminares para el problema 2

Aunque al ver las gráficas de los resultados preliminares (figuras 6.2 a 6.11) por primera vez es fácil pensar que el desempeño de ambos algoritmos es muy similar, hay algunos detalles que es necesario hacer notar. En primer lugar, el número de generaciones transcurridas difiere entre ambos algoritmos. En el caso de epiGA es posible observar que el número de generaciones se aproxima a alrededor de 4000, mientras que para SGA el número es más cercano a las 10,000 generaciones. Otro detalle notorio en aquellos problemas en donde es posible visualizar mejor la evolución de la optimización, epiGA muestra un comportamiento en el que la gráfica tiene una forma más escalonada, habiendo diferencias más notables en la mejora de las soluciones.

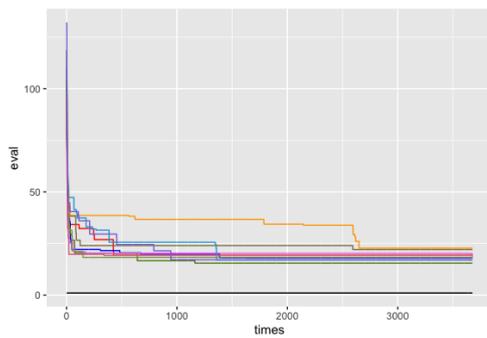


(a) epiGA

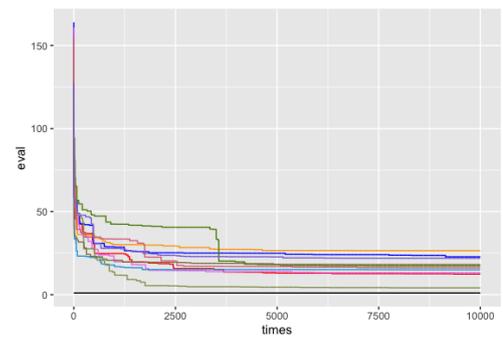


(b) SGA

Figura 6.4: Resultados preliminares para el problema 3

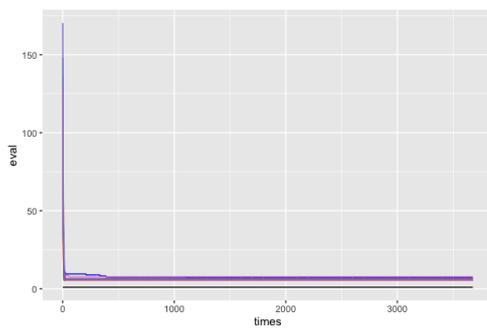


(a) epiGA

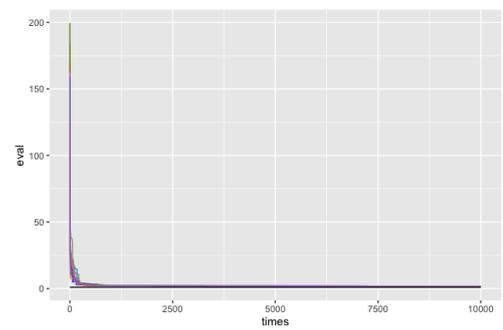


(b) SGA

Figura 6.5: Resultados preliminares para el problema 4



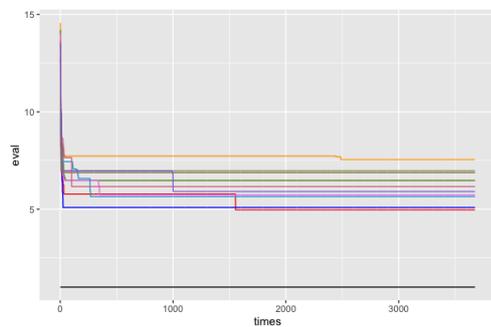
(a) epiGA



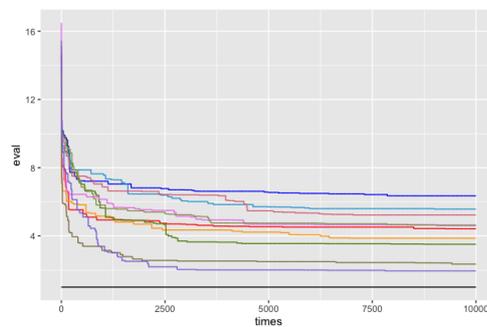
(b) SGA

Figura 6.6: Resultados preliminares para el problema 5

6. EXPERIMENTACIÓN Y RESULTADOS

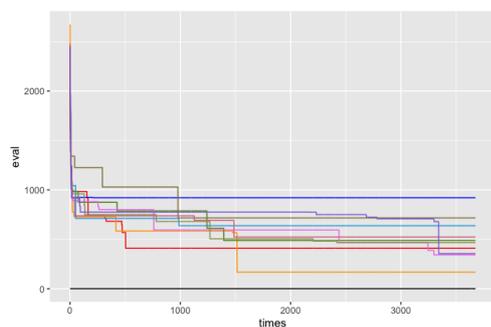


(a) epiGA

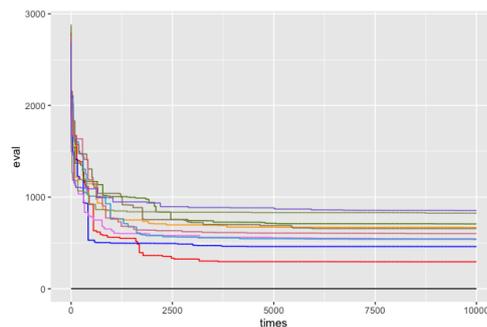


(b) SGA

Figura 6.7: Resultados preliminares para el problema 6

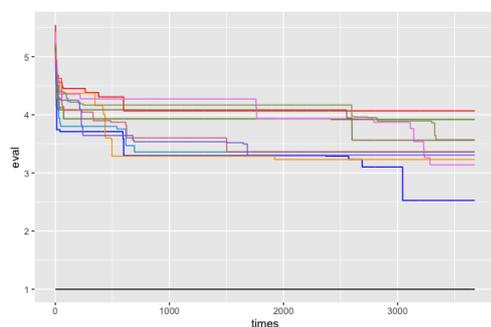


(a) epiGA

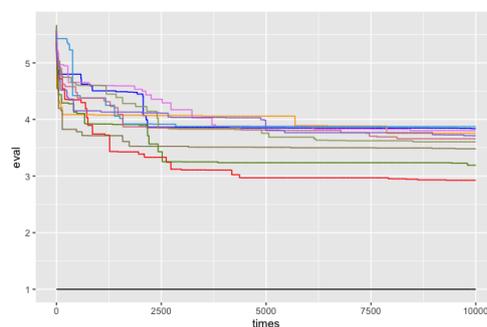


(b) SGA

Figura 6.8: Resultados preliminares para el problema 7

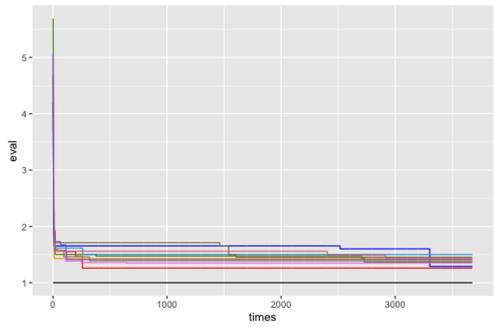


(a) epiGA

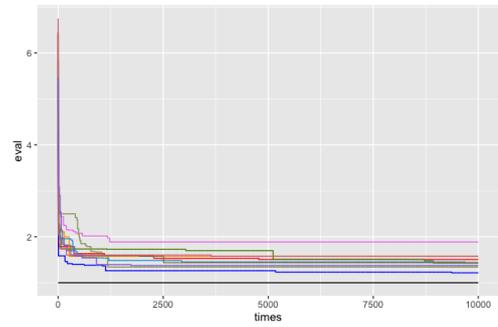


(b) SGA

Figura 6.9: Resultados preliminares para el problema 8

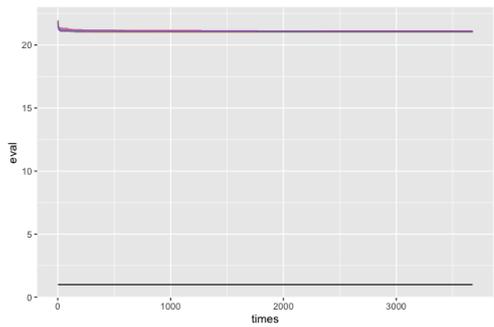


(a) epiGA

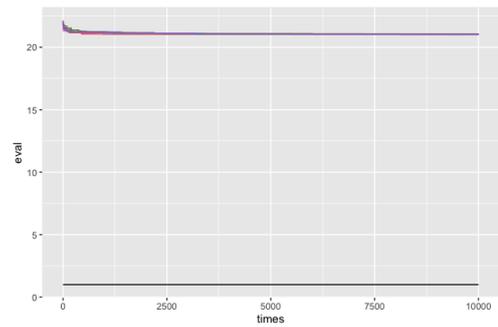


(b) SGA

Figura 6.10: Resultados preliminares para el problema 9



(a) epiGA



(b) SGA

Figura 6.11: Resultados preliminares para el problema 10

6.5. Resultados de epiGA con modificaciones

En esta sección se muestran las gráficas de desempeño de epiDA y su comparación con los otros dos algoritmos, después de realizar una calibración de parámetros mediante `irace` (en la tabla 6.3 se muestra la configuración encontrada). En las gráficas se utiliza el siguiente código de colores:

- **rojo** para la mejor solución en la población de hembras (población general en el caso de SGA y epiGA)
- **azul** para la mejor solución en la población de machos
- **morado** para el promedio de soluciones en la población de hembras en cada generación
- **verde** para el promedio de soluciones en la población de machos en cada generación (población general en el caso de SGA y epiGA)

Al igual que en los experimentos preliminares, se observa una marcada diferencia en el número de generaciones

No. individuos	No. células	Prob. metilación	Prob. entorno	Radio de nucleosoma
87	9	0.586223977	0.009903721	1

Tabla 6.3: Configuración de parámetros utilizada en las pruebas preliminares en epiDA

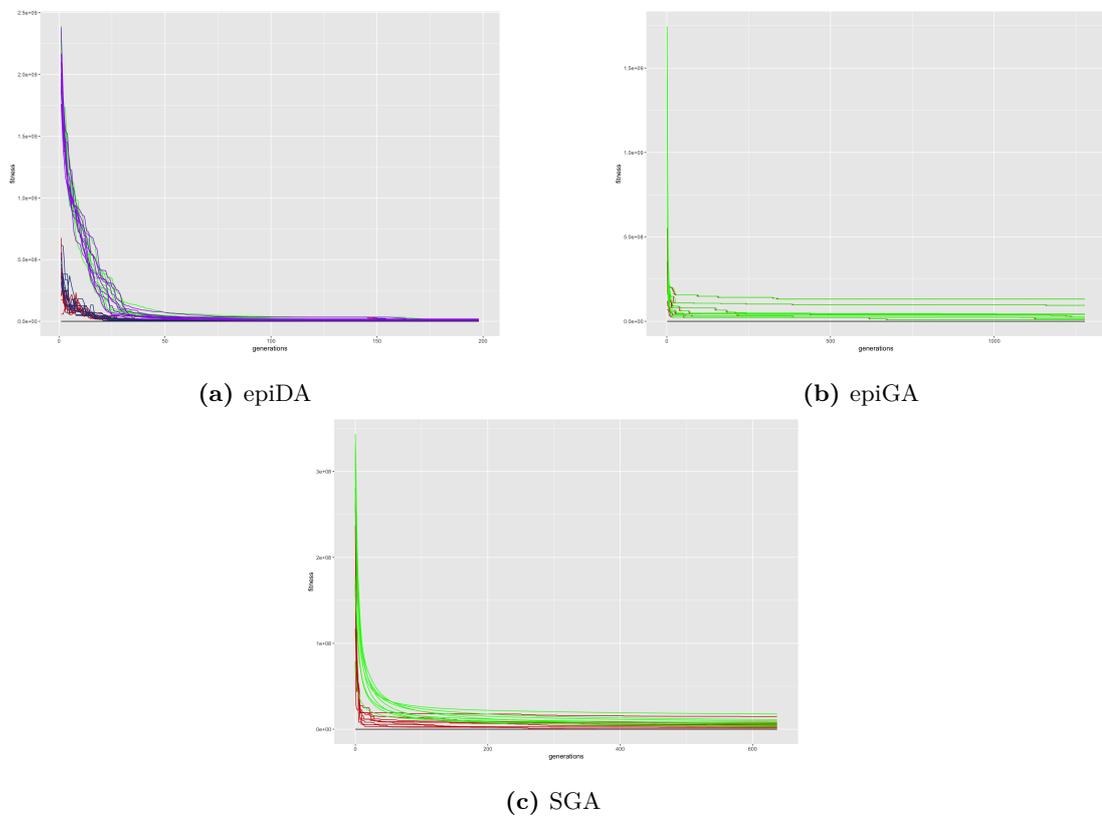
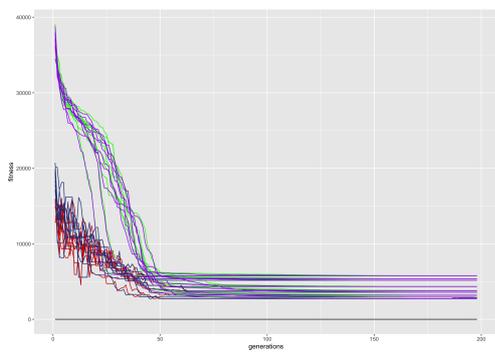
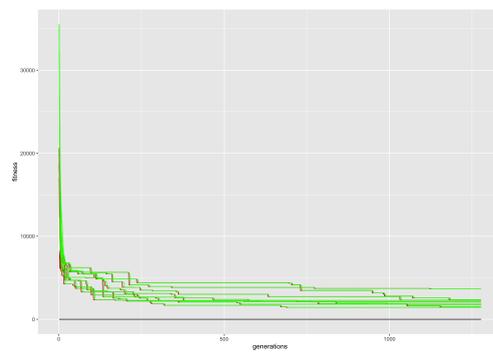


Figura 6.12: Comparación de resultados para el problema 1

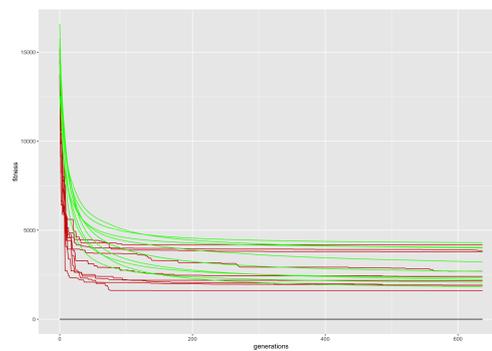
6. EXPERIMENTACIÓN Y RESULTADOS



(a) epiDA

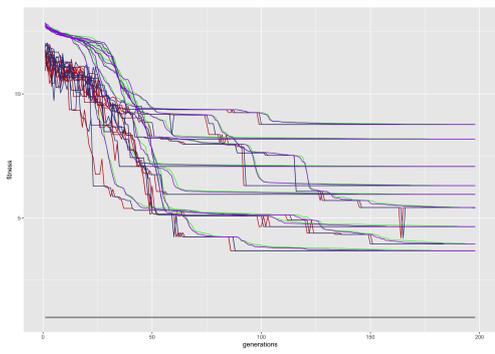


(b) epiGA

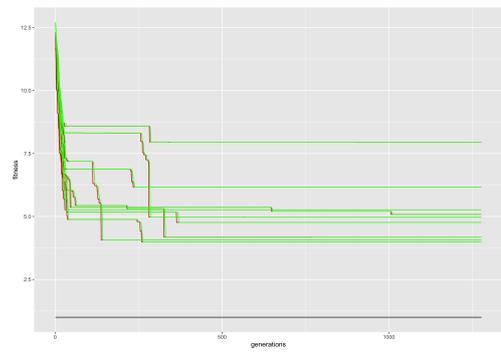


(c) SGA

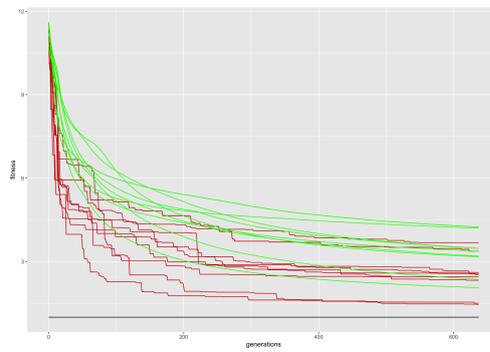
Figura 6.13: Comparación de resultados para el problema 2



(a) epiDA



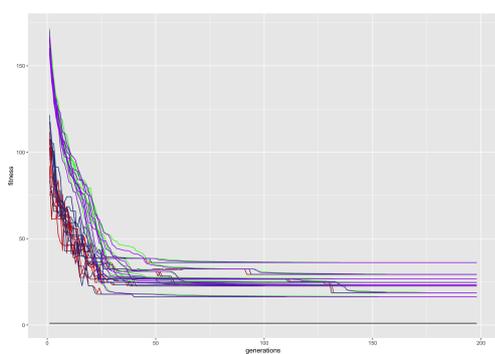
(b) epiGA



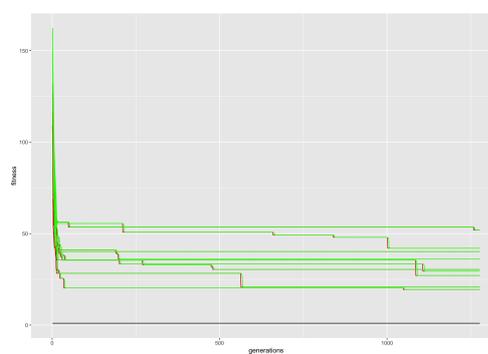
(c) SGA

Figura 6.14: Comparación de resultados para el problema 3

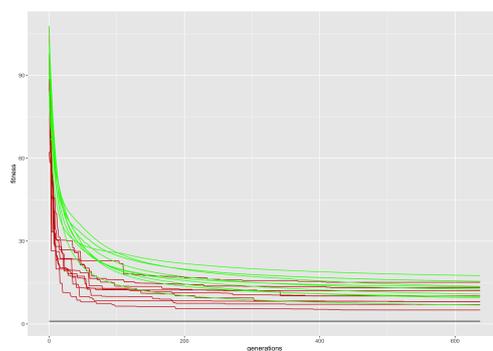
6. EXPERIMENTACIÓN Y RESULTADOS



(a) epiDA



(b) epiGA



(c) SGA

Figura 6.15: Comparación de resultados para el problema 4

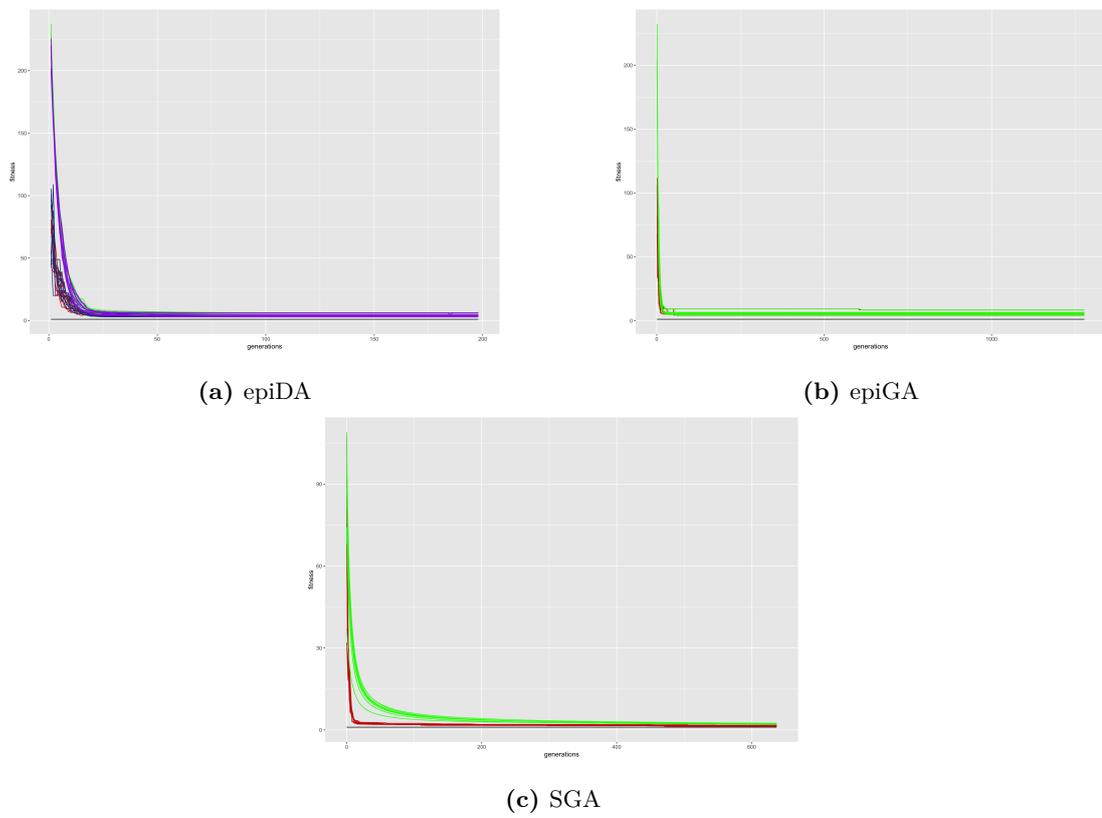
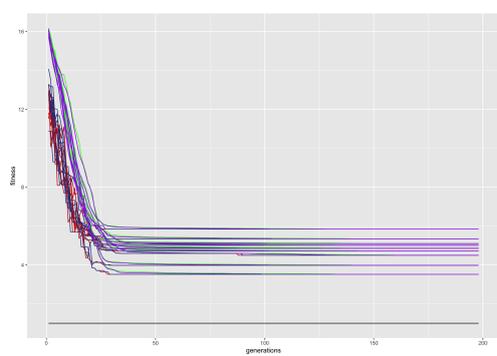
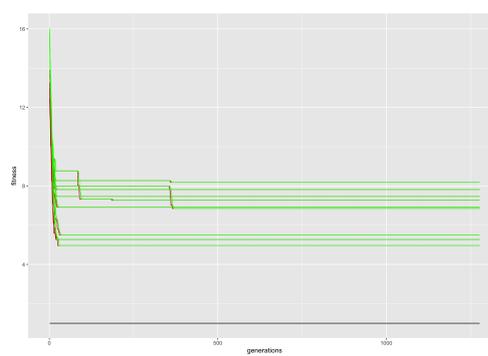


Figura 6.16: Comparación de resultados para el problema 5

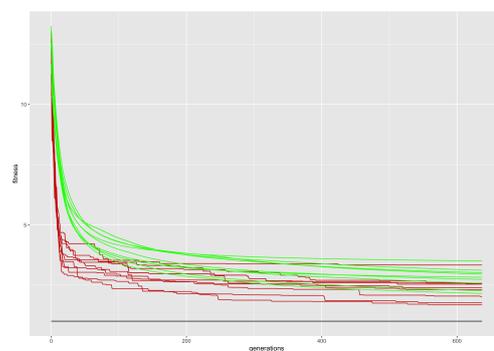
6. EXPERIMENTACIÓN Y RESULTADOS



(a) epiDA



(b) epiGA



(c) SGA

Figura 6.17: Comparación de resultados para el problema 6

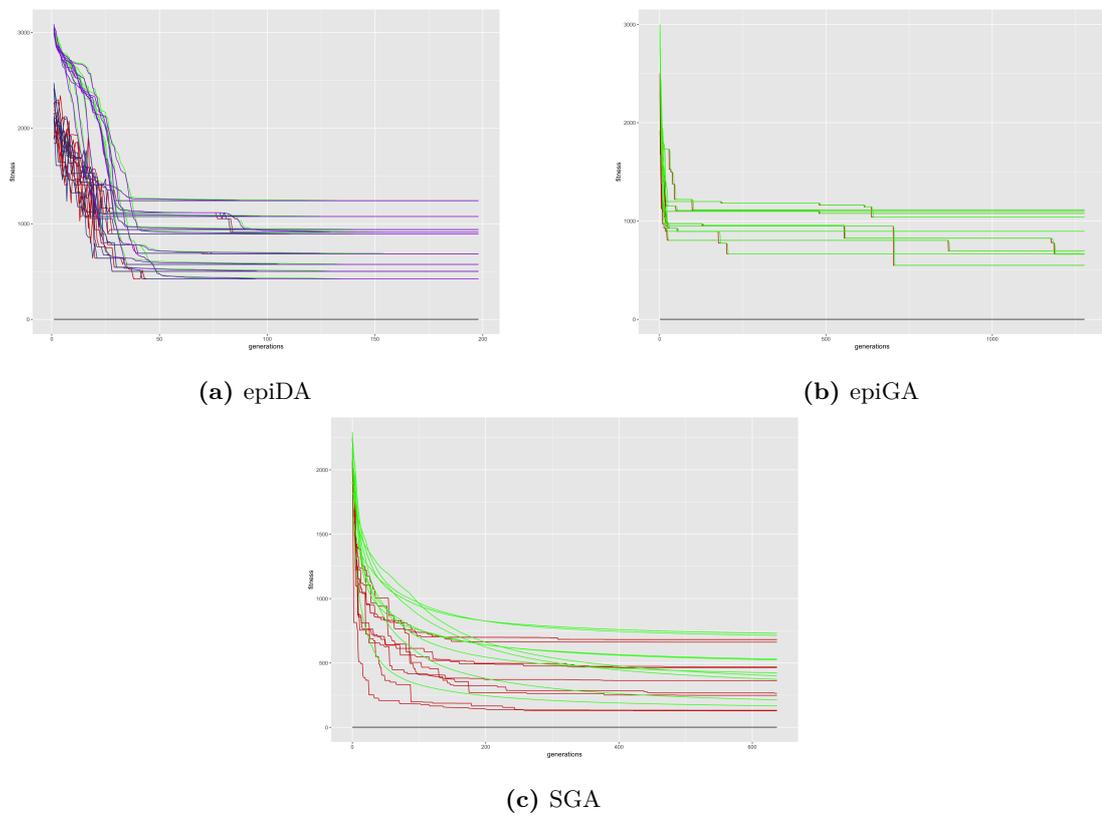


Figura 6.18: Comparación de resultados para el problema 7

6. EXPERIMENTACIÓN Y RESULTADOS

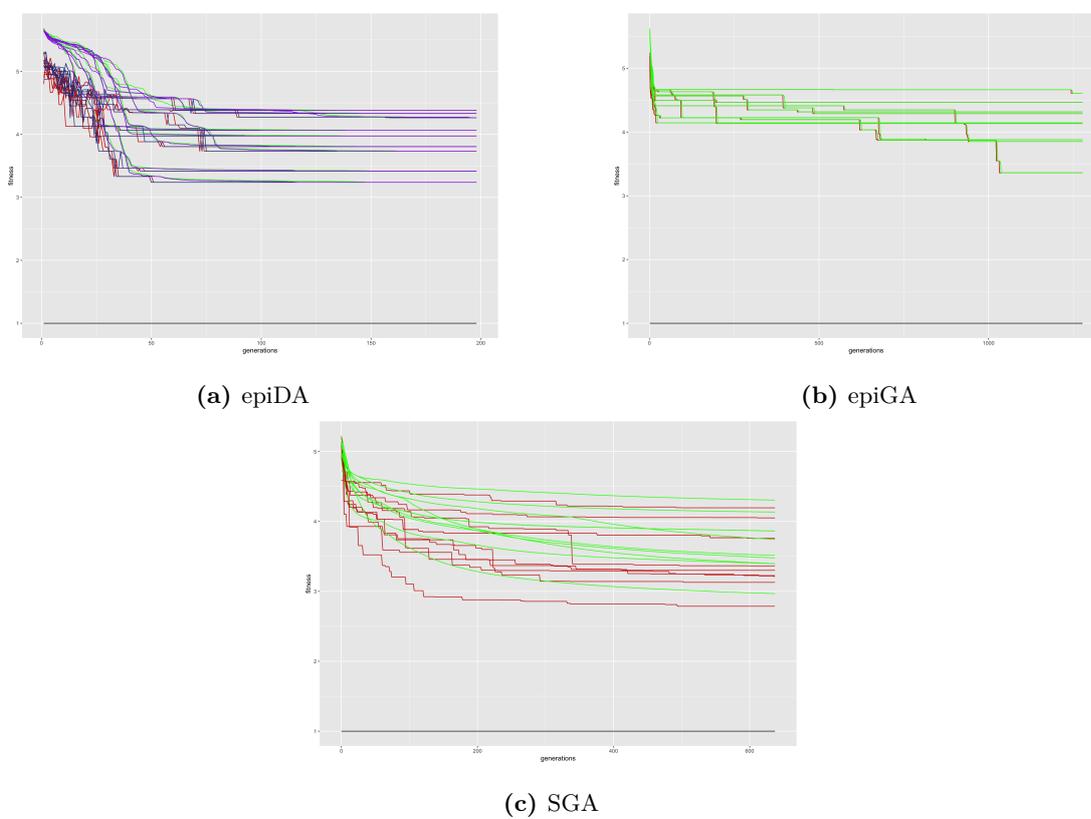


Figura 6.19: Comparación de resultados para el problema 8

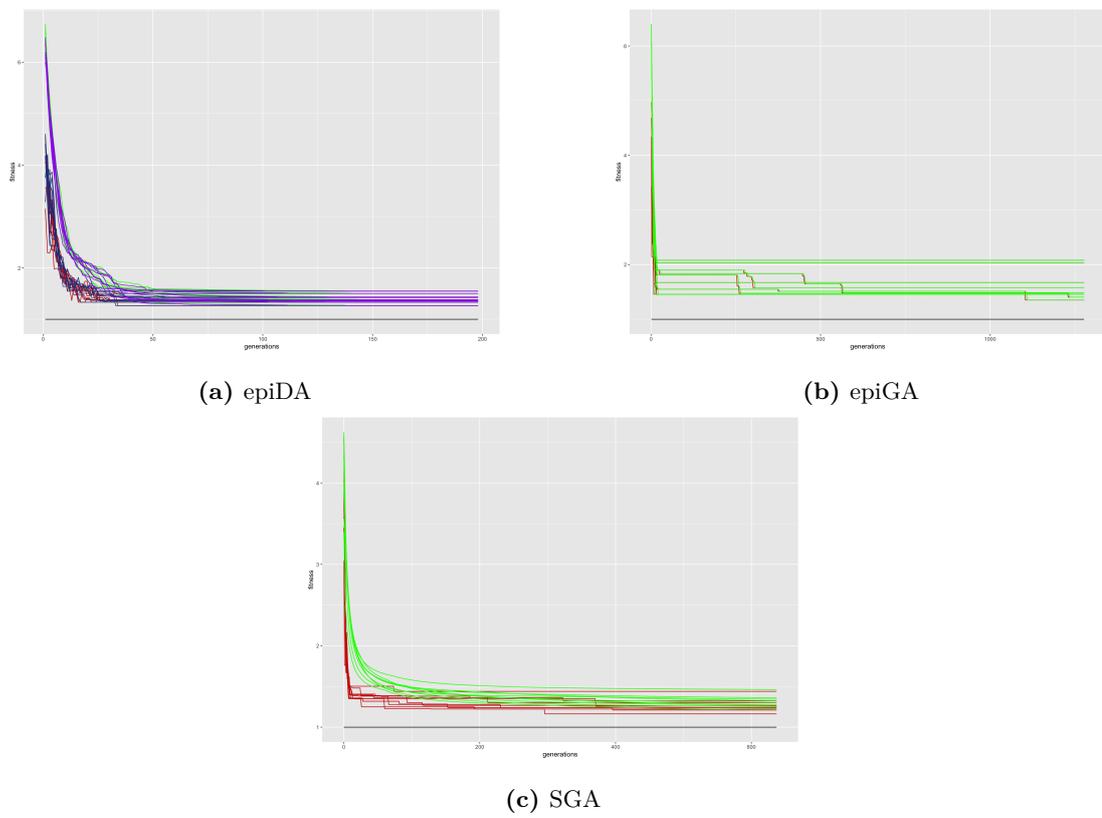


Figura 6.20: Comparación de resultados para el problema 9

6. EXPERIMENTACIÓN Y RESULTADOS

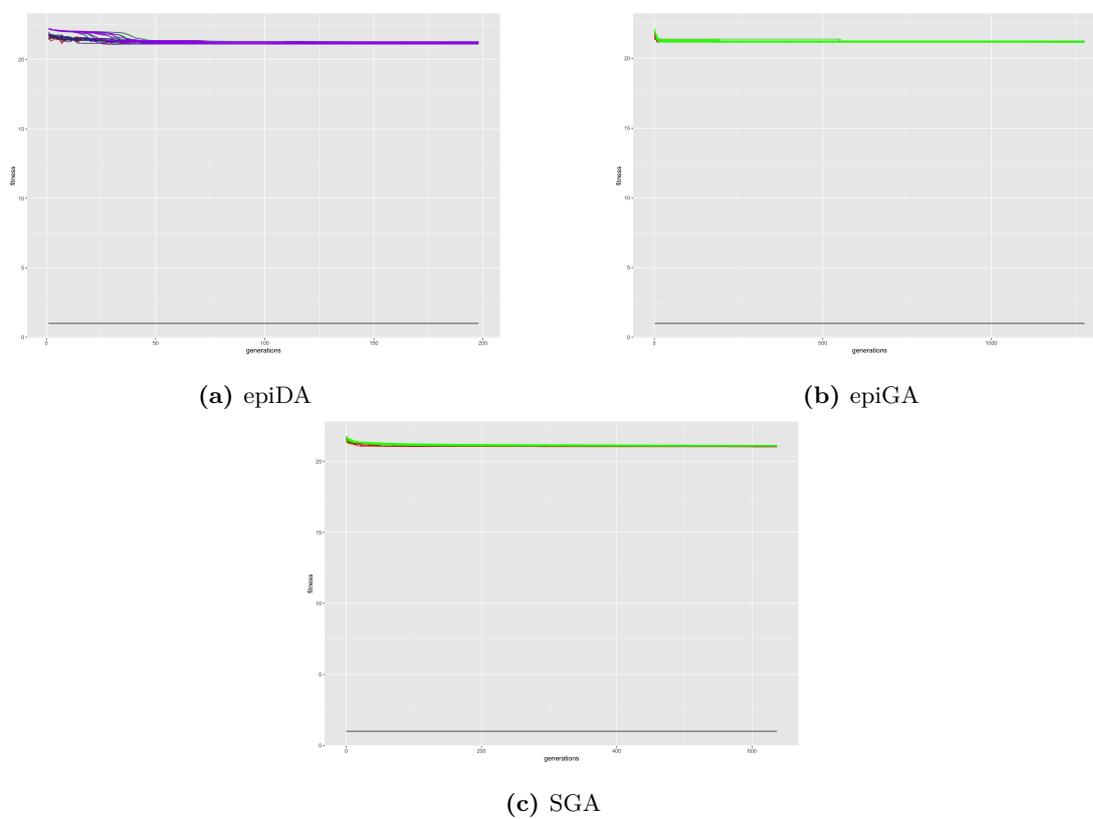


Figura 6.21: Comparación de resultados para el problema 10

	SGA	epiGA	epiDA	Solución óptima
P1 (Storn's Chebyshev Polynomial Fitting Problem)	1900136	9636719	4684252	1.000000000
P2 (Inverse Hilbert Matrix Problem)	1608	1439	2718	1.000000000
P3 (Lennard-Jones Minimum Energy Cluster)	1.458	3.987	3.680	1.000000000
P4 (Rastrigin's Function)	5.044	19.47	16.32	1.000000000
P5 (Griewangk's Function)	1.282	4.000	2.895	1.000000000
P6 (Weierstrass Function)	1.679	4.963	3.517	1.000000000
P7 (Modified Schwefel's Function)	128.3	550.03	422.6	1.000000000
P8 (Expanded Schaffer's F6 Function)	2.787	3.361	1.266	1.000000000
P9 (Happy Cat Function)	1.163	1.353	1.266	1.000000000
P10 (Ackley Function)	21.02	21.13	21.12	1.000000000

Tabla 6.4: Mejores soluciones encontradas por cada algoritmo

6.6. Análisis

Como se mencionó anteriormente, se utilizaron las pruebas omnibus de Friedman, Friedman alineado y Quade para determinar si existe al menos un algoritmo que tenga un desempeño significativamente diferente a alguno de los otros. En caso de ser así, se procederá a realizar la prueba de Bergmann-Hommel de comparación entre pares.

Friedman	Friedman alineado	Quade
0.0006755	0.005908	0.003706

Tabla 6.5: p -valores preliminares

Utilizando como condición $p < 0.05$, podemos observar en la tabla 6.5 que dicha condición se cumple en los resultados de las tres pruebas, lo que indica que existe diferencia estadística entre al menos dos algoritmos. Con esta información, es posible proceder a realizar la prueba *post hoc* de Bergmann-Hommel. A su vez, es posible apreciar en la tabla 6.6, utilizando los promedios de las posiciones obtenidas por los tres algoritmos en cada problema, que SGA ocupa la primera posición, seguido de epiDA y epiGA en último lugar.

6. EXPERIMENTACIÓN Y RESULTADOS

SGA	epiGA	epiDA
1.1	2.8	2.1

Tabla 6.6: Resultados de posicionamiento para cada algoritmo

	SGA	epiGA	epiDA
SGA	NA	0	0.03
epiDA	0.03	0.12	NA
epiGA	0	NA	0.12

Tabla 6.7: p -valores ajustados para prueba de Friedman

	SGA	epiGA	epiDA
SGA	NA	0	0
epiDA	0	0.58	NA
epiGA	0	NA	0.58

Tabla 6.8: p -valores ajustados para prueba de posiciones alineadas de Friedman

	SGA	epiGA	epiDA
SGA	NA	0.05	0.01
epiDA	0.05	0.31	NA
epiGA	0.01	NA	0.31

Tabla 6.9: p -valores ajustados para prueba de Quade

Las tablas 6.7, 6.8 y 6.9 muestran la comparación de significación estadística entre los tres algoritmos utilizando los resultados de las tres pruebas mencionadas, ajustados mediante la prueba *post hoc* de Bergmann-Hommel utilizando como condición $p < 0.05$. En las tres tablas es posible apreciar que existe diferencia significativa entre

SGA y epiGA y entre SGA y epiDA. Sin embargo, de acuerdo a los resultados de la prueba de Bergmann-Hommel para la información obtenida por las tres pruebas omnibus, no existe diferencia significativa entre epiGA y epiDA, a pesar del resultado de posicionamiento mostrado en la tabla 6.6.

Como se menciona en las definiciones del concurso, las funciones 4 a la 10 cuentan con un desplazamiento y una rotación que ubican la solución óptima en el rango $[-80, 80]$, a diferencia de los tres primeros problemas que no cuentan con estas modificaciones. Esto hace que la calibración de parámetros para el algoritmo resulte más complicada, por lo que la configuración obtenida por `irace` no será óptima para todos los problemas. Es de hacer notar el caso del problema 10, el cual resultó especialmente difícil de tratar por los tres algoritmos con las configuraciones de parámetros encontradas. Sin embargo, en el caso de los algoritmos participantes en el concurso del CEC 2019, fue el único problema en el que todos lograron encontrar la solución óptima. Otro aspecto a considerar es el número de evaluaciones. Para los experimentos realizados se fijó el límite de $1e6$ evaluaciones, tomando en cuenta que `jDE100` encontró la solución óptima en un rango de $1e5$ y $1e8$ en la mayoría de los problemas. Claramente los tres algoritmos comparados en esta fase experimental no logran alcanzar soluciones cercanas a la solución óptima dentro del límite establecido.

Capítulo 7

Conclusiones y discusión

A pesar de que los resultados del análisis estadístico no muestran diferencia significativa entre epiGA y epiDA, se considera que los datos mostrados en las tablas de posicionamiento y las gráficas apuntan a que con una profundización en los ajustes de los parámetros y la implementación de nuevas propuestas de mecanismos de selección, además de otros mecanismos epigenéticos, se puede tener un desempeño favorable por parte de algoritmos con enfoque epigenético. De manera especial la integración de los demás operadores epigenéticos mencionados por los autores resulta interesante y se propone como un punto de trabajo futuro, pues alguno de ellos podría presentar mejores resultados en diferentes tipos de problemas, sin dejar de tomar en cuenta la posibilidad de tener más de un mecanismo activo en la ejecución del algoritmo. Así, se propone hacer énfasis en el análisis del papel que juega el vector nucleosoma en la herencia pues es claro que es ideal evitar los casos extremos de su configuración (donde todos los elementos sean 0 o 1). Como se mencionó anteriormente, Yuen *et al.* [58] consideran que ninguna de las propuestas existentes de algoritmos inspirados en epigenética, incluido epiGA, logra implementar el conjunto de marcadores epigenéticos de manera adecuada. De este modo, otra propuesta a futuro es desarrollar nuevos enfoques en la implementación de estos marcadores, los cuales podrían sustituir al vector nucleosoma. Cabe mencionar que dado que las propuestas en este trabajo de tesis se enfocaron en la dinámica de selección y remplazo, no se realizaron modificaciones en la estructura de las células creada por Stolfi y Alba [48].

Un tema de discusión interesante en el rubro de los algoritmos inspirados en la naturaleza, en especial los evolutivos y bio-inspirados, es la explosión reciente en el número de metáforas que han servido de inspiración para nuevos algoritmos. El éxito de las primeras metaheurísticas bio-inspiradas ha motivado a explorar y buscar comportamientos interesantes en las especies animales que se puedan traducir a nuevas propuestas de

7. CONCLUSIONES Y DISCUSIÓN

modelado de algoritmos. Sin embargo, se considera que la gran cantidad de propuestas presentadas en años recientes son muy similares a sus predecesoras y no representan un avance significativo para el área de estudio. Basta dar un vistazo a artículos como el de Kumar y Singh [30] y proyectos como el *Evolutionary computation bestiary*¹ para comprobar la gran diversidad de opciones existentes. Además, se alega que quienes se dedican al desarrollo de dichas propuestas ponen más esfuerzo en hacer ver a la metáfora como el elemento valioso de su trabajo, y no el aspecto matemático y computacional del mismo. En el caso del fenómeno que sirve como metáfora para la propuesta de los métodos de selección y remplazo en esta tesis, se busca exponer un paradigma que puede significar un cambio relevante y moderno acorde a los nuevos descubrimientos que la investigación en genómica ha develado en años recientes. Buscar maneras de modelar la expresión genética, junto con la influencia del medio ambiente en la misma, abre las puertas a encontrar mejoras en el desempeño de los algoritmos genéticos, cambiando la forma en la que estos algoritmos exploran el espacio de búsqueda. Además, se utiliza un modelo matemático ya existente del fenómeno de DA como base para medir la aptitud de los individuos hembras de la población, en lugar de tratar de adaptar el fenómeno a algún método más tradicional dentro del diseño de los operadores.

Finalmente, se considera que los temas tratados en este trabajo de tesis podrían trasladarse al campo de los sistemas complejos, donde, con los diferentes enfoques de análisis en dicha área, podrían encontrarse resultados interesantes en la idea principal detrás del diseño del algoritmo propuesto en esta tesis (el papel de los efectos parentales en la evolución). Es posible que mediante la implementación de un entorno de simulación que incluya las interacciones estudiadas permita hacer una implementación más adecuada de las funciones de aptitud y de que pueda hacerse un avance hacia la propuesta de Haaland *et al.* [22] de crear un modelo más detallado que mida la evolución de las preferencias de los individuos. Tomando en cuenta esto último, se considera que la colaboración interdisciplinaria entre investigadores en biología y computación es crucial para el avance en el ámbito del cómputo evolutivo.

¹Disponible en <https://github.com/fcampelo/EC-Bestiary>

Bibliografía

- [1] Angela M. Achorn y Gil G. Rosenthal. *It's Not about Him: Mismeasuring 'Good Genes' in Sexual Selection*. Mar. de 2020. DOI: [10.1016/j.tree.2019.11.007](https://doi.org/10.1016/j.tree.2019.11.007). URL: [http://www.cell.com/article/S0169534719303416/fulltext%20http://www.cell.com/article/S0169534719303416/abstract%20https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347\(19\)30341-6](http://www.cell.com/article/S0169534719303416/fulltext%20http://www.cell.com/article/S0169534719303416/abstract%20https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(19)30341-6) (vid. pág. 12).
- [2] Lifeth Álvarez Camacho. «Modeling Epigenetic Evolutionary Algorithms: An approach based on the Epigenetic Regulation process». Tesis de maestría. Bogotá, D.C.: Universidad Nacional de Colombia, 2020 (vid. pág. 30).
- [3] Vladimir Espinosa Angarica y Antonio del Sol. «Bioinformatics tools for genome-wide epigenetic research». En: *Advances in Experimental Medicine and Biology*. Vol. 978. Springer New York LLC, 2017, págs. 489-512. DOI: [10.1007/978-3-319-53889-1_25](https://doi.org/10.1007/978-3-319-53889-1_25). URL: <https://pubmed.ncbi.nlm.nih.gov/28523562/> (vid. pág. 27).
- [4] Carlos Ansótegui, Meinolf Sellmann y Kevin Tierney. «A Gender-Based Genetic Algorithm for the Automatic Configuration of Algorithms». En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5732 LNCS (2009), págs. 142-157. DOI: [10.1007/978-3-642-04244-7_14](https://doi.org/10.1007/978-3-642-04244-7_14). URL: <https://link-springer-com.pbidi>.

- [unam.mx:2443/chapter/10.1007/978-3-642-04244-7_7B%5C_%7D14](https://doi.org/10.1007/978-3-642-04244-7_7B%5C_%7D14) (vid. pág. 21).
- [5] Maria J. Aristizabal *et al.* «Biological embedding of experience: A primer on epigenetics». En: *Proceedings of the National Academy of Sciences* 117.38 (oct. de 2019), pág. 201820838. ISSN: 0027-8424. DOI: [10.1073/pnas.1820838116](https://doi.org/10.1073/pnas.1820838116). URL: www.pnas.org/cgi/doi/10.1073/pnas.1820838116 (vid. pág. 11).
- [6] Sandra Brasil *et al.* «Artificial intelligence in epigenetic studies: Shedding light on rare diseases». En: *Frontiers in molecular biosciences* 8 (2021), pág. 314 (vid. pág. 28).
- [7] Janez Brest, Mirjam Sepesy Maučec y Borko Bošković. «The 100-Digit Challenge: Algorithm jDE100». En: *2019 IEEE Congress on Evolutionary Computation (CEC)*. 2019, págs. 19-26. DOI: [10.1109/CEC.2019.8789904](https://doi.org/10.1109/CEC.2019.8789904) (vid. pág. 43).
- [8] Janez Brest *et al.* «Dynamic optimization using Self-Adaptive Differential Evolution». En: *2009 IEEE Congress on Evolutionary Computation*. 2009, págs. 415-422. DOI: [10.1109/CEC.2009.4982976](https://doi.org/10.1109/CEC.2009.4982976) (vid. pág. 44).
- [9] N. Burley. «The differential-allocation hypothesis: an experimental test». En: *American Naturalist* 132.5 (oct. de 1988), págs. 611-628. ISSN: 00030147. DOI: [10.1086/284877](https://doi.org/10.1086/284877). URL: <https://www.journals.uchicago.edu/doi/abs/10.1086/284877> (vid. pág. 12).
- [10] Nancy Burley. «Sexual Selection for Aesthetic Traits in Species with Biparental Care». En: *The American Naturalist* 127.4 (abr. de 1986), págs. 415-445. ISSN: 0003-0147. DOI: [10.1086/284493](https://doi.org/10.1086/284493) (vid. pág. 12).
- [11] Nancy T. Burley. «Differential allocation». En: *Encyclopedia of Animal Behavior*. Elsevier, ene. de 2019, págs. 458-463. ISBN: 9780128132517. DOI: [10.1016/B978-0-12-809633-8.90120-1](https://doi.org/10.1016/B978-0-12-809633-8.90120-1) (vid. pág. 42).
- [12] Borja Calvo y Guzman Santafe. «scmamp: Statistical Comparison of Multiple Algorithms in Multiple Problems». En: *The R Journal* Accepted for publication (2015) (vid. págs. 44, 45).

- [13] Nasibeh Chenarani *et al.* «Bioinformatic tools for DNA methylation and histone modification: A survey». En: *Genomics* 113.3 (mayo de 2021), págs. 1098-1113. ISSN: 0888-7543. DOI: [10.1016/J.YGENO.2021.03.004](https://doi.org/10.1016/J.YGENO.2021.03.004) (vid. pág. 28).
- [14] Kornel Chrominski, Magdalena Tkacz y Mariusz Boryczka. «Epigenetic Modification of Genetic Algorithm». En: Springer, Cham, jun. de 2020, págs. 267-278. DOI: [10.1007/978-3-030-50417-5_20](https://doi.org/10.1007/978-3-030-50417-5_20). URL: http://link.springer.com/10.1007/978-3-030-50417-5%7B%5C_%7D20 (vid. pág. 29).
- [15] Francis Crick. «Central dogma of molecular biology». En: *Nature* 227.5258 (1970), págs. 561-563 (vid. pág. 8).
- [16] Joaquín Derrac *et al.* «A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms». En: *Swarm and Evolutionary Computation* 1.1 (mar. de 2011), págs. 3-18. ISSN: 2210-6502. DOI: [10.1016/J.SWEVO.2011.02.002](https://doi.org/10.1016/J.SWEVO.2011.02.002) (vid. pág. 44).
- [17] Robert E. Furrow, Freddy B. Christiansen y Marcus W. Feldman. «Epigenetic variation, phenotypic heritability, and evolution». En: *Epigenetics and Complex Traits*. Springer New York, ene. de 2014, págs. 233-246. ISBN: 9781461480785. DOI: [10.1007/978-1-4614-8078-5_10](https://doi.org/10.1007/978-1-4614-8078-5_10). URL: https://link-springer-com.pbidi.unam.mx:2443/chapter/10.1007/978-1-4614-8078-5_10 (vid. pág. 9).
- [18] Edgar Galván y Peter Mooney. «Neuroevolution in Deep Neural Networks: Current Trends and Future Challenges». En: *CoRR* abs/2006.05415 (2020). arXiv: [2006.05415](https://arxiv.org/abs/2006.05415). URL: <https://arxiv.org/abs/2006.05415> (vid. pág. 21).
- [19] Patricia A. Gowaty. «Compensation in reproduction». En: *Encyclopedia of Animal Behavior*. Elsevier, ene. de 2019, págs. 453-457. ISBN: 9780128132517. DOI: [10.1016/B978-0-12-809633-8.20833-9](https://doi.org/10.1016/B978-0-12-809633-8.20833-9) (vid. pág. 13).
- [20] Patricia Adair Gowaty. «Reproductive compensation». En: *Journal of Evolutionary Biology* 21.5 (sep. de 2008), págs. 1189-1200. ISSN: 1010061X. DOI: [10.1111/](https://doi.org/10.1111/)

- j.1420-9101.2008.01559.x. URL: <http://doi.wiley.com/10.1111/j.1420-9101.2008.01559.x> (vid. pág. 13).
- [21] Carlos Guerrero-Bosagna. «Evolution with No Reason: A Neutral View on Epigenetic Changes, Genomic Variability, and Evolutionary Novelty». En: *BioScience* 67.5 (mayo de 2017), págs. 469-476. ISSN: 0006-3568. DOI: [10.1093/BIOSCI/BIX021](https://doi.org/10.1093/BIOSCI/BIX021). URL: <https://academic.oup.com/bioscience/article/67/5/469/3746563> (vid. pág. 11).
- [22] Thomas R Haaland *et al.* «Differential Allocation Revisited: When Should Mate Quality Affect Parental Investment?». En: *The American naturalist* 190.4 (oct. de 2017), págs. 534-546. ISSN: 1537-5323. DOI: [10.1086/693484](https://doi.org/10.1086/693484). URL: <http://www.ncbi.nlm.nih.gov/pubmed/28937815> (vid. págs. 14, 16, 38, 66).
- [23] Ehtasham Ul Haq *et al.* «A novel selection approach for genetic algorithms for global optimization of multimodal continuous functions». En: *Computational Intelligence and Neuroscience* 2019 (2019). ISSN: 16875273. DOI: [10.1155/2019/8640218](https://doi.org/10.1155/2019/8640218). URL: <https://doi.org/10.1155/2019/8640218> (vid. pág. 22).
- [24] W. Edwin Harris y Tobias Uller. «Reproductive investment when mate quality varies: Differential allocation versus reproductive compensation». En: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1520 (2009), págs. 1039-1048. ISSN: 14712970. DOI: [10.1098/rstb.2008.0299](https://doi.org/10.1098/rstb.2008.0299). URL: <https://pmc/articles/PMC2666688/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2666688/> (vid. pág. 14).
- [25] Ahmad B. A. Hassanat y Esra'a Alkafaween. «On Enhancing Genetic Algorithms Using New Crossovers». En: (ene. de 2018). arXiv: [1801.02335](https://arxiv.org/abs/1801.02335). URL: <http://arxiv.org/abs/1801.02335> (vid. págs. 23, 24).
- [26] Yanghua He y Jiuzhou Song. «Bioinformatics Analysis of Epigenetics». En: *Bioinformatics in Aquaculture*. Chichester, UK: John Wiley & Sons, Ltd, feb. de 2017, págs. 247-273. DOI: [10.1002/9781118782392.ch15](https://doi.org/10.1002/9781118782392.ch15). URL: <http://doi.wiley.com/10.1002/9781118782392.ch15> (vid. pág. 27).

- [27] Konrad Hochedlinger y Kathrin Plath. «Epigenetic reprogramming and induced pluripotency». En: *Development (Cambridge, England)* 136.4 (feb. de 2009), págs. 509-523. ISSN: 0950-1991. DOI: [10.1242/DEV.020867](https://doi.org/10.1242/DEV.020867). URL: <https://pubmed.ncbi.nlm.nih.gov/19168672/> (vid. pág. 10).
- [28] Hassan Ismkhan. «Black box optimization using evolutionary algorithm with novel selection and replacement strategies based on similarity between solutions». En: *Applied Soft Computing Journal* 64 (mar. de 2018), págs. 260-271. ISSN: 15684946. DOI: [10.1016/j.asoc.2017.12.006](https://doi.org/10.1016/j.asoc.2017.12.006) (vid. págs. 22, 24).
- [29] Laurent Kappeler y Michael J. Meaney. «Epigenetics and parental effects». En: *BioEssays* 32.9 (sep. de 2010), págs. 818-827. ISSN: 1521-1878. DOI: [10.1002/BIES.201000015](https://doi.org/10.1002/BIES.201000015). URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/bies.201000015> <https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201000015> <https://onlinelibrary.wiley.com/doi/10.1002/bies.201000015> (vid. pág. 16).
- [30] Sachan Rohit Kumar y Kushwaha Dharmender Singh. «Nature-Inspired Optimization Algorithms: Research Direction and Survey». En: (feb. de 2021). arXiv: [2102.04013](https://arxiv.org/abs/2102.04013). URL: <http://arxiv.org/abs/2102.04013> (vid. pág. 66).
- [31] William La Cava *et al.* «Genetic Programming with Epigenetic Local Search». En: (2015). DOI: [10.1145/2739480.2754763](https://doi.org/10.1145/2739480.2754763). URL: <http://dx.doi.org/10.1145/2739480.2754763> (vid. pág. 29).
- [32] Kevin N. Laland *et al.* «The extended evolutionary synthesis: its structure, assumptions and predictions». En: *Proceedings of the Royal Society B: Biological Sciences* 282.1813 (ago. de 2015). ISSN: 14712954. DOI: [10.1098/RSPB.2015.1019](https://doi.org/10.1098/RSPB.2015.1019). URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2015.1019> (vid. pág. 7).
- [33] Manuel López-Ibáñez *et al.* «The irace package: Iterated Racing for Automatic Algorithm Configuration». En: *Operations Research Perspectives* 3 (2016), págs. 43-58. DOI: [10.1016/j.orp.2016.09.002](https://doi.org/10.1016/j.orp.2016.09.002) (vid. pág. 44).

- [34] Felicia M. Low y Peter D. Gluckman. «Epigenetic Mechanisms». En: *Neonatology: A Practical Approach to Neonatal Diseases*. Cham, Switzerland: Springer, jun. de 2016, págs. 1-8. ISBN: 978-3-319-18159-2. DOI: [10.1007/978-3-319-18159-2_152-1](https://doi.org/10.1007/978-3-319-18159-2_152-1) (vid. pág. 8).
- [35] Efrén Mezura-Montes y Omar Cetina-Domínguez. «Empirical analysis of a modified Artificial Bee Colony for constrained numerical optimization». En: *Applied Mathematics and Computation* 218.22 (jul. de 2012), págs. 10943-10973. ISSN: 00963003. DOI: [10.1016/j.amc.2012.04.057](https://doi.org/10.1016/j.amc.2012.04.057) (vid. pág. 40).
- [36] Sathish Periyasamy, Alex Gray y Peter Kille. «The epigenetic algorithm». En: *2008 IEEE Congress on Evolutionary Computation, CEC 2008*. 2008, págs. 3228-3236. ISBN: 9781424418237. DOI: [10.1109/CEC.2008.4631235](https://doi.org/10.1109/CEC.2008.4631235) (vid. págs. 29, 31).
- [37] Anna Qvarnström y Trevor D. Price. *Maternal effects, paternal effects and sexual selection*. Feb. de 2001. DOI: [10.1016/S0169-5347\(00\)02063-2](https://doi.org/10.1016/S0169-5347(00)02063-2) (vid. pág. 12).
- [38] M. M. Raghuvanshi y O. G. Kakde. «Genetic algorithm with species and sexual selection». En: *2006 IEEE Conference on Cybernetics and Intelligent Systems* (2006). DOI: [10.1109/ICCIS.2006.252229](https://doi.org/10.1109/ICCIS.2006.252229) (vid. pág. 21).
- [39] Irja Ida Ratikainen y Hanna Kokko. «Differential allocation and compensation: who deserves the silver spoon?» En: *Behavioral Ecology* 21.1 (nov. de 2009), págs. 195-200. ISSN: 1045-2249. DOI: [10.1093/beheco/arp168](https://doi.org/10.1093/beheco/arp168). eprint: <https://academic.oup.com/beheco/article-pdf/21/1/195/17279058/arp168.pdf>. URL: <https://doi.org/10.1093/beheco/arp168> (vid. pág. 14).
- [40] Elijah Reyes y Hope Klug. *Parental Investment and Sexual Selection*. 2021. DOI: [10.1007/978-3-319-19650-3_1691](https://doi.org/10.1007/978-3-319-19650-3_1691). URL: https://link-springer-com.pbidi.unam.mx:2443/referenceworkentry/10.1007/978-3-319-19650-3%7B%5C_%7D1691 (vid. pág. 12).
- [41] Esteban Ricalde. *A Genetic Programming System with an Epigenetic Mechanism for Traffic Signal Control*. 2019. arXiv: [1903.03854](https://arxiv.org/abs/1903.03854) [cs.NE] (vid. pág. 29).

-
- [42] Christina L. Richards, Koen J.F. Verhoeven y Oliver Bossdorf. «Evolutionary significance of epigenetic variation». En: *Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics*. Springer-Verlag Vienna, ene. de 2012, págs. 257-274. ISBN: 9783709111307. DOI: [10.1007/978-3-7091-1130-7_16](https://doi.org/10.1007/978-3-7091-1130-7_16). URL: https://link-springer-com.pbidi.unam.mx:2443/chapter/10.1007/978-3-7091-1130-7%7B%5C_%7D16 (vid. pág. 16).
- [43] *Roadmap Epigenomics Project - Overview*. URL: <http://www.roadmapepigenomics.org/overview> (visitado 26-03-2021) (vid. pág. 9).
- [44] Heather Ruskin y Ana Barat. «Recent advances in computational epigenetics». En: *Advances in Genomics and Genetics Volume 8* (dic. de 2017), págs. 1-12. ISSN: 1179-9870. DOI: [10.2147/agg.s115524](https://doi.org/10.2147/agg.s115524). URL: www.dovepress.com (vid. págs. 27, 28).
- [45] José Sánchez-Velazco y John A. Bullinaria. «A Gendered Selection Strategies in Genetic Algorithms for Optimization». En: *University of Bristol*. 2003, págs. 217-223 (vid. págs. 21, 22, 38).
- [46] Ben C. Sheldon. «Differential allocation: tests, mechanisms and implications». En: *Trends in Ecology & Evolution* 15.10 (oct. de 2000), págs. 397-402. ISSN: 0169-5347. DOI: [10.1016/S0169-5347\(00\)01953-4](https://doi.org/10.1016/S0169-5347(00)01953-4). URL: [http://www.cell.com/article/S0169534700019534/fulltext%20http://www.cell.com/article/S0169534700019534/abstract%20https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347\(00\)01953-4](http://www.cell.com/article/S0169534700019534/fulltext%20http://www.cell.com/article/S0169534700019534/abstract%20https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(00)01953-4) (vid. pág. 14).
- [47] Yang Shi. «Taking Epigenetics Center Stage». En: *Cell* 128.4 (feb. de 2007), págs. 639-640. ISSN: 0092-8674. DOI: [10.1016/J.CELL.2007.02.011](https://doi.org/10.1016/J.CELL.2007.02.011). URL: [http://www.cell.com/article/S0092867407001924/fulltext%20http://www.cell.com/article/S0092867407001924/abstract%20https://www.cell.com/cell/abstract/S0092-8674\(07\)00192-4](http://www.cell.com/article/S0092867407001924/fulltext%20http://www.cell.com/article/S0092867407001924/abstract%20https://www.cell.com/cell/abstract/S0092-8674(07)00192-4) (vid. pág. 4).
- [48] Daniel H. Stolfi y Enrique Alba. «Epigenetic algorithms: A New way of building GAs based on epigenetics». En: *Information Sciences* 424 (ene. de 2018),
-

BIBLIOGRAFÍA

- págs. 250-272. ISSN: 00200255. DOI: [10.1016/j.ins.2017.10.005](https://doi.org/10.1016/j.ins.2017.10.005) (vid. págs. 31, 33, 34, 37, 65).
- [49] I. Tanev y K. Yuta. «Epigenetic programming: An approach of embedding epigenetic learning via modification of histones in genetic programming». En: *2003 Congress on Evolutionary Computation, CEC 2003 - Proceedings 4* (2003), págs. 2580-2587. DOI: [10.1109/CEC.2003.1299413](https://doi.org/10.1109/CEC.2003.1299413) (vid. pág. 29).
- [50] Céline Tiffon. «The Impact of Nutrition and Environmental Epigenetics on Human Health and Disease». En: *International Journal of Molecular Sciences 2018, Vol. 19, Page 3425* 19.11 (nov. de 2018), pág. 3425. ISSN: 1422-0067. DOI: [10.3390/IJMS19113425](https://doi.org/10.3390/IJMS19113425). URL: <https://www.mdpi.com/1422-0067/19/11/3425/html><https://www.mdpi.com/1422-0067/19/11/3425> (vid. pág. 4).
- [51] Peter M. Todd y Geoffrey F. Miller. «Parental Guidance Suggested: How Parental Imprinting Evolves Through Sexual Selection as an Adaptive Learning Mechanism». En: *Adaptive Behavior* 2.1 (jul. de 1993), págs. 5-47. ISSN: 17412633. DOI: [10.1177/105971239300200102](https://doi.org/10.1177/105971239300200102). URL: <https://journals.sagepub.com/doi/10.1177/105971239300200102> (vid. pág. 20).
- [52] Aymeric Vie. «Qualities, challenges and future of genetic algorithms: a literature review». En: (nov. de 2020). arXiv: [2011.05277](https://arxiv.org/abs/2011.05277). URL: <http://arxiv.org/abs/2011.05277><https://arxiv.org/abs/2011.05277v2> (vid. págs. 19, 23).
- [53] Paolo Vineis. «The Epigenetic Landscape». En: *Health Without Borders: Epidemics in the Era of Globalization*. Cham: Springer International Publishing, 2017, págs. 67-74. ISBN: 978-3-319-52446-7. DOI: [10.1007/978-3-319-52446-7_8](https://doi.org/10.1007/978-3-319-52446-7_8). URL: https://doi.org/10.1007/978-3-319-52446-7_8 (vid. págs. 9, 11).
- [54] C. H. Waddington. «The Epigenotype». En: *International Journal of Epidemiology* 41.1 (dic. de 2011), págs. 10-13. ISSN: 0300-5771. DOI: [10.1093/ije/dyr184](https://doi.org/10.1093/ije/dyr184). eprint: <https://academic.oup.com/ije/article-pdf/41/1/10/1824404/dyr184.pdf>. URL: <https://doi.org/10.1093/ije/dyr184> (vid. pág. 9).

- [55] Guoying Wang *et al.* «Epigenetics and Early Life Origins of Chronic Noncommunicable Diseases». En: *Journal of Adolescent Health* 52.2 (feb. de 2013), S14-S21. ISSN: 1054-139X. DOI: [10.1016/J.JADOHEALTH.2012.04.019](https://doi.org/10.1016/J.JADOHEALTH.2012.04.019) (vid. pág. 4).
- [56] David H. Wolpert y William G. Macready. «No free lunch theorems for optimization». En: *IEEE Transactions on Evolutionary Computation* 1.1 (1997), págs. 67-82. DOI: [10.1109/4235.585893](https://doi.org/10.1109/4235.585893) (vid. pág. 19).
- [57] Angela Yen. «Computational epigenomics: gene regulation, comparative methodologies, and epigenetic patterns». Tesis doct. Massachusetts Institute of Technology, 2016 (vid. pág. 27).
- [58] Sizhe Yuen, Thomas H. G. Ezard y Adam J. Sobey. «Epigenetic opportunities for Evolutionary Computation». En: (ago. de 2021). arXiv: [2108.04546](https://arxiv.org/abs/2108.04546). URL: <https://arxiv.org/abs/2108.04546v1> (vid. págs. 31, 37, 65).
- [59] Yani Zhu, Zhongxiu Yang y Jiatao Song. «A Genetic Algorithm with Age and Sexual Features». En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4113 LNCS - I (2006), págs. 634-640. DOI: [10.1007/11816157_75](https://doi.org/10.1007/11816157_75). URL: https://link.springer.com/chapter/10.1007/11816157%7B%5C_%7D75 (vid. pág. 21).