



PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

CARACTERIZACIÓN ESTRUCTURAL DEL GEN HOMEBOX ABD-A DE *DROSOPHILA MELANOGASTER* Y SU GEN HOMEBOX ORTÓLOGO HOXC6 EN *HOMO SAPIENS*.

**TRABAJO DE TITULACIÓN
MAESTRÍA EN BIOLOGÍA COMPUTACIONAL**

CLARA YAMILET SERRANO DELGADO

DRA. DORIS VELA.

Quito, Ecuador

2023

TABLA DE CONTENIDO

1.	INTRODUCCIÓN	7
2.	JUSTIFICACIÓN.....	8
3.	OBJETIVOS.....	9
3.1.	Objetivo general.....	9
3.2.	Objetivos específicos	9
4.	PLANTEAMIENTO DEL PROBLEMA.....	10
5.	FUNDAMENTACIÓN TEÓRICA.....	12
5.1.	Antecedentes.....	12
5.2.	Definición, Clasificación y Estructura de los genes.....	14
5.2.1.	Genoma	14
5.2.2.	Estructura de los genes.	14
5.2.3.	Estructura del gen que codifican ARN mensajero	14
5.3.	Clasificación de los genes	15
5.3.1.	Homología.....	15
5.3.2.	Genes ortólogos.....	16
5.4.	Genes HOX	16
5.5.	Desarrollo y organización genética de <i>Drosophila melanogaster</i>	17
5.5.1.	Genes homeóticos de <i>Drosophila melanogaster</i>	19
5.5.2.	Genes HOX en el ser humano.....	22
5.5.3.	Gen HOXC6 en <i>Homo sapiens</i>	23
5.5.4.	Desregulación de genes HOX y cáncer.....	24
5.6.	Perspectiva a futuro.	27
5.6.1.	Genes HOX como dianas terapéuticas.....	27
6.	METODOLOGIA	27
6.1.	Identificación de Herramientas Bioinformáticas.....	28
6.2.	Búsqueda de la información.....	30
6.3.	Recuperación de secuencias.....	30
6.4.	Selección del identificador del gen a estudiar.....	31
6.5.	Análisis de los genes abd-A y HOXC6	31
7.	RESULTADOS	36
8.	DISCUSIÓN	63
9.	CONCLUSIONES	65
10.	RECOMENDACIONES.....	65
11.	REFERENCIAS BIBLIOGRÁFICAS	67
12.	ANEXOS	70

ÍNDICE DE FIGURAS

Figura 1. Organización del gen que codifica al ARNm en los genomas eucarióticos.....	15
Figura 2. Estructura del gen HOX y la proteína HOX. que codifica al ARNm en los genomas eucariótico	17
Figura 3. Genes HOX en la mosca <i>Drosophila melanogaster</i>	20
Figura 4. Complejo Antennapedia de <i>Drosophila melanogaster</i>	21
Figura 5. Complejo Ultrabithorax de <i>Drosophila melanogaster</i>	21
Figura 6. Genes Hox en el ser humano	23
Figura 7. Gen HOXC6 en el ser humano.....	23
Figura 8. Proteína HOXC6.....	24
Figura 9. Descripción general de la participación de HOX en diferentes tumores en humanos..	26
Figura 10. Página web de acceso a GenBank.....	28
Figura 11. Página web de acceso a UniProt.....	29
Figura 12. Página web de acceso a Ensembl.....	29
Figura 13. Página web de acceso a FlyBase	30
Figura 14. Secuencia de nucleótidos del gen HOXC6.....	30
Figura 15. Búsqueda de secuencias del gen HOXC6 en la plataforma Ensembl	35
Figura 16. Búsqueda de secuencias del gen abd-A en la plataforma Ensembl.....	35
Figura 17. Comparación del número de adeninas del gen HOXC6 del <i>Homo sapiens</i> y abd-A de <i>Drosophila melanogaster</i>	46
Figura 18. Comparación del número de timinas del gen HOXC6 del <i>Homo sapiens</i> y abd-A de <i>Drosophila melanogaster</i>	46
Figura 19. Comparación del número de nucleótidos resultantes del gen HOXC6 del <i>Homo sapiens</i> y abd-A de <i>Drosophila melanogaster</i>	47
Figura 20. Ubicación del gen HOXC6	47
Figura 21. Análisis del transcripto HOXC6-201	48
Figura 22. Ubicación del transcripto HOXC6-201.....	49
Figura 23. Ubicación del gen abd-A en el cromosoma 3 de <i>Drosophila melanogaster</i>	49
Figura 24. Región en detalle del gen abd-A de <i>Drosophila melanogaster</i>	50
Figura 25. Características del gen abd-A. motivo, unión al ADN, región.....	51
Figura 26. Proteína Homeobox. Sitio conservado.....	51
Figura 27. Segmento codificador del Gen HOXC6.	52
Figura 28. Estructura nucleotídica del gen HOXC6.....	53
Figura 29. Características del gen HOXC6: motivo, unión al ADN, región.....	53
Figura 30. Estructura de la proteína Homeobox. HOXC6.....	54
Figura 31. Secuencia de aminoácidos de la proteína Homeobox	54
Figura 32. Estructura del Gen HOXC6/203.	55
Figura 33. Nucleótidos de los exones del gen HOXC6-203	55
Figura 34. Resumen gráfico de la localización de la proteína homeobox.	55
Figura 35. Estructura de la proteína homeobox HOXC6.	56
Figura 36. Secuencia de aminoácidos de la proteína Homeobox.....	56
Figura 37. Secuencias de nucleótidos de exones e intrones del gen abd-A.	62
Figura 38. Estructura de la proteína homeobox abdominal A.....	62
Figura 39. Resumen del gen abd-A.....	63
Figura 40. Ortólogos humanos del gen abd-A.....	63

ÍNDICE DE TABLAS

Tabla 1. Transcritos del gen HOXC6	48
Tabla 2. Transcritos del gen abd-A	50

RESUMEN

Los genes homeóticos (HOX), reguladores maestros, tienen en común una secuencia conservada de 180 pares de bases, llamada caja homeótica, codifican factores de transcripción específicos que desempeñan un papel importante en la morfogénesis y diferenciación celular durante el desarrollo, además se consideran factores de transcripción implicados en el desarrollo de varios tipos de cáncer. El objetivo de este estudio fue analizar las secuencias nucleotídicas de los genes ortólogos homeobox abd-A de *Drosophila melanogaster* y del gen HOXC6 (homeobox C6) en *Homo sapiens*. La investigación tuvo un enfoque cuantitativo, descriptivo, transversal, observacional con aplicación de herramientas bioinformáticas; se analizó mediante Biopython información genética recuperada del National Center for Biotechnology Information (NCBI), además se utilizaron bases de datos como Ensembl, UniProt, y FlyBase. Los resultados de esta investigación determinaron que la longitud y composición nucleotídica de los genes analizados son diferentes con una supremacía del gen abd-A, además el gen HOXC6 se localizó en el cromosoma 12, mientras que el gen abd-A en el cromosoma 3, se determinó la presencia del homeodominio en ambas estructuras genéticas así como el motivo de unión al ADN y la proteína homeobox codificada, en conclusión, el estudio profundo de las estructuras genéticas a partir de herramientas bioinformáticas, permite una visión integral de su composición, nos lleva a determinar mutaciones genéticas y alteraciones en su expresión, permitiendo profundizar el conocimiento biológico y tomar decisiones acertadas para la prevención y tratamiento de diversas patologías.

Palabras clave: gen, HOX, HOXC6, abd-A, *Drosophila melanogaster*.

ABSTRACT

Homeobox genes (HOX), master regulators, have in common a conserved sequence of 180 base pairs, called homeobox, encode specific transcription factors that play an important role in cell morphogenesis and differentiation during development, and are considered to be transcription factors involved in the development of several types of cancer. The aim of this study was to analyze the nucleotide sequences of the orthologous genes homeobox abd-A from *Drosophila melanogaster* and the HOXC6 gene (homeobox C6) in *Homo sapiens*. The research had a quantitative, descriptive, cross-sectional, observational approach with the application of bioinformatics tools; genetic information retrieved from the National Center for Biotechnology Information (NCBI) was analyzed using Biopython, and databases such as Ensembl, UniProt, and FlyBase were also used. The results of this research determined that the length and nucleotide composition of the genes analyzed are different with a supremacy of the abd-A gene, in addition the HOXC6 gene was located on chromosome 12, while the abd-A gene on chromosome 3, the presence of the homeodomain was determined in both genetic structures as well as the DNA binding motif and the encoded homeobox protein, In conclusion, the in-depth study of genetic structures using bioinformatics tools allows a comprehensive view of their composition, leading us to determine genetic mutations and alterations in their expression, allowing us to deepen our biological knowledge and make the right decisions for the prevention and treatment of various pathologies.

Keywords: gene, HOX, HOXC6, abd-A, *Drosophila melanogaster*.

1. INTRODUCCIÓN

La Biología del desarrollo estudia los mecanismos que median la formación de un organismo, desde que una célula es indiferenciada, hasta que alcanza la forma adulta y definitiva. *Drosophila melanogaster* es el organismo modelo más estudiado en la Biología del desarrollo, siendo adoptada como animal de experimentación genética por Thomas Morgan, a principios del siglo XX. (Cortés, 2020)

Entre los genes del desarrollo altamente conservados se encuentran los genes HOX, los mismos que forman parte de una superfamilia familia de genes conocidos por su función en el control del establecimiento del plan corporal. Las proteínas codificadas por los genes HOX, se encuentran en todos los animales superiores y han permanecido virtualmente iguales con el paso del tiempo, éstas controlan la expresión de una variedad de genes dianas. Alteraciones en las funciones de estas proteínas son perjudiciales para el organismo humano, ya que la expresión anormal de estos genes puede afectar procesos biológicos importantes responsables del desarrollo corporal normal, la proliferación celular controlada, la diferenciación celular y la muerte celular programada.

Durante los últimos años, se ha descrito la existencia de numerosos genes que realizan funciones paralelas en la mosca y en el ser humano, lo cual ha permitido identificar genes responsables de rutas metabólicas comunes del desarrollo o genes importantes para la diferenciación celular. (Yamamoto et al., 2014)

Las mutaciones homeóticas determinan que estructuras que normalmente se encuentran en un segmento corporal de la mosca, se reemplacen por estructuras que normalmente se encuentran en otro segmento corporal. De igual manera, mutaciones en los genes HOX humanos pueden causar trastornos genéticos, ocasionados por la desregulación de estos genes, incluso se asocian con varias neoplasias malignas humanas, y se destaca su participación en el desarrollo del cáncer mortal.

Los genes homeóticos, reguladores maestros que dirigen el desarrollo de estructuras o segmentos corporales particulares, están involucrados en los mecanismos moleculares del cáncer ya que pueden actuar como oncogenes o supresores tumorales humanos, una mejor comprensión de sus reguladores aguas arriba y sus dianas aguas abajo permitirían entender la función de los genes HOX en la génesis del cáncer. (Feng et al., 2021)

El conocimiento de la estructura, función y alteraciones de la expresión de los genes HOX mediante el uso

de herramientas bioinformáticas representa un gran desafío, analizar dicha información genética, predecir las proteínas resultantes y determinar mutaciones existentes en la estructura de estos genes, nos permitirían tratar de entender y resolver problemas biológicos.

2. JUSTIFICACIÓN

Los genes homeóticos de *Drosophila*, tienen su contraparte en los mamíferos y son conocidos como genes HOX, juegan un papel fundamental en el desarrollo embrionario, participan en procesos de proliferación celular, diferenciación celular, y son capaces de unirse a secuencias específicas de ADN en genes blanco y regular su expresión (Parra-Soto et al., 2020)

En el ser humano el ordenamiento de los genes dentro de cada complejo HOX, es igual al complejo HOM que se encuentra en *Drosophila*, lo cual indica que se originaron por duplicaciones de un solo complejo primordial, el mismo que ha conservado su organización básica. (López-Romero et al., 2015)

Los genes HOX representan la evidencia de un sistema conservado de control del desarrollo; por ello es actualmente importante conocer la organización estructural y secuencias de ADN de estos genes, puesto que la posible homología que existe tanto en estructura como en funcionalidad entre genes de vertebrados e invertebrados, nos permitirá conocer, y comprender los mecanismos del proceso de desarrollo y también entender las alteraciones resultantes de modificaciones de secuencias genéticas normales.

En la actualidad, la combinación del análisis del genoma de *Drosophila* con la información del genoma humano constituye una herramienta de gran potencial, ya que se podrían identificar mutaciones que causan patologías sobre un sistema de interés en la mosca y posteriormente determinar si los genes que contienen las mencionadas mutaciones tienen genes homólogos en el ser humano y, lo que es más importante, si mutaciones en esos genes son responsables de causar enfermedades humanas. (Yamamoto et al., 2014). *Drosophila* constituye un excelente organismo con el que llevar a cabo rastreos in vivo a gran escala.

Si bien el cáncer tiene una etiología muy compleja, y es un proceso multifactorial, se ha comprobado que las mutaciones en los genes HOX humanos estarían asociados con el desarrollo de mecanismos moleculares del cáncer.

El conocimiento de la estructura normal de los genes HOX, y sus alteraciones permitirá abrir las puertas a la búsqueda de nuevos marcadores moleculares en procesos neoplásicos y el desarrollo de nuevas estrategias terapéuticas para tratar procesos malignos. (López-Romero et al., 2015).

Esto determina la importancia actual y emergente del conocimiento de los genes HOX como un objetivo terapéutico para superar las limitaciones de la terapia actual contra el cáncer, ya que identificadas nuevas moléculas efectivas contra los procesos tumorales, podría ser utilizadas experimentalmente en otros modelos animales o células humanas. (Tolosa, 2015)

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Analizar las secuencias nucleotídicas de los genes ortólogos homeobox abd-A de *Drosophila melanogaster* y el gen HOXC6 (homeobox C6) en *Homo sapiens*.

3.2. OBJETIVOS ESPECÍFICOS

- Analizar las secuencias nucleotídicas del gen abd-A de *Drosophila melanogaster* y del gen HOXC6, de *Homo sapiens*, mediante el uso de herramientas bioinformáticas.
- Determinar el largo de la cadena de los genes abd-A de *Drosophila melanogaster* y del gen HOXC6, de *Homo sapiens*, el porcentaje C-G de las secuencias de ADN del gen HOXC6, de *Homo sapiens*.
- Determinar la región cromosómica y el número de exones/intrones correspondientes al gen abd-A de *Drosophila melanogaster* y del gen HOXC6, de *Homo sapiens*.
- Identificar el número de transcritos codificantes y las secuencias de aminoácidos codificados por los genes abd-A de *Drosophila melanogaster* y del gen HOXC6, de *Homo sapiens*.

4. PLANTEAMIENTO DEL PROBLEMA

Se han descubierto genes con caja homeótica en todas las especies estudiadas incluida la especie humana, éstos genes están ordenados en los cromosomas de modo semejante a los de la *Drosophila* lo cual supone que los mecanismos genéticos responsables del desarrollo del plan corporal son similares en los organismos multicelulares; en los embriones de los vertebrados se expresan varios genes con caja homeótica, que presentan una organización metamérica análoga a la de los segmentos de *Drosophila* (Robertis, 2014).

Se han estudiado los genes HOX durante los últimos años en escenarios genéticos diferentes, tanto en la embriogénesis como en el individuo adulto, lo que contribuye a pensar que continúan desempeñando un papel en la identidad celular; alteraciones derivadas de mutaciones en estos genes han sido relacionadas con diversas patologías como la aniridia, simpolidactilia (HOXD13) y algunos tipos de tumores como el rabdomiosarcoma alveolar o tumores intestinales. (Perera et al., 2006)

Se han identificado genes HOX que se activan o reprimen en el desarrollo de los diferentes mecanismos moleculares del cáncer, de igual manera su expresión anormal es crucial en una variedad de procesos celulares malignos, como el cáncer colorrectal, cáncer de mama, próstata, glioblastoma, pulmón, nasofaríngeo, gástrico y ovárico. (Tang et al., 2019)

Las células normales proliferan en respuesta a inductores endócrinos o parácrinos necesarios para su crecimiento, pero las células neoplásicas muestran una menor necesidad de factores de crecimiento, bajas señales de proliferación exógena y varios genes HOX están desregulados en varios tipos de cáncer desempeñando un rol crítico en la proliferación tumoral. Los genes HOXC4, HOXB2, HOXB3, HOXC6 y HOXA13, tienen propiedades supresoras de tumores o promotoras de tumores, según el tipo de tumor en el que se expresen, por ejemplo, en condiciones normales el gen HOXC6 se expresa en líneas celulares de cáncer de próstata reduciendo considerablemente el crecimiento de células tumorales. (Paco et al., 2020)

La transcripción del gen HOXC6 se inicia a partir de dos promotores y genera dos proteínas que comparten el mismo dominio de unión al ADN, pero albergan una región N-terminal distinta, los dos productos HOXC6 pueden activar o reprimir la transcripción, según el contexto celular. (Tang et al., 2019)

Estudios actuales revelan que HOXC6 podría funcionar como un oncogen, puesto que sus versiones HOXC6-1 y HOXC6-2 se sobreexpresaron en tejidos cancerosos gástricos, promoviendo la migración, invasión y proliferación celular anormal. De igual forma se detectó una expresión disminuida del gen HOXC6-2 en tipos bien diferenciados de cáncer gástrico, concluyendo que HOXC6 funciona como un oncogén tumoral. Los oncogenes constituyen la versión mutada o alterada de los protooncogenes y actúan por dos mecanismos básicos: producción de proteínas anormales, o producción de cantidades mayores de proteínas, es decir actúan por ganancia de función.(Lin et al., 2020)

La sobreexpresión del gen HOXC6 en el carcinoma hepático, determinó su proliferación celular descontrolada, en tanto que la expresión negativa o disminuida de HOXC6 mediada por ARNip, que utiliza ARN pequeño de interferencia (siARN) para un potente silenciamiento de genes, además de inhibir la proliferación celular y la migración, también aumentó la quimiosensibilidad al 5-FU, (fluorouracilo) medicamento de quimioterapia citotóxico usado para tratar el cáncer, lo cual sería de gran beneficio en tratamientos de pacientes oncológicos. (Tang et al., 2019). Estas investigaciones han logrado sugerir que HOXC6 podría estar implicado en la iniciación y progresión de procesos neoplásicos.

De igual manera el gen HOXC6 está altamente expresado en células no pequeñas de cáncer pulmonar (NSCLC), promueve la proliferación, migración e invasión de células tumorales. La expresión fenotípica de HOXC6 podría estar relacionada con genes involucrados en los mecanismos moleculares de progresión de tumores, por ello, se convertiría en un marcador biomolecular de interés potencial para el diagnóstico y tratamiento del NSCLC. (Yang et al., 2019)

El glioblastoma (GBM) ocupa alrededor del 30 % de todos los tumores del cerebro y del sistema nervioso central (SNC) y casi el 80 % de los tumores malignos del cerebro y del SNC. Se encontró que HOXC6 se sobreexpresaba en tejidos y líneas celulares de GBM, y una expresión alta de HOXC6 predecía un mal pronóstico en pacientes con esta enfermedad. Usando modelos de células GBM de agotamiento y sobreexpresión de HOXC6, se demostró que HOXC6 promovió la proliferación y migración de células tumorales, principalmente a través de la vía de la proteína quinasa activada por mitógeno (MAPK), constituyéndose en un biomarcador pronóstico y un objetivo molecular prometedor para GBM.(Shenoy et al., 2022)

El gen HOXC6 regula la expresión del gen antiapoptótico BCL2, inhibiendo la vía apoptótica normal, es decir

sus niveles elevados de expresión determinarían un mecanismo positivo de la antiapoptosis. Sin embargo, la caída de HOXC6 da como resultado la regulación negativa de BCL2 y por consiguiente la activación de la vía de apoptosis intrínseca. (Moon et al., 2012)

Numerosos estudios moleculares y clínicos han demostrado que cualquier desregulación en los genes del grupo HOX ocasionaría un desequilibrio en las vías celulares, lo que favorecería a las células cancerosas hacia su progresión metastásica. Dentro de estos tenemos que la mutación sin sentido HOXD4 se observó en pacientes con leucemia linfoblástica aguda. (Quinonez & Innis, 2014) Las mutaciones hereditarias en el gen HOXB13 contribuyen significativamente al aumento del riesgo de cáncer de próstata, el gen HOXA9 es un oncogen implicado en el desarrollo de la leucemia, sin embargo, inhibe el crecimiento tumoral y la metástasis en la glándula mamaria (Gilbert et al., 2010).

Si es que las modificaciones genéticas y epigenéticas en los genes HOX son determinantes en el desarrollo de metástasis, la investigación debería tratar de identificar características moleculares de los genes HOX en su nivel transcripcional. (Shenoy et al., 2022)

El conocimiento y análisis del perfil genético de los grupos de genes HOX permitiría que estos sean utilizados como un biomarcador molecular en la temprana detección de procesos tumorales, determinando que las técnicas moleculares de detección temprana del cáncer sean ventajosas sobre métodos convencionales de detección de tumores. (Shenoy et al., 2022).

5. FUNDAMENTACIÓN TEÓRICA.

5.1 ANTECEDENTES.

El estudio de la Biología del Desarrollo utiliza organismos modelo por su fácil manipulación, homología filogenética con el organismo humano y el gran conocimiento acumulado durante años de estudio.

Drosophila melanogaster fue escogida como el modelo experimental ideal para el estudio genético y molecular de una gran variedad de procesos biológicos, por su fácil manejo, un genoma con pocas secuencias repetidas y 4 pares de cromosomas, bajo costo de mantenimiento y un ciclo de vida rápido si se compara con otros organismos pluricelulares usados en investigación. Se determinó que *Drosophila melanogaster* en la etapa larvaria

presenta estructuras que darán origen a las patas o tienen capacidad regenerativa, pero en la etapa adulta, las patas ya no pueden regenerar. (Uribe, 2019)

Drosophila melanogaster conocida como la “mosca del vinagre” es el modelo animal que marcó áreas de la biología evolutiva contemporánea. Descrita en el año 1830 por Johann Wilhelm Meigen: su muy breve descripción de la especie fue traducida como: “Cabeza, tórax y patas amarillas; vientre negro, halterios blancos, alas incoloras.” (Keller, 2007)

En 1974 Richard Lewontin, utilizó “moscas del vinagre” con el objeto de determinar la variabilidad genética presente en diferentes poblaciones, para ello utilizó una técnica de separación de proteínas a partir de su tamaño y carga eléctrica, (electroforesis de enzimas) demostrando que hay más variación genética de lo que teóricamente se podría esperar. (González-Astorga, 2021)

La primera mutación homeótica fue descrita por Bridges a principios del siglo pasado. Thomas Morgan encontró mutaciones en *Drosophila* donde la parte anterior del tercer segmento torácico había sido reemplazada por la parte anterior del segundo segmento torácico. Lewis, a finales del año 70 encontró el gen responsable del fenotipo bithorax, a partir de lo cual se aislaron proteínas con el homeodominio en su estructura, algunas de las cuales se encuentran relacionadas con mutaciones en los genes homeóticos. (Perera et al., 2006)

El grupo de Hogness en Stanford marcó la nueva era en la Biología al lograr la clonación del primer gen homeótico, el gen Ultrabithorax, posteriormente se clonaron otros genes homeóticos como Antennapedia, abdominal-A, abdominal-B y posteriormente el resto, logrando identificar la naturaleza de los productos génicos de los genes HOX. (Morata, 2001)

Se logró determinar que los genes y proteínas de *Drosophila melanogaster* comparten homología con genes y proteínas de los vertebrados, lo cual permitió a la genética y biología del desarrollo identificar procesos que se desarrollan en la mosca, y que están conservados en el desarrollo de organismos superiores, dando paso al estudio de genes involucrados en enfermedades humanas. (Betti, 2016)

Las bases de datos de información genómica, revelaron que la mosca *Drosophila* y el ser humano comparten más del 60% de los genes, y de este porcentaje un 75% causan enfermedades en los humanos, adicionalmente poseen más del 90% de genes relacionados con la génesis del cáncer en los seres humanos. (Yamamoto et al., 2014)

5.2. DEFINICIÓN, CLASIFICACIÓN Y ESTRUCTURA DE LOS GENES.

5.2.1. GENOMA.

El genoma constituye, las características genéticas y hereditarias, que determinan las diferencias existentes entre los seres vivos. Su estudio, ha permitido desarrollar importantes avances y mejorar la calidad de vida de los seres humanos y de otras especies en general. El mismo se encuentra en el ADN o ácido desoxirribonucleico y se caracteriza por ser único e irrepetible; el genoma humano contiene 6×10^9 pares de bases.

5.2.2. ESTRUCTURA DE LOS GENES.

El gen constituye la secuencia de ADN cromosómico que contiene la información requerida para fabricar una molécula de ARN y en el caso del ARN mensajero una proteína, ocupa un sitio particular del cromosoma denominado locus, comprenden alrededor del 10% del ADN nuclear; los genes constituyen entidades biológicas mediante las cuales los rasgos físicos se transmiten de padres a hijos. (Robertis 2014).

Los genes poseen segmentos superfluos de ADN llamados intrones, intercalados entre segmentos útiles llamados exones. La molécula completa de ARN, que surge de la transcripción se llama transcripto primario, se transforman en ARNm funcional antes de salir del núcleo, en esta molécula se copian no solo los exones sino también los intrones que forman parte del segmento codificador. El gen posee partes funcionales que le ayudan a regular su actividad:

- **Promotor:** sector genético que señala a partir de que nucleótido el gen debe comenzar a transcribirse e inicia la transcripción, se encuentra localizado cerca del extremo 5' del segmento codificador.
- **Regulador:** sector genético que señala cuando y cuantas veces el gen debe comenzar a transcribirse, localizado lejos del segmento codificador, comprende secuencias amplificadoras y secuencias inhibidoras.
- **Señal de terminación:** marca la conclusión de la síntesis de ARN.

5.2.3. ESTRUCTURA DEL GEN QUE CODIFICAN ARN MENSAJERO

Generalmente el promotor tiene dos elementos. La combinación más común comprende las secuencias llamadas cajas TATA y CAAT, ubicadas cerca del codificador. La caja TATA (TATAAAA), se encuentra a 25

nucleótidos del primer nucleótido del segmento codificador. La caja CAAT(GGCCAATCT) se encuentra en el mismo lado, aproximadamente 75 nucleótidos, del segmento codificador, es decir 50 nucleótidos de la secuencia TATA.

Los reguladores, tanto amplificadores e inhibidores también se encuentran corriente arriba respecto del extremo 5' del segmento codificador, a miles de nucleótidos de distancia del segmento codificador.

En el segmento codificador, los segmentos de ADN utilizables y los segmentos no funcionales se alternan, igual que en los transcritos primarios se denominan exones e intrones. No se puede identificar la secuencia de terminación, sin embargo, en un sector previo a ella encontramos la secuencia AATAA, conocida como señal de poliadenilación necesaria para completar la síntesis del transcrito primario.

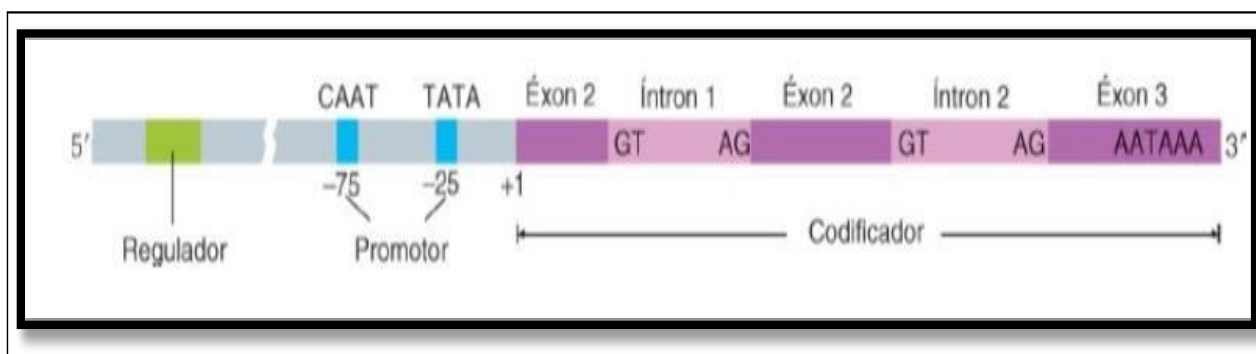


Figura 1. Organización del gen que codifica al ARNm en los genomas eucarióticos. Se observa la estructura del gen con sus diferentes elementos: promotor, regulador, codificador. (Tomado de *Biología celular y molecular*, por E. Robertis, 2014, El Ateneo).

5.3 CLASIFICACIÓN DE LOS GENES.

Dentro del escenario de investigación biológica, representa un desafío interesante, conocer las características funcionales y evolutivas de los genes, debido a la aparición frecuente y constante de datos de nuevas secuencias genéticas. Por ello es importante una clasificación evolutiva de los genes basada en la ortología y paralogía que permitan describir relaciones de homología entre los mismos.

5.3.1. HOMOLOGÍA.

Las secuencias que están relacionadas evolutivamente son conocidas como homólogas, son aquellas que, aunque difieren en los detalles derivan del mismo plan corporal. Se puede comparar dos genes, realizando alineamientos de sus secuencias, lo que nos permitiría identificar si poseen suficiente similitud como para determinar

la existencia de homología entre ellos. Es importante diferenciar entre similitud y homología, la primera es independiente de la evolución, mientras que la homología es una característica que describe la historia evolutiva. (Webber & Ponting, 2004)

5.3.2. GENES ORTÓLOGOS.

Dos genes son ortólogos, cuando son homólogos de diferentes especies y derivan de un gen en el último ancestro común, presentando las mismas funciones biológicas. Los genes ortólogos son secuencias homólogas derivadas de un evento de especiación a partir de una sola secuencia ancestral en el último ancestro común de las especies que se comparan. Los ortólogos cumplen funciones semejantes en especies que están estrechamente relacionadas. (Fernández, 2012.)

Los genes ortólogos están relacionados evolutivamente y su caracterización está comúnmente relacionada con los árboles filogenéticos. Es decir, demanda la obtención de homólogos, la generación de múltiples alineamientos correctos para cada grupo de dominios homólogos, la construcción de un árbol filogenético por cada grupo y posteriormente la extracción de ortólogos. Mientras que los genes parálogos son copias duplicadas dentro de un genoma que pueden ser producidos por el fenómeno de poliploidización o duplicaciones en tándem. (Solís, 2008).

5.4. GENES HOX.

Los genes HOX forman parte de la familia de genes homeobox, los mismos que poseen una secuencia de 180 pb que se llama "homeobox". Esta secuencia fue descubierta y descrita por primera vez en el genoma de *Drosophila melanogaster*, a partir de ello también se encontró en organismos pluricelulares, como plantas, hongos, esponjas, al igual que en los vertebrados, incluido el ser humano. (Verena, 2017). Las secuencias de homeobox descubiertas y posteriormente analizadas fueron muy parecidas a las de *Drosophila*, determinándose que el "homeobox" codifica una secuencia de 60-aminoácidos de largo, segmento polipeptídico que fue denominado como "homeodominio". (Gehring et al. 1994).

Los genes HOX son pequeños, están formados de dos exones separados por un intrón, tienen una longitud genómica de 5-10 kb en comparación con la media para genes codificadores de proteínas de 27 kb; el exón está localizado cerca del extremo 3' y contiene el homeobox de 180 pb que codifica el homeodominio de la proteína HOX.

Estas proteínas homeobox actuarían como factores de transcripción específicos con el homeodominio como motivo de unión al ADN. (Grier et al. 2005)

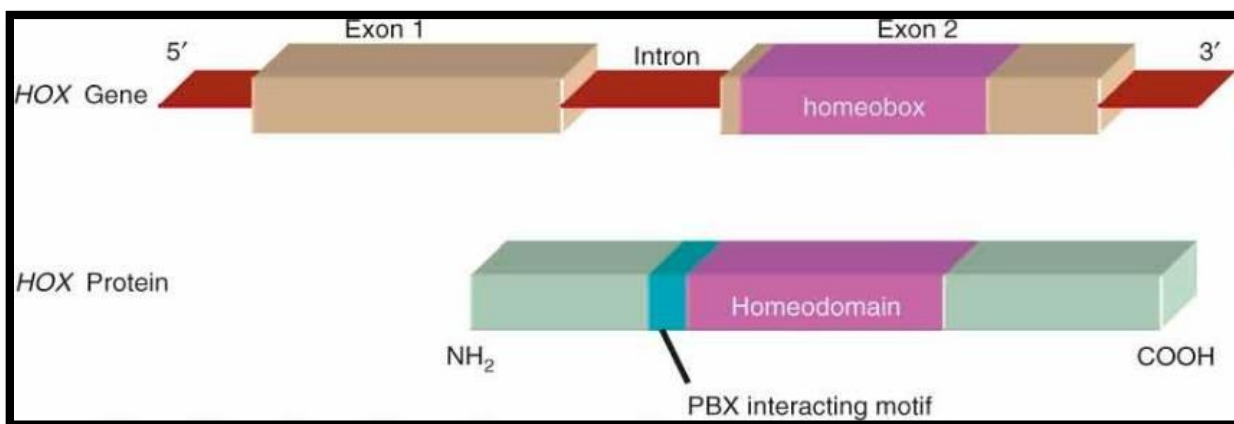


Figura 2. Estructura del gen HOX y la proteína HOX, que codifica al ARNm en los genomas eucarióticos. Se observa el gen HOX con sus dos exones, separados por un intrón; además se representa la proteína HOX con el Homeodominio. (Tomado de *The pathophysiology of HOX genes and their role in cancer*. (p.154-171), por Grier et al., 2005, *The Journal of Pathology*, 205).

El homeodominio posee un motivo de unión al ADN de tipo hélice-giro-hélice. La mayor parte de las proteínas HOX, contienen una secuencia de seis aminoácidos cerca del homeodominio, el mismo que funciona como sitio de unión para cofactores que incluyen proteínas TALE como PBX y MEIS. Sin embargo, algunas proteínas HOX carecen de este hexapéptido. (Erselius, 1990)

En los seres humanos se han aislado 39 genes HOX, los mismos que están clasificados en cuatro grupos diferentes: HOXA, HOXB, HOXC y HOXD. Estos genes HOX son importantes en varios procesos biológicos como señalización mediada por receptores, muerte celular programada, adicionalmente las proteínas HOX están implicadas y activas en mecanismos moleculares importantes como replicación y reparación del ADN, traducción del ARNm y la degradación de proteínas mediante ubiquitinas, revelándose actividades no transcripcionales de las proteínas HOX. (Rezsosazy et al., 2015)

5.5. DESARROLLO Y ORGANIZACIÓN GENÉTICA DE *DROSOPHILA MELANOGASTER*

Drosophila es un pequeño insecto de 1 mm de largo, formado por un millón de células, con un ciclo biológico muy corto de 9-10 días, fácil de cultivar en el laboratorio. El modelo tridimensional que marca el establecimiento del plan corporal de *Drosophila*, comprende la expresión génica jerárquica, mediante la cual grupos genéticos rectores controlan la expresión de genes subordinados siguiendo una secuencia escalonada de eventos.

La mosca *Drosophila* se desarrolla, a partir de una larva, como producto de la fecundación, luego de formarse la célula huevo y atravesar el período embrionario. La larva está formada por la sucesión de segmentos: uno cefálico, tres torácicos y ocho abdominales, que le otorgan una marcada polaridad espacial, ya que tan pronto como aparecen se configuran los ejes cefalocaudal, dorsoventral y medio lateral del cuerpo larvario.

La larva se convierte en mosca a partir de grupos celulares llamados discos imaginales, los mismos que dan origen a las estructuras externas de la mosca, por ejemplo, de un par de discos surgen las antenas y los ojos, de otro las alas y parte del tórax, de otros las estructuras que originan el abdomen; las partes mencionadas correctamente ensambladas, forman el cuerpo adulto llamado imago.

El plan corporal mencionado, está regulado por una compleja red de genes reguladores, que ejercen sus funciones apenas se forma la célula huevo. Primero actúan los genes de la polaridad de la célula huevo, luego actúan los genes segmentarios que son de tres clases: genes de hendidura, genes de la regla par y genes de la polaridad de los segmentos, estos se expresan, en el orden mencionado, activados por señales posicionales previamente establecidas. (Robertis, 2014).

Los genes de hendidura dividen al embrión en regiones grandes, vistas como bandas transversales en el eje anteroposterior, y estos activan a su vez la expresión de los siguientes genes en cascada, los genes de la regla par y los genes de la polaridad de los segmentos. Los genes de la regla par se expresan de manera alterna, y controlan segmentos pares es decir uno de cada dos, determinando la naturaleza de éstos, la acción de los genes “gap” y de la regla par activa la expresión de los genes de polaridad de segmento, que definen los límites y la polaridad de los mismos. (Uribe, 2019). Finalmente se expresan los genes homeóticos, activados por productos de genes que actuaron anteriormente, originan los discos imaginales y las estructuras externas de la mosca adulta.

Todos los genes mencionados participan en la construcción del plan corporal, mediante una sucesión de activaciones en cascada, que va marcando la diferenciación correspondiente y al mismo tiempo preparando al gen próximo en activarse, siguiendo una dirección céfalo-caudal; por otra parte, el producto de cada gen influye sobre los genes que se expresaron anteriormente, imprimiendo en las células un determinado e indeleble valor posicional (Robertis, 2014).

Los genes involucrados en la formación del plan corporal son los llamados genes rectores, ya que controlan

la expresión de genes que actúan siguiendo un orden jerárquico, estos codifican factores de transcripción específicos, que tienen en común un segmento de 60 aminoácidos llamado homeodominio, codificados por genes homeóticos que tienen una secuencia de 180 pares de nucleótidos llamada caja homeótica por ello su nombre.

5.5.1. GENES HOMEÓTICOS DE *DROSOPHILA MELANOGASTER*.

Gracias a importantes investigaciones realizadas sobre la familia de genes HOX, se han descubierto, varios hallazgos fundamentales sobre las características, funciones y los mecanismos de regulación de esta familia de genes conservados en el desarrollo, la enfermedad y la evolución.

Los genes Homeobox (HOX), son factores de transcripción altamente conservados a través de la evolución, se originan mediante ciclos repetidos de duplicaciones en tándem a partir de un gen ancestral denominado ProtoHOX, mutación necesaria para generar un nuevo límite de expresión a lo largo del eje embrionario y adquisición de una nueva función de patrón HOX. (Gaunt & Gaunt, 2016).

Debido al importante papel que cumplen estos genes como reguladores generales del desarrollo, es necesario conocer cuántos de estos genes existen en el genoma de *Drosophila*. Se ha descubierto que el conjunto de genes homeóticos primarios, llamado complejo HOX, consta de ocho genes relacionados con el desarrollo de los segmentos cefálicos, torácicos y abdominales. (Morata, 2001). Estos genes no están distribuidos al azar en los cromosomas, se encuentran agrupados en dos complejos génicos: el Complejo Antennapedia (C-ANT), que incluye desde la región proximal a distal, y el Complejo Bithorax (C-BX), que incluye a Ultrabithorax (Ubx), abdominal-A (abd-A) y abdominal-B. (Del Saz Soler, 2018). El complejo de genes HOX, localizado en el cromosoma 3, comprenden el ANT-C que contiene 5 genes y el BX-C que contiene 3 genes HOX. (Bofarull, 2005).

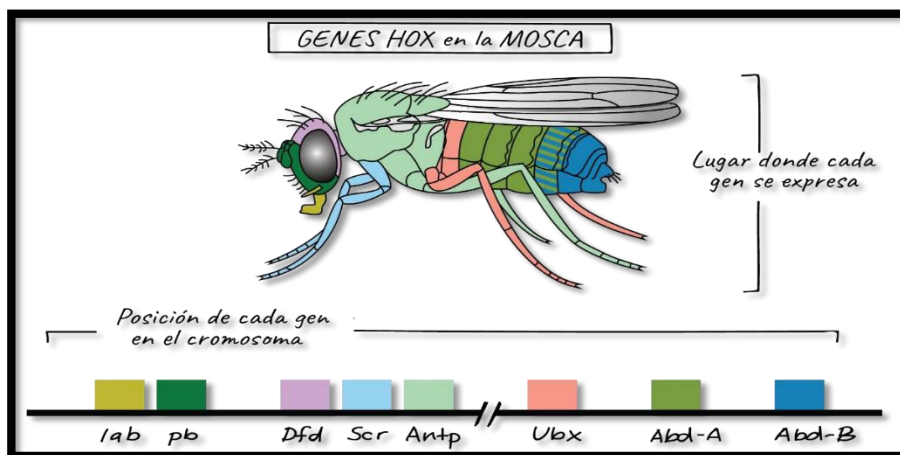


Figura 3. Genes HOX en la mosca *Drosophila melanogaster*. Posición de los genes HOX a lo largo del cromosoma y la correspondiente zona del cuerpo de *Drosophila melanogaster* en donde se expresan. (Tomado de Hoxgenesoffruitfly.svg, por Philip, 2017, imagen de dominio público) (<https://commo.ns.wikimedia.org/wiki/File:Hoxgenesoffruitfly.svg>)

El complejo bitórax (BX-C) marca la identidad de la porción del cuerpo de *Drosophila*, contiene 3 genes denominados, Ultrabithorax (Ubx), abdominal-A (abd-A) y abdominal-B (abd-B) que contienen un homeobox. Los genes abd-A son responsables de la identidad específica de los parasegmentos 7 a 9 y también son necesarios para el correcto desarrollo de los parasegmentos 10 a 13. Los parasegmentos son surcos que marcan las áreas corporales mediante la acción de genes de la regla par. (Busturia et al., 1989)

El análisis molecular del complejo BX determinó que contiene tres unidades de transcripción correspondientes al 2% de la secuencia total. El 98% restante está formado por varias regiones reguladoras que otorgan la especificidad espacial y temporal de la expresión de los productos proteicos (González, 2007)

El gen abd-A de la familia *Drosophila*, presenta una región codificadora de proteínas y dos regiones reguladoras en cis. La región reguladora 1, 3' a la unidad de transcripción, contiene elementos reguladores positivos y negativos. La región reguladora 2, 5' a la unidad de transcripción, establece el nivel correcto de actividad abd-A en los metámeros abdominales. (Busturia et al 1989.) De manera general, las regiones reguladoras-cis están potencialmente estructuradas y muestran una compleja organización molecular formada por aisladores, silenciadores, amplificadores y promotores basales (Levine, 2003).

Si consideramos que los genes homeóticos son responsables de otorgar la identidad de segmentos o determinar el desarrollo del plan corporal, es indudable que, cuando en estos genes ocurren mutaciones, se inactiven

o se expresan en lugares inusuales causando que segmentos del cuerpo adquieran nuevas o distintas identidades.

Habitualmente el gen homeótico *Antennapedia*, se expresa en el segundo segmento del tórax de la mosca, desde que la mosca es un embrión diminuto y se mantiene en la mosca adulta; el gen actúa como un regulador maestro, poniendo en marcha el programa genético que origina el segundo par de patas de la mosca y otras estructuras específicas del segmento. Sin embargo, una mutación genética de este gen podría ser responsable de que su expresión, se extienda a la cabeza de la mosca, creciendo patas en la cabeza de la mosca en lugar de antenas. Por otra parte, la mutación del gen *Ultrabithorax*, determinará que pequeñas estructuras llamadas halterios responsables de ayudar a mantener el equilibrio de la mosca, se conviertan en un segundo conjunto de alas ubicadas detrás del par normal. (Del Saz Soler, 2018)

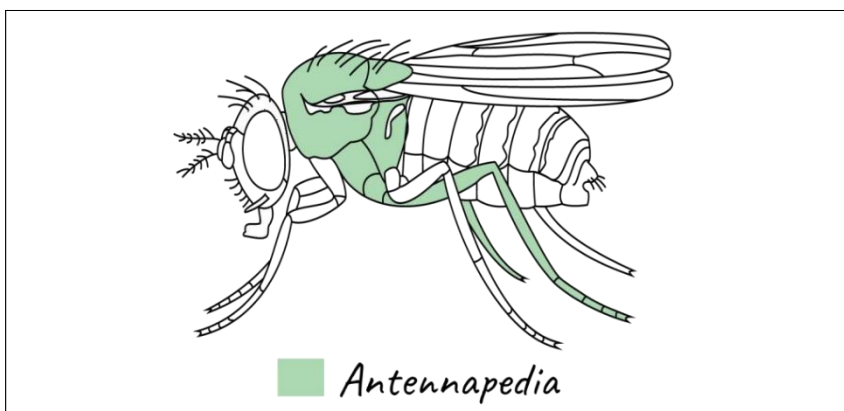


Figura 4. Complejo *Antennapedia* de *Drosophila melanogaster*. Ubicación del complejo *Antennapedia*. (Tomado de *Hox genes of fruitfly*. Svg por Philip, 2017, imagen de dominio público. (<https://commons.wikimedia.org/wiki/File:Hoxgenesoffruitfly.svg>))

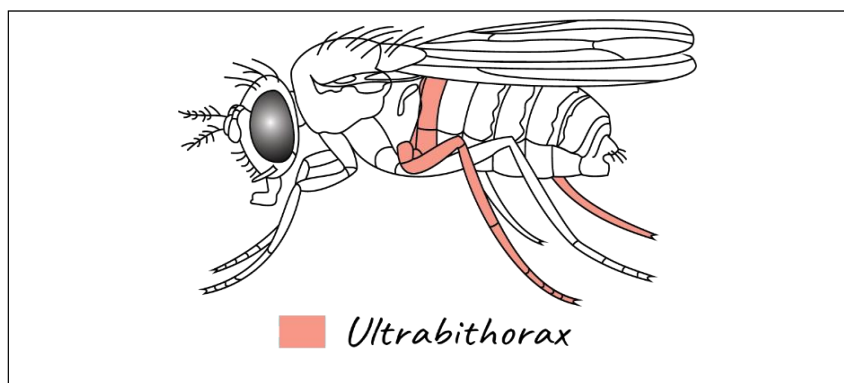


Figura 5. Complejo *Ultrabithorax* de *Drosophila melanogaster*. Ubicación del complejo *Ultrabithorax* de *Drosophila melanogaster*. Tomado de *Hoxgenesoffruitfly.svg*, Philip, 2017, dominio público. (<https://commons.wikimedia.org/wiki/File:Hoxgenesoffruitfly.svg>))

5.5.2. GENES HOX EN EL SER HUMANO.

Los genes HOX se encuentran presentes en todos los animales estudiados donde determinan el desarrollo de cada parte del cuerpo en el lugar apropiado del mismo. Se ha confirmado que un gen HOX humano puede sustituir a su equivalente de la mosca, obviamente el desarrollo de un órgano es mucho más complejo y es el resultado de muchas interacciones entre genes. (Castagnino, 2004)

La superfamilia de genes HOX, codifican para una gran variedad de homeoproteínas que actúan como factores de transcripción. El homeodominio de los genes HOX puede unirse a secuencias específicas de ADN en sus genes blanco y regular su expresión. (López-Romero et al., 2015). Estas proteínas, se originan o están codificadas por una secuencia génica común de 180 pares de bases que presentan un dominio de unión de tipo hélice-vuelta-hélice, conformado por 61 aminoácidos que permite su unión al ADN; determinando la activación o represión de genes. (Saha et al., 2017)

En el genoma humano, de acuerdo a la semejanza de secuencia y la similitud de ubicación en el cromosoma, los 39 genes pertenecientes a la familia HOX están distribuidos en cuatro grupos: HOXA, HOXB, HOXC y HOXD, ubicados en distintos cromosomas: 7p15, 17q21.2, 12q13, 2q31, respectivamente. (Paco, 2020). Cada grupo presenta de 9 a 11 genes que se alinean en 13 grupos parálogos, (HOX1 a HOX13) según la similitud de secuencia y la posición dentro del grupo. (Rux & Wellik, 2017).

Además de activarse durante el desarrollo embrionario, los genes HOX se expresan en diversos tejidos adultos, donde estarían implicados en una variedad de rutas biológicas que incluye la homeostasis, diferenciación celular y el mantenimiento de la función orgánica; sin embargo, sus funciones no han sido totalmente definidas y aún queda mucho por investigar.

En relación a los estudios realizados en dichos genes, se ha determinado que existen diferencias en los patrones de expresión en cáncer de mama, próstata, hígado, cabeza y cuello, leucemia, y otros. Sin embargo, el mecanismo molecular exacto por el cual estos cambios ocurren, no se conocen con precisión (Platais et al. 2015). En un estudio de cáncer de cuello uterino se determinó la expresión alterada de 4 grupos HOX en comparación con muestras histopatológicamente normales. (Saha et al., 2017)

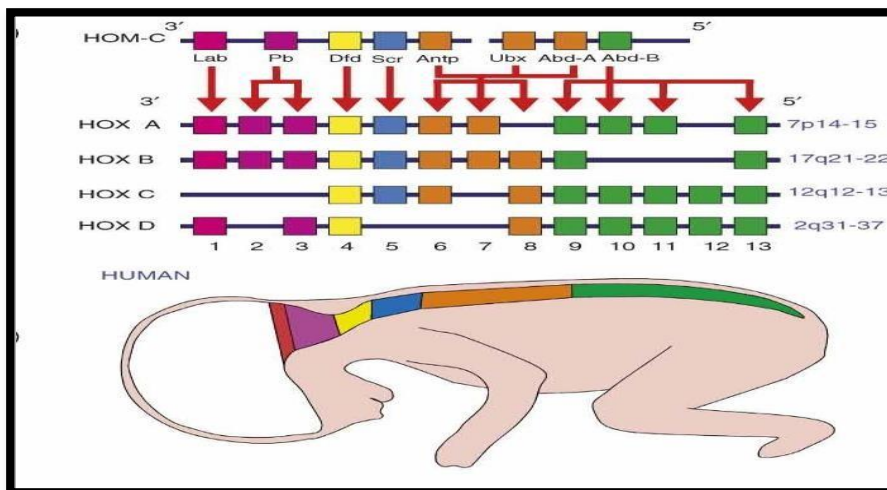


Figura 6. Genes Hox en el ser humano. Origen y agrupación, espacio-temporal de los genes HOX en el ser humano. (Tomado de *The pathophysiology of HOX genes and their role in cancer*. (p.154-171), por Grier et al., 2005, *The Journal of Pathology*, 205)

5.5.3. GEN HOXC6 EN *HOMO SAPIENS*

Este gen pertenece a la familia homeobox, codifica una familia altamente conservada de factores de transcripción que desempeñan un papel importante en la morfogénesis de todos los organismos multicelulares. Los mamíferos poseen cuatro grupos de genes homeobox similares, HOXA, HOXB, HOXC y HOXD, que se encuentran en diferentes cromosomas y constan de 9 a 11 genes dispuestos en tándem.

HOXC6, es uno de varios genes HOXC ubicados en un grupo en el cromosoma 12q13.13. Tres genes, HOXC5, HOXC4 y HOXC6, comparten un exón no codificante 5'. Las transcripciones comprenden el exón compartido empalmado con los exones específicos del gen, o pueden incluir solo los exones específicos del gen. Adicionalmente, se han encontrado variantes de transcripción empalmadas que codifican diferentes isoformas para HOXC6. (NCBI). Codifica la proteína Homeobox HOXC6 Homeobox HOXC6.

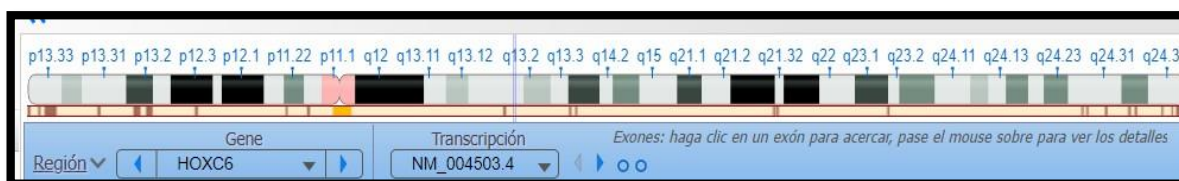


Figura 7. Gen HOXC6 en el ser humano. Ubicación del gen HOXC6 dentro del cromosoma 12 en el ser humano. (Tomado de Biblioteca Nacional de Medicina (EE. UU), Centro Nacional de Información Biotecnológica; (NCBI) 2004 – [citado 2022-12-29]. Disponible en: <https://www.ncbi.nlm.nih.gov/gene>)

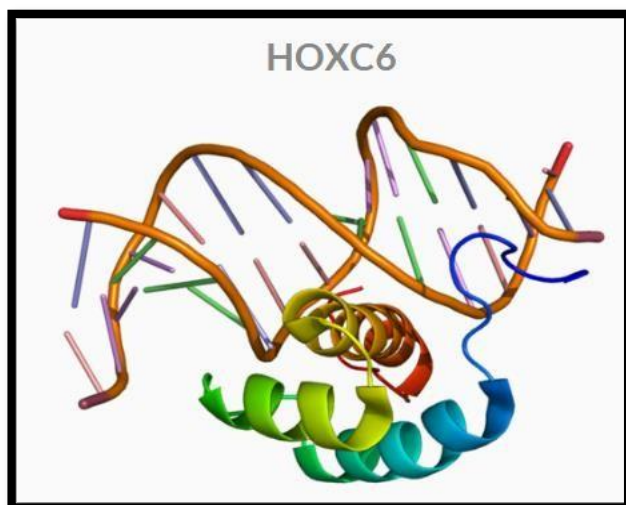


Figura 8. Proteína HOXC6. Estructura tridimensional de la proteína HOXC6 codificada por el gen HOXC6 en el ser humano. (Tomado de proteína HOXC6 PDB 1ahd.png, (2015).

5.5.4. DESREGULACIÓN DE GENES HOX Y CÁNCER.

El cáncer aumenta día a día su frecuencia que se calcula que millones de personas mueren a causa de esta enfermedad, constituye la segunda causa de muerte en el mundo, después de las enfermedades cardiovasculares, sin embargo, la invasión y metástasis corresponden a la primera causa de muerte relacionada con el cáncer. (Paço et al., 2020). Se caracteriza por un crecimiento celular incontrolado, cuya base es genética, las células se multiplican, aunque el organismo no precise de ellas, surge luego de sucesivos estados precancerosos por los cuales atraviesan generaciones de células. (Soliz, 2011)

Los genes HOX en procesos tumorales muestran un patrón de desregulación temporoespacial diferente al de los tejidos y órganos normales. Además, el aumento patológico del nivel de expresión de los genes HOX en tejidos específicos, desencadena el mecanismo que permite explicar el relevante papel en el cáncer.

En condiciones normales los genes HOX se comportan como genes supresores de tumores en varios tipos de cáncer; su expresión anormal puede afectar la proliferación celular, la diferenciación, la apoptosis, la motilidad, la angiogénesis, la autofagia y la señalización del receptor celular. En general la alteración del nivel de expresión relacionado con la región cromosómica puede conducir a enfermedades y defectos del desarrollo, especialmente cáncer humano. (Feng et al., 2021)

La expresión anormal de los genes HOX está típicamente asociada con oncogénesis y puede variar de

acuerdo con el tipo histológico y la etapa de progresión del cáncer, incluida la metástasis. (López-Romero et al., 2015)

Diferentes tipos de cáncer incluidos cáncer de pulmón, próstata, mama, colon, vejiga y tiroides, así como cáncer de ovario, muestran diferentes cambios en la expresión de los HOX, sea por un incremento de expresión cuando comúnmente esta es baja en los tejidos normales, o como una expresión de novo cuando comúnmente no existe en los tejidos sanos. Los genes HOX son responsables de la diferenciación celular normal y el mantenimiento de células diferenciadas y su desregulación es importante en la transformación celular o en la progresión tumoral. (López-Romero et al., 2015)

La expresión anormal de los genes del grupo HOX constituyen una parte integral de mecanismos reguladores relacionados con la homeostasis del tejido adulto normal, el mantenimiento y la activación del proceso de autorrenovación de células madre, fundamental para la transformación maligna. La expresión desregulada de genes pertenecientes a la familia HOXC, particularmente HOXC4 a HOXC9, HOXC11 y HOXC13, muestran un aumento predominante en la mayoría de los tipos de tumores sólidos, excepto en cánceres de ovario. (Saha et al., 2017)

Mutaciones en los genes HOX, están relacionadas con diferentes tipos de cáncer, y se han relacionado con el inicio y la progresión de procesos tumorales ya que estos pueden activar o inhibir diferentes procesos celulares e incluso exhibir comportamientos antagónicos en diferentes tejidos. (Paco et al., 2020)

Una característica distintiva del cáncer influenciado por los genes HOX, es la inflamación, ya que las células cancerosas estimularían la proliferación celular anormal; el movimiento celular que promueve los procesos de migración e invasión dependen del suministro de enzimas que degradan la matriz, como las metaloproteinasas, grupo de enzimas que pueden descomponer proteínas, como el colágeno, que normalmente se encuentran en los espacios intercelulares. La división celular descontrolada que marca el inicio del cáncer implica no solo la alteración de la proliferación celular y la resistencia a la apoptosis, sino también cambios en el metabolismo energético para impulsar el crecimiento y la división celular. (Paço et al., 2020)

También es importante considerar que todo el locus HOXC está asociado a una gran cantidad de tipos de cáncer tales como sarcoma, glioblastomas, pulmón, vejiga, melanomas y carcinomas transicionales (Cantile et al., 2009) Adicionalmente, se ha identificado expresión de novo en líneas celulares de cáncer de colon, así como una tasa de expresión mayor asociada a tumores primarios. El clúster C es considerado uno de los más activos en neoplasias

de tracto urogenital, sobre todo aquellos genes que se encuentran en las regiones próximas a 5' (García, 2018).

HOXC6 juega un papel fundamental en el desarrollo del linfoma, estimula la proliferación de células de cáncer gástrico al actuar como un oncogén; sin embargo, la disminución de su expresión llevaría a una menor proliferación de ciertas células cancerosas. Con la isoforma HOXC6-1, no hubo efectos estadísticamente significativos sobre la migración, la invasión, la apoptosis o la proliferación cuando se reguló a la baja. (Lin, 2020). Según otros estudios de células cancerosas revelan que la supresión del gen HOXC6 juega un papel en el bloqueo de la cascada TGF- β /SMAD, lo cual conduce al debilitamiento de la transición epitelial a mesenquimatoso para el carcinoma de cuello uterino. (Zhang et al., 2018)

El gen HOXC6 como supresor tumoral se inactivaría por metilación del promotor y/o una mutación somática, que sería la causa principal de la inactivación del gen y la desregulación de sus funciones. afectando las vías de diferenciación, proliferación, crecimiento y renovación celular normal. (Guerrero-Preston et al., 2014)

El gen HOXC6, estaría involucrado en varios tipos de cáncer, incluidos el cáncer de pulmón, osteosarcoma y el cáncer de mama, además está asociado con varias progresiones tumorales; además el silenciamiento del gen HOXC6 llevaría a la inhibición de la proliferación celular anormal y la promoción de la apoptosis en células malignas. (Zhang, 2018)

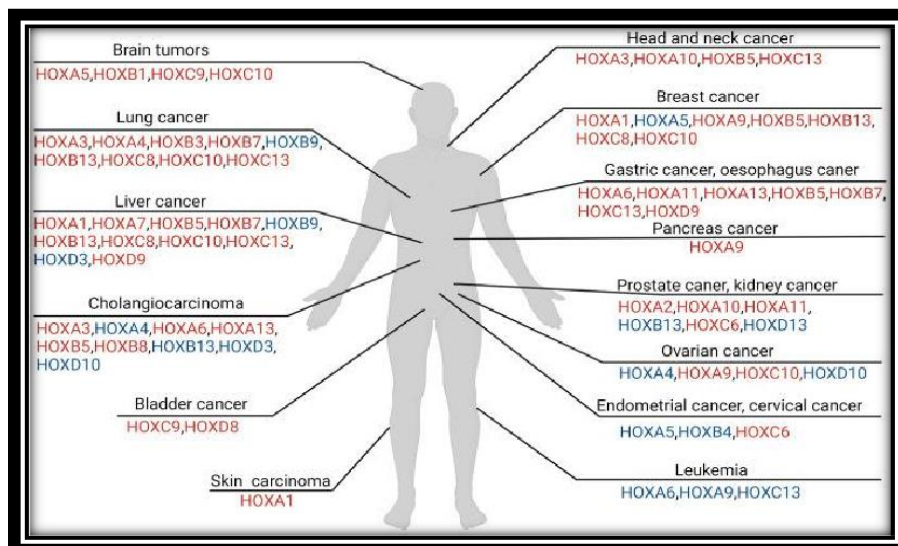


Figura 9. Descripción general de la participación de HOX en diferentes tumores en humanos. Descripción general de la participación de los genes HOX en tumores humanos, las letras en rojo indican que los genes HOX están regulados al alza en el tumor, mientras que las letras en azul significan que los genes HOX correspondientes están regulados a la baja en comparación. (Tomado de Homeobox Genes in Cancers: From Carcinogenesis to Recent Therapeutic Intervention, por Feng, Y. et al. *Frontiers in Oncology*, 11.).

5.6. PERSPECTIVA A FUTURO.

5.6.1. GENES HOX COMO DIANAS TERAPÉUTICAS

La desregulación de los genes HOX está íntimamente asociada con numerosas neoplasias malignas humanas, lo que destaca la participación indispensable en el desarrollo del cáncer mortal. La mayoría de los genes HOX se activarían como oncogenes o supresores de tumores en el cáncer humano, por ello conocer sus reguladores aguas arriba y aguas abajo contribuirían a dilucidar la función de los genes HOX en el desarrollo de procesos neoplásicos.

Es indiscutible la función de los genes HOX en el desarrollo del cáncer, cumplen un papel fundamental en dos procesos agresivos en el desarrollo del tumor, como son la invasión y la metástasis, responsables de la mayoría de las muertes por cáncer. Por otra parte, controlar la expresión de los genes HOX, podrían representar una estrategia prometedora para prevenir la invasión y la metástasis y disminuir el número de muertes relacionadas con el cáncer. La detección temprana de mutaciones en los genes HOX. podrían ayudar a prevenir el desarrollo de la enfermedad; un diagnóstico precoz de procesos tumorales disminuiría la mortalidad por esta enfermedad.

6. METODOLOGIA

Este trabajo se desarrolló mediante un enfoque cuantitativo a nivel descriptivo de tipo observacional y de corte transversal retrospectivo, con información almacenada en bases de datos públicas primarias como: GenBank, UniProt, Ensembl, FlyBase. Para analizar las características de los genes HOXC6 y abd-A se utilizó el lenguaje de programación Biopython que permitió el acceso a servicios en línea y el procesamiento de la información.

El procedimiento general consistió en búsqueda del gen de interés, procesamiento de datos, y salida de la información, lo cual involucró los siguientes pasos:

- a) Identificación de las herramientas bioinformáticas,
- b) búsqueda de la información.
- c) recuperación de secuencias.
- d) selección del identificador del gen.
- e) análisis de los genes abd-A y HOXC6.

6.1 IDENTIFICACIÓN DE LAS HERRAMIENTAS BIOINFORMÁTICAS

Biopython: lenguaje de programación de alto nivel que se utilizó para analizar las secuencias nucleotídicas, las mismas que fueron manejadas por el objeto Seq. Las funciones de Biopython incluyen analizadores para varios formatos de archivos de bioinformática como BLAST, Clustalw, FASTA, Genbank, acceso a servicios en línea como NCBI. (Chang et al., 2021.) El objeto Seq tiene funciones incorporadas como: complemento, reverse complement, transcribe, back_transcribe y translate. (Anexo 1)

GenBank: (<http://www.ncbi.nlm.nih.gov/genbank/>) es la base de datos de secuencias genéticas del National Institutes of Health (NIH) de Estados Unidos, una colección anotada de todas las secuencias de ADN disponibles públicamente (Nucleic Acids Research, 2008 Jan; 36 (Database issue): D25-30). La base de datos está alojada en los servidores del Centro Nacional Para la Información Biotecnológica (The National Center for Biotechnology Information) en EEUU.



. **Figura 10.** Página web de acceso a GenBank. Página de entrada a la Biblioteca Nacional de Medicina (EEUU) Centro Nacional de Información Biotecnológica; (NCBI) 2004 (Disponible en: <https://www.ncbi.nlm.nih.gov/gene/>)

UniProt: Universal Protein Resource, funciona como un recurso centralizado de secuencias proteicas (<http://www.uniprot.org/uniprot/>), permite el libre acceso a secuencias de proteínas, proporciona información funcional derivada de proyectos de secuenciación del genoma. Contiene información sobre la función biológica de las proteínas, derivada de la literatura de investigación, es una de las bases de datos de proteínas más utilizadas.



Figura 11. Página web de acceso a UniProt. Página de entrada a UniProt: la base de conocimiento universal de proteínas en 2021. (Disponible en <https://www.uniprot.org/help/publications>)

Ensembl: Esta base de datos recoge los genomas de gran número de organismos que se anotan mediante una serie de programas que localizan en las secuencias distintos tipos de características y, entre ellas, la localización de genes, exones e intrones. La base de datos se pudo consultar “on line” a través de la web, utilizando librerías escritas en lenguaje perl. Ensembl es un proyecto conjunto entre el EMBL-EBI y el Wellcome Trust Sanger Institute, y facilita el acceso a diferentes genomas eucariotas para analizar.



Figura 12. Página web de acceso a Ensembl. Página de acceso al Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>

FlyBase: (<http://flybase.org/>) base de datos del organismo modelo *Drosophila melanogaster*, se utilizó para obtener información del gen abd-A incluida: alelos, aberraciones y otros objetos genéticos, fenotipos, secuencias, existencias, imágenes, términos controlados e investigadores de *Drosophila* utilizando las herramientas disponibles en el menú desplegable "Herramientas" en la barra de navegación.

FB2022_05, publicado el 29 de septiembre de 2022
Gen: Dmellabd-A

Hogar Instrumentos Descargas Enlaces Comunidad Especies Sobre Ayuda Archivo

Información general

Símbolo	Dmell abd-A	Especies	<i>D. melanogaster</i>
Nombre	abdomen A	Símbolo de anotación	CG10325
Tipo de función	gen_codificación_de_proteínas	ID de FlyBase	FBgn0000014
Estado del modelo genético	Actual	La disponibilidad de stock	20 disponible públicamente

Resumen de genes
abdominal A (abd-A) codifica un componente del factor de transcripción que contiene homeobox del complejo bitórax. Contribuye al destino del desarrollo de los segmentos embrionarios. [Fecha de la última revisión: 2019-03-07] (FlyBase Gene Snapshot)

Todos los resúmenes

Instantánea de genes Alianza Resumen automático Grupo de genes UniProtKB
libro Rojo Vuelo interactivo

También conocido como
abdA, lab-2, lab-4, lab4, Hab

Figura 13. Página web de acceso a FlyBase. Fuente: Consorcio FlyBase. (2022)

6.2. BÚSQUEDA DE LA INFORMACIÓN: se utilizaron las siguientes palabras clave: nucleotide, gen, HOX, HOXC6, abd-A, *Drosophila melanogaster*, protein, humano, transcripción, hasta ubicar el gen respectivo.

6.3 RECUPERACIÓN DE SECUENCIAS: se recuperó la secuencia de los genes abd-A de *Drosophila melanogaster* y del gen HOXC6 de *Homo sapiens* en formato FASTA del Centro Nacional de Información Biotecnológica (NCBI).

Homo sapiens cromosoma 12, ensamblaje primario GRCh38.p14
Secuencia de referencia NCBI: NC_000012.12

Gráficos GenBank

>NC_000012.12:54016888-54030823 Homo sapiens cromosoma 12, GRCh38.p14 Ensamblaje primario
GCAGAACAGTCCCTCCCTGTAAGAGCTAACCATTTGCCAGGGAACCTGCCCTGGCGCTCCCTTCATTAG
CAGTATTTTTTTAAATTAATCTGATTAAATATTTTCCCCCATTTAATTTTTTCTCCAGGTG
GAGTTGCCAAGCTGGGGGAGCTGGGAGGGTGGGATGGGAGGGGAGAGACAGAAAGTTGAGGGCATCT
CTCTCTTCTTCCCGACCTCTGGCCCCCAAGGGGAGGAGGAATGCAGGAGCAGAGTTGAGCTTGGGA
GCTGCAGATGCTCCGCCCTCTCTCTCCAGGCTCTCTCTCTGCCCTCTCTTGAACCTCTCTCTAA
TTTTGTTGGCTTTTGGATGATTAAATATTTTAAATTTTGAATTTATATAAGTATATGTGTGGT
GTGGAGCTGAGACAGGCTCGGACGGCACAGAATAGGGAGAGCAGAGAGAGAGTGGAGAGAGAGAG
GCAGAGAGGAGAGAGGAGAGTGACAGCAGCGCTCGTAAGTTTTCTTATTGTTCAAAATGTAAC
TAATGGTCCGTGGCGTTTCAAGAAAGGAGCAGCAGGCTCTCTTGGTTAGGAGAGAGGGGGTGG
CGAGAGGGTGGGCGGTAGTTAATTTATGGGAGGAAAAATGGCGGGGAGAGAAACGGGTTAACTGG
AATACCGGTAAGTGGTGGGCGATTGTGTCTGATGCTCACTGTGTGCAAGTTGGGGGTGGGGA
GGCAAGGGAACAGCACCTGTGGGCTCAAGCGGGCGGTAGGGGCCAGCTCTCTGCTGCCACTGTGT
CAGTACCTGCTTATTATAGATCTTCGGTCTCTGGCTAATGGAAGTGGGTCGAGCGGAAGTGAAGCTC
AAGCTCCAGGAGGAGAGAACTTAGAGCTGTGCTTATTTGCCCTTAAGCAAGTTAGCCCCCAAC
CCCAACAACTTTCAGCTTCCCTCCCTCTGGCAGCCCCCTCTCCAGAGAGCGGACTGCTAG
AGCTACATGCGAGTGTGGGCGGAGGCGGCGGCGGAGAGAGAGCGGCGAGCTAGGCGGCG
GGGGGCTGTAAATTTGGAATAGGGGGAGGCTGTGTGCTGTGTGCGCTCAGCGAGAGAGAGGCTC
TGCCACCCCTGCTCCGCCCTCACTGGGCGGATGGAGGGTGGAGGTGCTGCTTGGGTGAGG
GTGGAGTTGAGGTTGGGGTGGGGATGCTACTCAAAAGCATCTGTGCTCAGAGAAAGAGGCC
TACCGGCTTTCTTCCGGGCTCGGCGGCTTACCTCCAGCGCGGCTCCAGCGGAGTGGCCAC
TGGACCGGATGCCGCTCGCCAGCATGCTGCTGTGGCTAGGACTGCTGCTCGGCTCGGGGAGGCTG
GAAATGGGGAAGGCCCTAGGCTGGGCGGGAGAGGCCAGCTCTCTGTTAGAGTCTCTAGGTGCTGTG

Sequencia completa
Región seleccionada de: 54016888 a: 54030823
Actualizar vista

Personalizar vista

Analiza esta secuencia
Ejecutar BLAST
Escoger imprimaciones
Resalte las características de la secuencia

Información relacionada
Asamblea
Bioproyecto
Proteína
PubMed
Taxonomía

Figura 14. Secuencia de nucleótidos del gen HOXC6. (Tomado de Biblioteca Nacional de Medicina (EE. UU), Centro Nacional de Información Biotecnológica; (NCBI) 2004 Disponible en: <https://www.ncbi.nlm.nih.gov/gene/>)

6.4. SELECCIÓN DEL IDENTIFICADOR DEL GEN A ESTUDIAR: Posteriormente se eligieron dos transcritos del gen HOXC6 del repositorio Ensembl cuya característica común es que en todos ellos evaluaban el número de exones, intrones, dominios y características, alelos variantes y la asignación a un determinado número de sondas oligo. Se seleccionó los transcritos HOXC6-201, HOXC6-203. Además se eligió el transcripto del gen abd-A-RB que presenta el homeodominio conservado, cuyo identificador es FBtr0083388.

Los datos de expresión de los genes previamente mencionados se obtuvieron de la plataforma UniProt, en donde se encontró la siguiente información de los genes HOXC6 y abd-A: nombres y taxonomía, ubicación subcelular, enfermedad y variantes, procesamiento, expresión, interacción, estructura, familia y dominios, secuencia e Isoforma, proteínas similares. Se accedió a la opción taxonomía del gen abd-A, y se encontró la base de datos FlyBase, en donde se realizó una búsqueda de ortología utilizando las herramientas disponibles en el menú desplegable "Herramientas" en la barra de navegación, en donde se encontró el Gen HOXC6.

6.5 ANÁLISIS DE LOS GENES ABD-A Y HOXC-6

Una vez obtenida la secuencia de los genes HOXC6 de *Homo sapiens* y abd-A de *Drosophila melanogaster* se trató los objetos Seq como si fueran cadenas normales de Python, y se obtuvo la longitud de los elementos de la cadena. Se agregaron las librerías necesarias para realizar el análisis de la secuencia del gen y se importó la secuencia del gen HOXC6 y abd-A del NCBI.

Se aplicó el método de transcripción integrado del objeto Seq, se transcribió la hebra de codificación en el ARNm correspondiente. Adicionalmente utilizando uno de los métodos biológicos del objeto Seq que usan objetos de tabla de codones derivados de la información del NCBI se logró traducir el ARNm resultante en la secuencia de aminoácidos de la proteína correspondiente. Finalmente se importó la librería matplotlib, y se comparó el número de nucleótidos, adeninas, timinas, y aminoácidos codificados por los genes HOXC6 y abd-A

```
In [7]: from Bio import SeqIO
from Bio.SeqUtils import GC
for seq_record in SeqIO.parse("/usr/local/bioinf/sampled/data/t-coffee/sequence.fasta", "fasta"):
    print(seq_record.id)
    print(repr(seq_record.seq))
    print(len(seq_record))
```

```
NC_000012.12:54016888-54030823
```

```
Seq('GCAGAACAGTCCTCCCTGTAAGAGCCTAACCATTGCCAGGGAACCTGCCCTGG...TTA')
```

```
13936
```

Se guardó la secuencia para su manejo de manera más sencilla

```
In [7]: seq_1= seq_record.seq
```

Se imprimió la secuencia para comprobar la integridad de la misma.

```
In [8]: print(seq_1)
```

Se obtuvo el número de Adeninas y se guardó en una variable.

```
In [10]: a_seq1=seq_1.count('A')  
print(a_seq1)
```

Se obtuvo el número de Timinas y se guardó en una variable.

```
In [11]: t_seq1=seq_1.count('T')  
print(t_seq1)
```

Se obtuvo el complemento de la secuencia de ADN

```
In [11]: Seq1_C=seq_record.seq.complement()  
print(Seq1_C)
```

Se transcribió la secuencia del gen HOXC6 para lo cual primero, se creó seqobjetos para las hebras de ADN de codificación y plantilla.

```
In [21]: trans_seq=seq_record.seq.transcribe()  
print("\n")  
print(trans_seq)
```

Posteriormente mediante la traducción del ARNm se obtuvo la proteína de nuestra secuencia y la cantidad de aminoácidos lo guardamos en una variable.

```
In [24]: seq_tradu=seq_1.translate(stop_symbol='')
         print(seq_tradu)
         lenseq_tradu=len(seq_tradu)
         print(lenseq_tradu)
```

Posteriormente se importó la secuencia del gen abd-A

```
In [13]: for seq_record_2 in SeqIO.parse("C:/Users/momta/Downloads/sequence_mosca.txt", "fasta"):
         print(seq_record_2.id)
         print(repr(seq_record_2.seq))
         print(len(seq_record_2)) # NUMERO DE NUCLEOTIDOS
```

Se guardó la secuencia del gen abd-A para su manejo de manera más sencilla

```
In [14]: seq_2=seq_record_2.seq
```

Se guardó el número de nucleótidos del gen abd-A en una variable para su comparación.

```
In [15]: nuc_mosca=len(seq_record_2)
         print(nuc_mosca)
```

Se obtuvo el número de Adeninas en la secuencia del gen abd-A

```
In [16]: a_seq2=seq_2.count('A')
         print(a_seq2)
```

Se obtuvo el número de Timinas en la secuencia del gen abd-A

```
In [17]: t_seq2=seq_2.count('T')#TIMINAS
         print(t_seq2)
```

Se obtuvo la proteína de nuestra secuencia y la cantidad de aminoácidos se guardó en una variable

```
In [18]: seq_tradu_2=seq_2.translate(stop_symbol='')
print(seq_tradu_2)
lenseq_tradu_2=len(seq_tradu_2)
print(lenseq_tradu_2)#numero de aminoacidos,codones
```

Se generó un gráfico de barras para la comparación de nucleótidos:

```
In [19]: import matplotlib.pyplot as plt
fig, ax = plt.subplots()
plt.title("NUMERO DE NUCLEOTIDOS")
plt.xlabel('SECUENCIAS')
plt.ylabel('NUCLEOTIDOS')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [nuc_homo, nuc_mosca])
plt.show()
```

Se generó un gráfico de barras para la comparación de adeninas, timinas y aminoácidos

```
In [20]: fig, ax = plt.subplots()
plt.title("NUMERO DE ADENINAS")
plt.xlabel('SECUENCIAS')
plt.ylabel('ADENINAS')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [a_seq1, a_seq2])
plt.show()
```

```
In [21]: fig, ax = plt.subplots()
plt.title("NUMERO DE TIMINAS")
plt.xlabel('SECUENCIAS')
plt.ylabel('TIMINAS')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [t_seq1, t_seq2])
plt.show()
```

```
In [22]: fig, ax = plt.subplots()
plt.title("NUMERO DE CODONES")
plt.xlabel('SECUENCIAS')
plt.ylabel('CODONES')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [lenseq_tradu, lenseq_tradu_2])
plt.show()
```

Adicionalmente se obtuvo la tabla de aminoácidos correspondiente al código genético.

```
In [35]: standard_table = CodonTable.unambiguous_dna_by_name["Standard"]
print(standard_table)
```

Para cumplir el tercer objetivo, la identificación de la región cromosómica, el número de transcritos codificantes correspondientes al gen abd-A de *Drosophila melanogaster* y del gen HOXC6, de *Homo sapiens*, se utilizó un navegador genómico como Ensembl, se pudo obtener la secuencia y localización del gen de interés. En la página

principal de Ensembl: <http://www.ensembl.org>. se eligió “Human” y en el campo de texto de búsqueda se colocó el gen de interés: HOXC6:

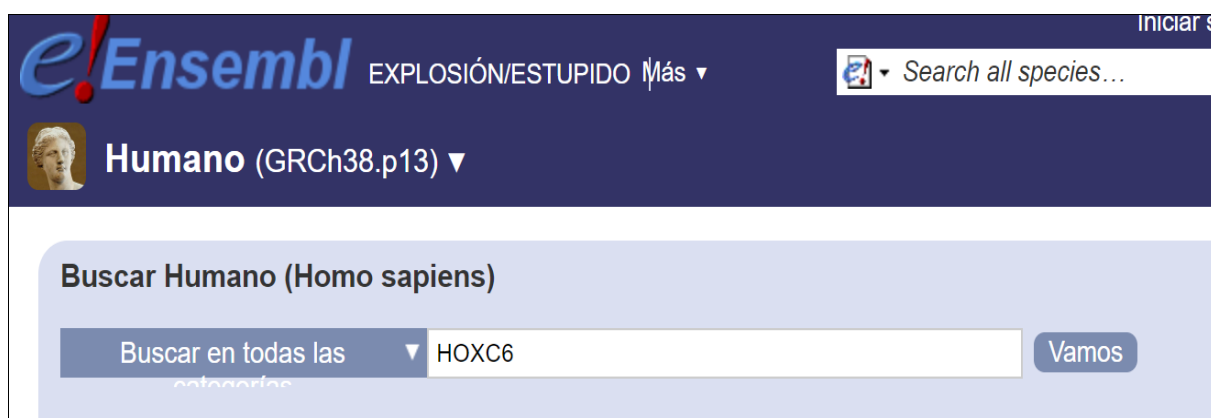


Figura 15. Búsqueda de secuencias del gen HOXC6 en la plataforma Ensembl. Entrada en la plataforma Ensembl para la búsqueda del gen HOXC6. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

De igual manera para el gen abd-A, en la página principal de Ensembl se eligió la opción Gen y en el campo de texto de búsqueda se colocó el gen de interés: abd-A:



Figura 16. Búsqueda de secuencias del gen abd-A en la plataforma Ensembl. Entrada en la plataforma Ensembl para la búsqueda del gen abd-A. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

Para el cumplimiento del cuarto objetivo mediante el navegador Ensembl se analizaron 2 transcripciones del gen HOXC6: HOXC6-201, HOXC6-203, se encontró el número de exones, de intrones, la longitud de transcripción, la secuencia ascendente 5' y descendente 3'. Posteriormente la búsqueda y análisis de la proteína codificada por el gen HOXC6 y abd A se realizó en la base de datos Uni-Prot., se encontró la proteína Homeobox codificada por los genes HOXC6 y abd-A, así como la secuencia de aminoácidos de las proteínas resultantes.

7. RESULTADOS

En respuesta al primer objetivo de investigación, en el gen HOXC6 se encontró 3266 adeninas, 3469 timinas, la secuencia de nucleótidos resultante, la cadena de ADN complementaria, la secuencia transcrita en ARNm, la secuencia traducida del ARNm con sus 4476 aminoácidos se muestra a continuación:

```
GCAGAACAGTCTCCCTGTAAGAGCCTAACCATTGCCAGGGAAACCTGCCCTGGGCGCTCCCTTCATTAGCAGTATTTTTTTTAA
TTAATCTGATTAATAATTATTTTTTCCCCATTTAATTTTTTTTCTCCAGGTGGAGTTGCCGAAGCTGGGGGCAGCTGGGGAGGG
TGGGGATGGGAGGGGAGAGACAGAAGTTGAGGGCATCTCTCTTCTTCCCGACCTCTGGCCCCCAAGGGGCAGGAGGAATGCA
GGAGCAGGAGTTGAGCTTGGGAGCTGCAGATGCCTCCGCCCTCTCTCTCCAGGCTCTTCTCTGCCCCCTTCTTGCAACTCT
CCTTAATTTTGTGTTGGCTTTGGATGATTATAATTATTTTTATTTTGAATTTATATAAAGTATATGTGTGTGTGTGGAGCTGA
GACAGGCTCGGCAGCGCACAGAATGAGGGAAGACGAGAAAGAGAGTGGGAGAGAGAGAGGCAGAGAGGGAGAGAGGGAGAGTGAC
AGCAGCGCTCGGTAAGTGTTTCTTATTGGTTCAAATATGTAACATAATGGTCCGTGGCGGTTTCAAGAATAGGACAGCAGCGGCTT
CTCTTGGTTAGGGAGAAGAGGGGGCTGGCGAGAGGGTGGGCGCGTAGTTAATTATGGGGAGGAAAATTGGCGGGGGAGAGAAAACG
GGGTTAACTGGAATACCGGTAAGTGGTAGTGGGGCATTGTGTTGTCTGATGTCACTGTGTGCCAAGTTGGGGGTGGGGAGGCA
AGGGAACAGCACCTGTGGGGCTCAAGCCGGCGGTCTAGGGCGCCAGCTCTCTGCTGCCACTGTGTGAGTACCTGCTTATTATAGA
TCTTCGGTCTCTTGGCTAATGGAACCTGGGTCCGAGCGGAACCTGGAAGCTCAAGCTCCAGGCAGGAGAGAGAACTTAGAGGCTGTGC
CCATTGCCCCCTAAGCAAAGTTAGCCCCCAACCCCAAACTTTCCAGCTTCCCTCCCTGTCTGGCAGCCCCCTCCCTCCC
AGAGAGCGGACTGCTAGAGCTCACACATGCGCAGTGTGGGCCAGGGCCGGGCCCGGAGCAGGAAGCCGGCGCAGCTAGGCGGC
CGGCGGGGCTGTAAATTGGCAATTAGGGGGAGGCTGGTGGCTGGTGCAGCGTACAGCGAGAGGAGAGCGTCTGCCACCCCTGC
TCCCGCCCCACTCGGGCGGATGGAAGGGTGGGAGGTGCCCTGCGTTGGGTGGAGGGTGGAGGTTGTAGGGTGGGGGTGGGGGATG
CTGTACTCAAAGCCATCTTGTGCTCAGAGAAAAGAGCCTACCGGCTTTCCTTCCGGGTCCGGCGCCCCCTCACCCCCAGCCGC
GGCCATCCCAGCCGGGATGCCACTGGACCGGGATGCCCGCTCGCCACGCATGGCTGCTCTGGGTAGGACCTGCCTCGCCTCGGG
GGGAGGTGGAATGGGAAGGCCCTAGGCTGGGCGGGGACAGGCCAGTCTCTGGTAGAGTCTCTAGGTGCTGTGGAGGCAGT
GAGGATTCGGGGGCTGCTGCTTATGGGGTAGAGCGCTCAGCCTGGGGCGGGGAGGGTCTTGGGCACTCGGGATCTCTCTAAT
AATTGGTCTAATGGCTGTTGTATAGTCTTAACCCGACGCATATGGTTTCATAACACAGTGGCTTAATTTCTTCCAAATGTACGT
GGCCCTGAATGTGTGAGTGTGATATATGACCCCGCCTGCTCCCGGCCTGGGTCCGAGGCTCCGAGGCGGCCCATAAAGTGTG
AATTTCTGTTTATAAACGGCGACGCACACGCAACGCCTACACACCAGGCAGACACACCCGGAGGCGTGGGGGAACATATTCCAG
ACGAGGTCCGAGAGAACCCTGGGGGCCAGACTTACCTTAATCTAAGCCAAGACACCCCTCTACCCCCACCCCTCTCCAGAGGCGCG
ACTGGGATTTGCGCCAAGTCCCAGTGGGGGAGGGATCGCTCCACGATTTTCTGTTTAAACCCCTGACCCGTGCGCTGTGGGGGGG
CGGGATGGGGGAAGAAACCTGTTTTATCAAATTTTAAAAATACAGTTTGCATTAGATCAAACCGTTTGGGCCAGAGCTGACG
GTGAATTTTATCTCCCTCCCCCACACCCCTTCCCTTCTTTTTCTGCCCCCGCTAAATTTCCCTTTCGAGCCATCATAAAGG
CCTCTGAGGTATTCTCCCGCGGAAGAAATGGCATTCTCTCCCCCTCCCCACCCCCACCCCACTAGGACTTGTGTGTCGG
CAGAGGCGTCTGACGAGGGGCTAATTTGCTTTTCTTGTACGATCGAGAAAAGCGCTTGGCTCTCCCCAAATGGGCCAGCCCCG
CCGCTGCTCAGGCTGCCGCGCGCTCTCCCGCCGCTCAGGCACGGCCAGTGGCTGATGCCCGCGGATCCAGCCTGCGGCAGGCC
GGAACCCGGAGTCTGTGGGACCTGAAAGGGCAAAGGGAGTGCAGTGGGACACAAGAGGGGCCCGAGAGGCCGAAAAAGGGAGCC
TTTCTCTGGAGTAATGTGGGAATCGCCGACCGGTGAGGCCTCTATTTCTCTCTTTCGCTGCATTTTGTCCGCCCCCGCCCTGT
GGATGGCCGCTGATGGGGGGAACACAAGTGTCTTTTGTCCGTCGCCGAAAGAAAACCTTTTCCAGAGAAGGCCATTAGGCCAGT
GGAGATTACAGATAATATTAGAAGGGCCCCCTCCCAACCGAGGGCGGGCTGGGCGAGGCTGAGTTTCTCTTAATCTTAGA
ACGACCTCACTGAGTGTACCCACAGCGAGATTTGGAGGTGCTAGAGTTCGAGGTTGGCCGCTTAACAGCACCATCTGAGGTGTCT
TTATATAATATTTAATTGATTTTCTGTCCCTTCTGGCTACAGCCCTGCAGCTCCCCACCCCCATCTATGCCCTTACCCCTA
GACCCATCCTTACAAGACAGAGGCAAGCAGAGAAGGCATAGCAGCAGCGACCGGCGCTCTGTTTTCAATTTCCACTCTGGCCAGGG
GATAAACTGGACCCAGTGGACTCCAACACCAACTCTCTGCTTCCAGTTACTTTTCGATTTTCTTTCCCTTCTCTTCTCTC
CCTCTCTGGCATTAAACAGAAAGCACAAGAGGACCCACCAGTAGGCAACTCTAAGCCTCCTTGCAGGCCACATCCCTGCTGACGTG
TCTCTTCTGCTCAGGAGAGCCCTTCTGTGATTATCTAATAATAGCCAGGGTGGCTGAGAATAGAATCTGGCTCAAGGACTCTGCACA
CCCGGGGCGAGAGACCTGAATCTGTGGTCAAGGCCCTGGCAACATTTCTTGGAGCCTCCCCCTCTGAGCCCACTCTTTGGGCACCC
TTAGGCTTGCCTGCCCCCTTATAAGGGGCTTTGTGACGCCCTGCCATCCGCTGTATTTTCATGTGCTGGTTCCTAGCTAAGGAGAAT
TGGGCAGAGCAAACCTTTCTCACCTGGTCTGACAGTTGTGATCCGACAGAATGCGACCTTGTTTGGAGAAGGCGACCTTGTG
GGTTGGCTGTGTTCTGAACACAGGCTTGGGAGTGTATATATTTATATATACAGGCTTAAATATACAGACTCAGGCTTATGCATG
AATATATACAGACTCATATTTCTTGGTGGTAAATTGTATATTACCATGCCTATATTCGCTCAAACACCATTTTATGTTTCATGCAA
AATTTTGTAGTCTCTACTAGATGAGAGTGGGCTCATGGAGGCACCGGGTGCAGTGAGCAGGCCCTCTGTGCATGCTCTGCTGGAA
ATATGTGCTCACTGGGAATCGGAGTATCGGGTTCCAGTAACCAATGGATGCCATAAAGGCTTTCTCAGACCTTATTCGCTTTTGTG
CCCAAGCAGCTGGACTGGGGAATACTAAATCCCTTCTGCTTTCTCTCCACACCTCCTGGCAAACCTCTCCCACTTAGAAATTAT
TTACAAGGGAGACAGTTGTCTTGAATTTGAGTTCTGCGGGATTCTTTTACGGGCATGGTGGGGTGCAAGCAGTTCCGAGTAGT
CAGGGCCCTGGCGCTGGAAGCCTCACTGGAGCGATGAAGGGATCAGAGTGGTTAATAAGGTGGCCTCTGGACAGTGTCTGGGGCA
GCTGCTGCCAGGGTGTCTGGAGAAGCTTGTGGAAGGCAAGCCAAAGGGGTAGGGATGCAGTGAATATCTGTTTCTCCCCCTG
CTTGCCCCAGCCGGCGCTCACCCCGCCCCAGCCGATTTGCTGCTATCCGTGGCGGCAGATTGGGAGTTCTCTAAGGTCCCCAA
```

ATACCCAGATCTGGTTACTGGATGAACTGCGTTTTAATCCTCGTCGTATCTGAGCCAGGACAAGACCTGGTCAGGGGAGACGGCTG
 GTCCGGGCAGCACGGGAGAGGACGCTGGACCCGCTGCCTGTGGCCTGGAGCACTGAGTATGGGCTACCAAGGCTGAGAATTGATGG
 GCCTCCTTTCCCAGCCTTTTTCCCTTCCAGAGACACAAAAGACAGAGCAGAGGGGAAATATCAATGTCAGTGGAACCCCATAA
 AAGAAAGCAGGAGGACGCGGGCTGGAGGGTTTTCCGCTACTCTCTAGGTGATCGGGTTCTCCCAAGACTCTGTGCCAGA
 CAGGCTGCACCAAGAGTGTGCAGCTTCTTCCACCTTCCCTCCGCCCCCTCAAGGAGGGCAGCCTTAGGCTTTCTGGGCTTGGTC
 ACCCCAGGCACTGGGGTGACATTACTCCAGGATTGTGCCCTCGCCAGCCACGGCCTCTGCCACGGCGCCGACGCGCTCCTGT
 TTATTGTGTCTGTGAGTTGTTTACTTGTGTGTCTGTGCCACCTGCGTCAGAAACCCGCCAAACCAGCGAACTCGCCAGGAGA
 ACAAAGTGTGTGCTATCCCCGACTCCCTCCCTTTCCAGCAACCCCTCCTGCCCTCACAAGCTCCTGCATTCTCCCTTTCCGGA
 GGGGAGGAAGCAGAAATAGTCTTAAACAGGGCAGCTGTTTATAGGGTCAGGGGAATTATGGGGACACCATAAACCTCTGGGTGTC
 CCTGTGGTGGACAGACTTCTGAGCAGCCCTTCCCTTGCCTAGGCTTTTCTTCTTCCCTGGTGACCCTTGGTTAAGGGGGCCAC
 ACTCATTGAGAAAGTTGAGCCACCCCATCAACCCAGTCCCCCACTACCCCACTCCTTTCTGCAGGAGTTTCCCTGAGACACAGG
 TTGTTGGGTTTTAAATCAGGGTAATGTTGGTGGATTCTATAAAGCTAGATATAGCATGTCATTCTTCTCCCACTCCCTCTCCTCA
 GCCCCCAAAATTGCAGTGAATTTAAGCAAATCCTACAAAATTAATTTGCCCTTTGCTAAAGATCCTATATGTCTCCAATCCTGC
 CAGGTGTAGAGTCAGATAGTTGGCTTTTGGCCTTTAATTTTGTCTGTAGCTATTAAATTTAAAAGAGTATCATTGAAGTAAGTAAT
 ATTCTCATTTCTGAAATTAATTTTCCCTCCCACTCTCTCTCTCTCTCTTTTGGCCAGCACAGATTAATGGTTCTTTCAGATCTG
 CAAACAATAAAGGATAAAATATTATATAGGGAAATGAGGATGCTGTTTTATTGTTGGCATATTGTTCCAGAACATAACAACATTTT
 ATTTTCAAGCAGGAGAGTTAGACCAGGGGGAGAAAGAGGGGGAAAGGAAAAACATGAAAGGAGGGCAGGAACCCACACAAAA
 CTCCTGCCATGTATAAGTGGTAGGACATTTACAATTTCTATATGACAACTCACTGTGATGTCTATAAGTACGAGTTCTGAAATTTTA
 TAATTTCTCATGAACCTCTCAAATGGGTCCAAGCAGTGCTTCATGTGGAGGGGTATAGAAGAAGCCAGGAGGCTGGGGATCCTTTGA
 CCCAGTGCAGAGCAGCACACCCCTATTGCGGCTTCTTCTCATGTGACATCCTCTTGAGGCCCTATTCTGGGGGCTGAAAATATT
 CTGGATATTCTGCATTTAATTTTTCTCCTGGGCCAAGCAGGGCCACTTGCTTTGCCCCCTCAAGTGCCCTAGGCATCTAAGCT
 CTCTCCCCATTCCCAGATCACAGAGAGCTGTGTCCCCAGGCTCTCCCCAAGCCCCCTCCTCAGCTAATACTCAGCTGCTTGAATA
 GAGGTGGGAGGCTTGGGCACCCAGCCCCAAAAAGCCAGAAATGATTTTTTCCAGATGGAGCTGGAAGGGGGCTCAGGGCTGAGCGT
 GATGGGTAGAGATGTTTGGGACCTATTTTACGTTTTAACCTTAAGTCAGAGAAGTGAATTGAGAAGCCAAACAAATGCCAATTGTG
 AGAGTTTGTATCTGCTGATGGTAGCAGCAATAAGTGAAGTGCTGGTAAGGAAAGGTGAGGCAGAAATCTGGAAAGAGAGAGGGAGG
 AGAGGAAATGTGACATGAGTTAGCAGAGCCCTTCCATATACCTTCTCTCCGTTTAATTGTAATAAAGCAAAGCAGCTCGGGTA
 GCCTCTCAACCTCTTAGCTCAAAGAGGAATCACCAGCGCAGGCCCTGGAGGTGTTTGGGGCAGTTCTCAGCAGCTTAACATGCC
 CAATCTGCCCTCTTCTCCTTTTCCCCCTCCCTGCTGATCCCCAGCTCCCTTCAACTGGTCAGCCCCGGGCCCTGTCCCCAAATATT
 AGCACCCCAAAATGATATCCATTTCTGTGTTAGGTGTGTCAAATTTCACTTCTCTCCATGCCTAAACTTTGAAAGGAGGAGTGTAC
 CTTTTTGAATGCGGTGTGTGTAGCGGGGCGAGGTATGTGCTCAGACAGCTTCCCCATCATATCTATGGGAAACTGGAAGCCCT
 TAGGTGAGCCATGCTTTGGTGTTCCTCCAAACAGCAATCCCTGCATTCAGGTTCTCTGCGCAGACAGGAGGCTCTTTGCCATCTGATC
 CTTTGACTAGGAGTTCCATGGAGATCAAAGGAGCTTATTTATACACAAAACACCAGGCATTTGGAGTAGAATTTATTTTTTAATTT
 CTAAATAAAAAAGTCTAATCATAGACCAATCGGGCCAACTGAGCAGTCGCTCCAGAACAGCAAAGCCTTAACCAATAGTGTGATT
 ATGAGGGGAGCTGAGCAATTACTGAACTCGGCTTTCTCCGTAAGAGGGATTTTAGCCCTCGCCCAGTGGCACAAGGGGGCCAGC
 GGGGCAACCGTTCTTCTTCTTGGCTCAGAAAAGTGAGACTTGGGGGTGGTGGGGGCAGGGTCGGGCAGCTGCGGTGCGCTCCAGGCAG
 CTGAGCACCCAGGCCAGGCGCTGAGGCTGAGCACTGAGTCTGAGGCTCCAGTGCTTGGGTCCCCGAGCAGCCTCAGGAGGCACCG
 AGCTGTAGTTGCAGCGCGCAGTCAGGAATGGAGTATGTGTATCTGTGTGCGAGGGTTGGTGAAGTTACAACATCCGCGCAGAGTT
 AGGGGATCTGGGGAGAGAGAGAGACAGGGAGAACACAACCTAAGGAGGACCTCTGGAAATATTGTCCCCAAATAAAAATCCAGAG
 TCTCATCCCTCTCACTTCAAATATCCCTCCCCAGGAAATTAATATATTCCCAGAGCTGAGTGAAAGCAAAGGCCAGCAAGGTGA
 GGCCTGCCACGAATCTTTCAGGCTGGGTCCAGTCCCACCCACCTTTCCCCAGTTGAGAGCTCACATGCCCCAGCCTTTCCCT
 GGCCTAAAGGGGAGTAAGGCTGGAGTGGTAGCATTAACTTGCCCTAGGGAATCCCCAAGTGCCAGGCTGCAATGCCACCCTTAT
 CTTTGGGGACTCCAGGCCAGTGACTGCACCCCACTCAGGTCCAAGGCGCAGGCAGTAAATTTCTCTCCTATAATTAGCGCTATCAA
 AAGCATGTGTTTTCATTGGAGAGGACTGGAGAAGCAAGTGGCAATTTTTCCCTTGATTTCTTTTTCTGTGTTTTTGAATTTCTGT
 TTTTGAATCCCTAATAAGCTTCTCTGAAAGTGAAATCTGCTCTTGACACCAAGCCCTTCCGATCTGATTTTGGGAGGCT
 AAGGGGCCAGAGCAGTAGCAAGGAGACCTGCAAGCCTGATGGGGTTGGGGGTAGTGTCTCTGATTTCTTTAGGCTCGTTTTATA
 GAAACAAAGAACAAGTTCTCTCCCCCTCCTTCCCAAGACTTGTGCATATTTACAGAGCACAACCTGCAGTTTTAATAAAACATCTGTT
 CTTGCTTCTGCTGATGAGTTTCCATAATTGTTTTTCCAGACAAATTTAACAGGAAAATATAAATATATTTAGAAAAATCTAAGTCCCA
 GCTTTCCCCCTATAGACAGCCATTCTTTTGCCAAAAAATAAAATTAACATGAAATGGGAAAAAGAAAATATAATCCAATGTAGGG
 AAGATCGGGCTCATCTGACCTGCTCTCCCCCTTGGAGCGAATCCTTTTCAGCAAATTTTCAGCTGCATAATTAAAGATCTAAGAGAA
 AGAAGAGGGCTTCTTCTTCTCAGGAATGAAAGGTAATTGGGGGGGGGGAGGTGTTGAAAATTAATTTGCCAAAGGTCTTTAA
 GCAGTGGAGGAATTTAATTTGTATTTCTGTTTCCCTTCTCTCCCCCTCCAGCATTTCTGCCTTTTTTACCACCCAGCAGCTGGTTTCT
 GTTCATTATAAATCGGCGAGACCTTTTCAGTCTCTGCCCTCTGTTTAGGGACGTCATAAAACACTTAACTTTCTGGGGCTGAGGCTC
 ACATTTGCGCTCTCAGGTGCGCAACCTCCCTGAGCCCACTGTTAGGCCCAATAGAGAATCTCCCCAACTTTGGGGCCCCCTTT
 CCTCTCCAGACATGGTTGAGGAGGGGCTCAAGGAAACCTAGGCATCCTTGCTGGGGGCTGTAGTCCATGCTGCCCCAGAAGCCCT
 TCTTTCTTCCAAGGTCTTCCCTCCCCCAGACATAAAAGAGATTTTTTAAATACATAGAAAATCAACACTCATTGAAAAGTGAA
 ACCTCATTAAGACATCCTATCTTCCATCATTCAATTTGTTGTACATTGCAACAATTTAATATCTTGAAGGAAATATCACTGACATG
 ATTTATCCCTCTAGCAATCTCTCTGAAATTGGAGGGGGGACAGAGTTACTCTCTCTTCCGGCCTCTCTACTATTGTTACACCAC
 TACCTTAGGGCAGAAAAGGAGGCTGGGAACCTCTGAGAGGGGCTACTTGTGTCCAGGGGTGCATGACTGTTTCCCTAGCTCAGTCC
 CCAGAGTACTCCTGCCTTGTGTCTCTCCCCACCTCTTGCCCTCCAGGGGGCCTGAAGCCTCAGCCAGCTGTTGGAACCTCTGC
 TCAGGCATGCCATGAATTCGAGGACAGCTGGATTCTAGGGCAAAAAGCAACTCACATTGGTGTCTATGTTGTGTACCTGCCAAGT
 CACCCCAAAAATGAGATAACTCATCTAAAATTTTTTTCTTAATTTAAAAATCCTCCTCTACCCCTTTCCCTTCAATTTCCCTGGGG
 AGAAGAAAGTGTGATCTAGAAGGAGGAACCTTTATAAATTATAGGAGTCTTTCCCTTCCCTTCCAAATACCGACAAATTTCTGTCC
 TTTCTGGAGGATTGGGGCTGTATTTTCTAAAACCAACACAAATTGTCTGGCAGGAAGACTCTTCAACCCCTTCAACGAGGTTCATGT
 ATTAACTAACATGTTGTCTCCACCAGCTCCTGCACACTGTCTGCTTTTGGGTTGAAACTTTATTTTTTCCCCCTAGCGACACTCC

CGTCTTGTCAGGAGGGACATTCTCGGATTGGTAACGGTCCCTTTGGACGGGACCCGCGAGGGAAGTAATCGTCATAAAAAAAATTT
AATTAGACTAATTATTAATAAAAAGGGGTAAATTAAAAAAAGGAGGGTCCACCTCAACGGCTTCGACCCCCGTGACCCCTCCC
ACCCCTACCCTCCCCTCTCTGTCTTCAACTCCCGTAGAGAGAGAAGGAAGGGCTGGGAGACCGGGGGTTCCCCGTCTCTCTTACGT
CCTCGTCTCTCAACTCGAACCCCTCGACGTCTACGGAGGCGGGGAGAGAGAGGGTCCGAGAAGGAGGACGGGGGAAGAACGTTGAGA
GGAATTAATAAACAAACCGAAACCTACTAATATTAATAAAAAATAAACTTAATATATTCTATATACACACACACACACTCGACT
CTGTCCGAGCCGCTCGCCGTGTCTTACTCCCTTCTGTCTTTCTCTCACTCTCTCTCTCGTCTCTCCCTCTCTCCCTCTCACTG
TCGTGCGAGCCATTACAAAGGAATAACCAAGTTTATACATTGATTACCAGGCACCGCCAAAGTTCTTATCTGTGTCGCGGAA
GAGAACCAATCCCTCTTCTCCCCGACCGCTCTCCACCCGCGCATCAATTAATACCCCTCCTTTTAACCGCCCCCTCTCTTGC
CCCAATTTGACCTTATGGCCATTGACCATCACCCCGCTAACACAACAGGACTACAGTGACACACGGTTCAACCCCCACCCCTCCGT
TCCCTTGTGCTGGACACCCCGAGTTTCGGCCGCCAGATCCCGCGGTTCGAGAGACGAGGTTGACACAGTCTCATGGACGAATAATATCT
AGAAGCCAGAGAACCGATTACCTTGACCCAGCTCGCTTGACCTTCGAGTTCGAGTCCGTCCTCTCTTGAATCTCCGACAG
GGTAACGGGGGATTCTGTTCAATCGGGGGGTTGGGGGTTTGTTTGAAAGGTGCAAGGGGAGGGACAGACCGTGGGGGAGGGAGG

TCTCTCGCGCTGACGATCTCGAGTGTGTACGCGTCACACCCGGGTCCCGGCCCGCGGCTCGTCCCTTCGGCCGCGTCGATCCGCCG
 GCCGCCCCCGGACAATTAACCGTTAATCCCCCTCCGACCACCGACCACGCGCAGTCGGCTCTCCTCTCGCAGACGGGTGGGGGACG
 AGGCGGGGGGTGAGCCCCGCTACCTTCCACCCCTCCACGGGACGCAACCCACCTCCACCTCCAACATCCACCCCCACCCCTAC
 GACATGAGTTTTTCGGTAGAACACGAGTCTCTTTTCTCCGGTAGGCCGAAAGGGAAGGCCCGAGGGGAGTGGGGGTCCGGCG
 CCGGTAGGGTCGGCCCTACGGGTGACCTGGCCCTACGGCGGAGCGGTGCGTACCAGCAGACCCGATCCTGGACGGAGCGGAGCCC
 CCCTCCACCTTTACCCCTTCCGGGATCCGACCCGCCCTGTCCGGGTGAGGAGGACCATCTCAGAGGATCCACGACACCTCCGTCA
 CTCCTAAGCCCCCGACGACCGAATACCCCATCTCCGAGTCGGACCCCGCCGCCCTCCAGGACCCGTGAGCCCTAGAGAGATTA
 TTAACCACGATTACCGGACAACATATCAGAATTGGGCTGCGTATACCAAAGTATTGTGTACCGAATTAAAGAAGGTTTACATGCA
 CCGGGACTTACACAACGTCAAACCTATATACTGGGGCGGGACGAGGGGCCGGACCCAGGCTCCGAGGCTCCGCCGGGGTATTACAC
 TTAAGACAAATATTTGCCGTGCGTGTGCGTTTTCGGATGTGTGGGTCCGTCTGTGTGGGCTCCGACCCCTTGTATAAGGTC
 TGCTCCAGGCTCTCTTGACCCCCGGTCTGAATGGAATTAGATTTCGTCTGTGTGGGGAGATGGGGGTGGGAGAGGGTCTCCGCCG
 TGACCTAAGACGCGGTTACGGGTACCCCCCTCCCTAGCGAGGTGCTAAAGGACAAATTTGGGACTGGGCAGCGGACACCCCCC
 GCCCCTACCCCTTCTTTGGACAAAAATAGTTTAAAAATTTTATGTCAAACGTAAGTCTAGTTTGGCAAACCCGGTCTCGACTGC
 CACTTAAATAGAGGGGAGGGGGGTGTGGGGGAAGGAAGGAAAAGACGGGGGCGATTAAAGGGAAAGCGTCGGTAGTATTTCC
 GGAGACTCCATAAGAGGGCGCCCTTCTTTACCGTAAAGAGAGGGGGAGGGGGTGGGGGGTGGGGTGCATCCTGAACACACAGCC
 GTCTCCGAGACTGCTCCCCGATTAAAGCGAAAGGACAAATGCTAGTCTTTTCGCGAACCGAGAGGGGTTTACCCGGGTCCGGG
 GCGGACGAGTCCGACGGCGCGCGAGAGGGCGGCGAGTCCGTGCCGGTACCGACTACGGGGCCCTAGTTCGGACGCCGTCCGG
 CTTGGGCTCAGACACCTGGACTTTTCCCGTTTCCCTCACGCCACCTGTGTTCTCCCGGGCTCCGGGCTTTTCCCTCCG
 AAAGAGGACCTCATTACACCTTAGCGGCTGGCCACTCCGGAGATAAAGAGAGAAAGCGGACGTAAGACAGGCGGGGCGGGGACA
 CCTACCGGCGACTACCCCCCTTGTGTTACAGGAAACAGGCAGGGGCTTTCTTTTGGAAAAGTCTCTTCCGGTAAGTCCGGTCA
 CCTCTAAGTGGATCTTATAATCTTCCCGGGGAGGGTGGGCTCCCGGCCCGACCCGCTCCGACTCCATAAGAGAAATTAGGATCT
 TGCTGGAGTGACTCACATGGGTGTGCTCTAAACCTCCACGATCTCAAGCTCCAACCGGCGAATTGTGCTGGTAGGACTCCACAGA
 AATATATTATAAAATTAACATAAGACAGGGAAGGACCGATGTCCGGACGTCGAGGGGGTGGGGGTAGGATACGGGAAGTGGGGAT
 CTGGGTAGGAATGTTCTCTCCGTTCGTCTCTTCCGTATCGTCTCGTGGCGGAGACAAAAGTAAAGGTGAGACCGGTCCCTCC
 CTATTTGACCTGGGGTCACTGAGGTTGTGGTTGAGAGACGAAGGTTCAATGAAAGCTAAAAGAAAGGGAAGAGAAAGAGAG
 GGAGAGACCGTAAGTTTGTCTTCGTGTTCTCCTGGGTGGTACCGTTGAGATTCCGAGGAACGTCCGGTGAAGGGACGACTGCAC
 CAGGAAGGACGTCTCTCGGGAAGACACTAATAGATTATTATCGGTCCACCGACTCTATCTTAGACCGAGTTCCTGAGACGTGT
 GGGCCCCGTCTCTGTGACTTAGACACCCAGTCCCGGACCGTTGTAAGGAACCTCGGAGGGGGAGACTCGGGGTGAGAAACCCGTGGG
 AATCCGAACGAGCGGGGGAATATTTCCCGAAACAGTCGGGACGGTAGGCGACATAAAGTACACGACCAAGGATCGATTCTCTTA
 ACCCGTCTCGTTTGGGAAAGAGTGGGACAGACTGTCAACACTAGGGCTGTCTTACGCTGGAACAAACCTCTTCCGTGGAACAA
 CCAACCGACACAAGACTTTGGTCCGAACCCCTCACAATATATAAATATATATGTCCGAATTTATATGTCTGAGTCCGAATACGTAC
 TTATATATGTCTGAGTATAAAGGAACCAACATTAACATATAATGGTACGGATATAAGCGAGTTTGTGGTAAAATACAAGTACGTT
 TTAACAACATCAGGAGTGATCTACTCTACCCGAGTACCTCCGTGGCCACGTGACTCGTCCGGGAGACGTAAGAGACGACCTT
 TATACACGAGTGACCTTAGCCTCATACCCCAAGGTCAATTGGTTACCTACGGTATTTCCGAAAGAGTCTGGAATAACGAAAACAAC
 GGGTTCGTGACCTGACCCCTTTATGATTTAGGGAAGACGAAAGAGAGGTGTGGAGGACCGTTTGGAGGGGTTGAATCTTTAATAA
 AATGTTCCCTCTGGTCAACAGAACTTTAACTCAAGACGCCCTAAGAAAAGTGCCCGTACCACCCCGACGTTTCGTCAAGCCTCATCA
 GTCCCGGGACCGGACCTTCGGAGTGACCTCGTACTTCCCTAGTCTACCAATTATTCACCGGAGACCTGTCACTAGACCCCGT
 CGACGACCGGTCCCACGAACCTCTTCGAACACCTTTCCGTTCCGTTTCCCATCCCTACGTCACTTATAGACAAAAGAGGGGAC
 GAACGGGTGCGCCGGCGAGTGGGGCGGGTCCGGCTAAACGACGTAAGGACCCCGCTCAACCCCTCAAGAGATTCAGGGGGTT
 TATGGGTCTAGACCAATGACCTACTTGACGCAAAATTAGGAGCAGCGATAGACTCGGTCTGTCTTGACCAGTCCCTCTGCCGAC
 CAGGCCGTGCTGCGCTCTCCTGCGACCTGGGCGACGGACACCGGACCTCGTGACTCATACCCGATGGTTCGACTCTTAACCTACC
 CGGAGGAAAGGGTCGGAAAAAGGGAAGGTCTCTGTGTTTTTCTGTCTCGTCTCCCTTTATAGTTACAGTCACTTTTGGGGTATT
 TTCTTTCTCGTCTGGTCTCGCCCCGACCTCCCCAAAAGGCGATGAGAGAGATCCACTAGCCCCAAGGAGGTTCTGAGAACAGGTTCT
 GTCCGACGTGGTTCTCCACGTCGAAAGAGGTGGAAGGGCGGGAAGTTCTTCCCGTCGGAATCCGAAAGACCCGGAACCCAG
 TGGGGTCCGTGACCCCACTGTAATGAGGTCTTAACACGGGGAGCGGGTCCGGTCCCGGAGACGGTGCCGCGCGCTCGGCAGGACA
 AATAACACAGACACTCAACAAATGAACAAAACAGACGACGGTGGGACGCGAGTCTTTGGGCGGTTTGGTTCGCTTGAGCGGGTCTCT
 TGTTTTACGACACGATAGGGGCTGAGGGAGGGGAAAGGTCTTGGGGGAGGACGGGAGTGTTGAGGACGTAAGAGGGAAAGCCT
 CCCCTCCTCGTCTTTATCAGAATTTTGTCCCGTCGACAAATATCCAGTCCCTTAATACCCCTGTGGTATTTGGAGACCCACAG
 GGACACCACTGTCTGAAGACTCGTCGGGAAGGGAACGGATCCGAAAAAGGAAGGGGACCACTGGTGAACCAATTCGCCCGGTG
 TGAGTAACCTTTCAACTCGGGTGGGGGTAGTTGGGTGAGGGGTGATGGGGTGAGGAAAGACGTCCTCAAAGGGACTCTGTGTCC
 AACAACCCAAAATTTAGTCCCATTAACACCACTAAGTATTTTCGATCTATATCGTACAGTAAGAAGAGGGGTGTGAGGGAGAGGAGT
 CGGGGGGTTTTAACGTCACTTAAATTCGTTTAGGATGTTTTAAATTTAACGGGGAACGATTTCTAGGATATACAGAGGTTAGGACG
 GTCCACATCTCAGTCTATCAACCGAAAACCGGAAATTAACACGACGATCGATAATTTAAATTTTCTCATAGTAACCTTCAATTATTA
 TAAGAGTAAAGACTTTAAATTAAGGGGAGGGTCCAGAGAGAGAGAGAAAAACGGTCGTGTCTAATTTACCAAGAAAGTCTAGAC
 GTTTGTTATTTTCTATTTTAATAAATATCCCTTTACTCCTACGACAAAATAACAACCGTATAACAAGGTCTTGTATTGTTGTA
 TAAAGTTCGTCTCTCCAATCTGGTCCCCCTCTTTCTCCCCCTTTCTTTTGTACTTTCTTCCCGTCTTGGGTGTGTGTTTT
 GAGGACGGTACATATTGACCATCTGTAAATGTTAAAGATATACTGTTGAGTGACACTACAGTATTATGGCTCAAGAACTTAAAT
 ATTAAGAGTACTTGGAGAGTTTACCCAGGTTCTGCTACGGAAGTACACTCCCATATCTTCTCGGTCTCCGACCCCTAGGAACT
 GGGTACGTCCTGCTGTTGGGGATAACGCCGAAGAAGGAGTACACTGTAGGAGAACTCCGGGATAAGACCCCCGACTTTTATAA
 GACCTATAAGACGTAATAAATAAGAGGACCCGGTTCGTCCCGGTGAACGAAACGGGGGAGTTACAGGGATCCGTAGATTGCA
 GAGAGGGGTAAGGGTCTAGTGTCTCTGACACAGGGGTCCGAGAGGGGTTCGGGGGAGGAGTCGATTATGAGTCGACGAACCTTAT
 CTCCACCTCCGAACCCGTGGTCCGGGGTTTTTCCGTCTTTACTAAAAAAGGTCTACCTCGACCTTCCCCGAGTCCCGACTCGCA
 CTACCCCATCTCTACAAACCTGGATAAAATGCAAAATTGGAATTCAGTCTCTTCACTTAACCTTTCGGTGTGTTTACGGTTACAC
 TCTCAACATAGACGACTACCATCGTCTTATTCACTTCACGACCATTCCTTTCCAGTCCGTCTTTAGACCTTTCTCTCTCCCTCC

TCTCCTTTTACTACTGTACTCAATCGTCTCGGGAAGGTATATGGAAGAAGAGGCCAAATTAACATTATTTTCGTTTTTCGTCGAGCCCAT
 CGGAGAGTTGGAGAATCGAGTTTTTCTCCTTAGTGGTCCGCTCCGGAACCTCCACAAACCCCGTCAAGAGTCGTCGAATTGTACGG
 GTTAGACGGGAGAGAGGAAAGGGGGAGGGACGACTAGGGGTTCGAGGGAAGTTGACCAGTCGGGCCCCGGACAGGGGTTTTATAA
 TCGTGGGGGTTTTACTATAGGTAAAGACACAATCCACACAGTTTAAAGTGAAGAGAGGTACGGATTTGAAACCTTTCCTCCTCACATG
 GAAAAACCTTACGCCACACACATCGCCCCCGTCCATACACAGTCTGTCTCGAAGGGGGTAGTATAGATACCTTTTACCTTCGGGGA
 ATCCACTCGGTACGAAACCACAAGGGTTTGTCTGAAGGGACGTAAGTCCAAGAGACCGTCTGGTCTCCGAGAAACGCTAGACTAG
 GAAACTGATCCTCAAGGTACCTCTAGTTTCTCGAATAAATATGTGTTTTGTGGTCCGTAAACCTCATCTTAAATAAAAAATTAA
 GATTTATTTTTTTCAGATTAGTATCTGGTTAGCCCCGTTGACTCGTCAGCGAGGTCTTGTCGTTTTCGGAATTGGTTTTATCACACTAA
 TACTCCCCCTCGACTCGTTAATGACTTTGAGCCGAAAGAGGCATTCTCCCTAAAATCGGGAGCGGGTCACCGTGTTTTCCCGCGTGC
 CCCCCGTTGGCAAGAAGAGAACGAGTCTTTTCACTCTGAACCCCCACCACCCCGTCCAGCCCGTCGACGCCAGCGCAGGTCCGTC
 TACTCGTGGGTCCGGTCCGCGACTCCGACTCGTGACTCAGGACTCCGAGGTACAGAACCCAGGGGCTCGTCGGAGTCCCTCCGTGGC
 TCGACATCAACGTCGCGCCGTAGTCTTTACCTCATACACATAGACACACGCTCCCAACCACTCAATGTTTGTAGGCGCGTCTCAA
 TCCCCCTAGACCCCTCTCTCTCTCTGTCCCTCTTGTGTTGATTCTCTCTGGAGACCTTTTATAACAGGGGTTTTATTTTTAAGGTCT
 AGAGTAGGGAGAGTGAAGTTTATAGGGAGGGGTCTTTAATTATATAAAGGGTCTCGACTCACCTTTTCGTTTCCGGTCGTTCCACT
 CCGGACGGGTGCTTGAAGAAGTCCGACCCAGGTACAGGTGGGTGGGAAAGGGGTCAAGTCTCGAGTGACGGGGGTTCGAAAGGGA
 CCGGATTTCCCTCATTTCCGACCTCACCATCGTAAATTGAACGGATCCCTTTAGGGGTTACAGGGTCCGACGTTACGGTGGGAATA
 GGAACCCCTGAGGTCCGGTCACTGACGTGGGGTGAGTCCAGGTTCGCGTCCGTCATTAAAGAGAGGATATTAATCGCGATAGTT
 TTCGTACACAAAACGTTAACCTCTCCTGACTCTTCGTTACAGCTTAAAAAGGGAACATAAGAAAAAGACGAAAAAACTTAAGACA
 AAAACTTAGGTGATTAATTTTGAAGAGACTTTTCACTTTTAGACGAGAACTGTGGTTCGGGGAAGGCTTACGGAATAAACCTCCGA
 TTCCCCGCTCTGTCATCGTTCTCTGGACGTTCCGACTACCCCAACCCCATCACAAAGAGACTAAAGAAATCCGAGCAAAATAT
 CTTTGTCTCTTGTTCAGAGAGGGGAGGAAGGGTCTGAACACGTATAAATGTCTCGTGTGACGTCAAAATTATTTTGTAGACAA
 GAACGAAGACGACTACTCAAAGGTATTAACAAAAGTCTGTTTAAATTGTCCTTTTATATTTATATAAACTTTTTTAGATTACAGGT
 TTCAAGGGGGATATCTGTCGGTAAGAAAACGGTTTTTTTAAATTGTAAGTCTTTACCTTTTTCTTTTATATTAGGTACATCCC
 CCGTAGCCCGAGTAGCGACGAGAGGGGGAACCTTCGTTTAGGAAGTACGTTTTAAAGTCGACGTATTAATTTCTAGATTCTCTT
 TCTTCTCCCGAAGAAGAAAGGAGTCTTACTTTCCATTAAACCCCCCCCCCTCCACAACCTTTTAATTAAAACGGTTTTCCAGAAATT
 CGTCACCTCCTTAAATAAAACATAAAGACAAAGGGGAAGAGAGGGGGAAGGTCTGTAAGACGGAAGAAAGTGGTCTCGACCAAAGA
 CAAGTAATATTTAGCCGCTCTGGAAAGTCAGAGACGGGAGACAAATCCCTGCAGTATTTGTGAATTGAAAGACCCGACTCCGAG
 TGTAAGACGGAGAGTCCAGCGGTTGGAGGGACTCGGGTGGCTGAAATCCGGGGTTATCTCTTAGAGGGGTGAAACCCCGGGGAAA
 GGAGAGGTCTGTACCAACTCCTCCCCGAGTTCTTTGGATCCGTAGGAACGACCCCGACATCAGGTACGACGGGGTCTTCGGGGA
 AGAAGAAGGTTCCAGGAAGGGAGGGGTGGGTCTGTATTTCTCTAAAAATTTATGTATCTTTTAGTGTGTAGTAACCTTCACTT
 TGGAGTAATCTGTAGGATAGAAGGTAGTAAGTTAAACAACATGTAAAGCTTGTAAATTATAGAAGCTTTCTTTATAGTCAAGTGTAC
 TAAATAGGGAGATCGTTAGAGAGAGACTTTAACCTCCCCCCTGTCTCAATGAGAGAGAAGGCCGAGAGAAGATGAACAATGTGGTG
 ATGGGAATCCCGTCTTTCTCCGACCCTTGAGACTCTCTCCGATGAACACAGGTCCACGTCGACTGAACAAGGGATCGAGTCAGG
 GGTCTCATGAGGACGGAACAACAGAGAGGGGTGGAGAACGGGAGGTCCCCCGACTTCGGAGTCGGTCGACAACCTTTGGAGACG
 AGTCCGTACGGTACTTAAGCTCCTGTGACCTAAGATCCCGTTTTTCGTTGAGTGTAACACAGATACAAACACATGGACGGTTCA
 GTGGGGTTTTTACTCTATTGAGTAGATTTTAAAAAAGAATTAATTTTTTAGGAGGAGATGGGAAAGGGGAAGTAAAGGAACCC
 TCTTCTTTCACACTAGATCTTCTCCTTGGAATATTAATATCCTCAGAAAGGGGAAGGGAAGGTTTATGGCTGTTTAAAGACAGG
 AAAGACCTCCTAACCCCGACATAAAAGATTTTGGTTGTTTTAAACAGACGCTCCTTCTGAGAAGTTGGGGGATTGCTTCAAGTACA
 TAATTGATTGTACAACAGAGGTGGTTCGAGGACGTGTGACAGACGAAACCCCAAGCTTTGAAATAAAAAAGGGGGGATCGCTGTGAGG
 TCCCTTTGGAATTGCAATGTCTTCTACTTATTTGCTCAAAAAGAGTCGTATCAGGGCAAAATACCGAAATAACGATGGGTAACTAA
 TGAGGCGAAACAATGTGCTTCTCTTAGTATTTTAGGCCGCTGTAGGCCAAGAACAATATCGGTTCGAAAGGGGGGGGGGTTG
 GGTGGGTTTTTACCACCCCCCCCCCTATACTCGAAAGAGACGAGGGGCGTGTGAGGGGTGGTGGGAAGAAACAGTTGAGTTCGG
 TAAAGTCGGATATAGTGCTATAGTTTAATTAACACCTGATTAGTCTCTTTTGGGGGTAGACGTGTGAAAGGCGACAAAGTCAAG
 CGGATGACGCCGAGGAGTGTTCGGAGCCTAAGTCGACGAGAAAGCATGCGAGAGAAGGGAAGAAATCGGGGTTTACCCCTT
 ATTTTACCCGCAACTCCAACGACTCTCCGGGTAGGGGTGGTGCCCACTGACCGTGAACCCACTCTCGAGGAATCTTGGGGAAGAC
 TTCAGAGGTTTCGAGTATCGAGTCCAGGTCCCCGCGGAGGTGACCGCCACGGGACTCAGACCTCCCCCGTCCCCCGAAGACTGGT
 TCCGTCCCCGAACCTCGGACGTCTTCGATTGACAGGTAGTCTCGGGGTGGTTCGGAAGAGAAAGACCGACCTAGGATACACTTAATG
 GGGTGTCCGGGAGTGACCGAGGTTTATTTAAGAGTGGTTGTATTGTCTCACTACAGAGATCTTTCTAAGAGTCTCAAAAG
 CAAGAAGGAGAGAGAGAGAAAGAGAGACAGAAAGGGGGAGGCCACGGTGGAGGGATGTAATGAAGAAAAAGAGGAAAAAGATAA
 CGTAAAAAATATGTAAAAGGGAAAAATCCGTTTTTACGACCTTCTCAAAAGGGTTGTCTCCCCAGGACTCTTCTCCGTCTTCCCTC
 GGTCTCCGTTTTCTCTCGTCTACGCACCGGGTAACCGGTGAGGTGAGGTGACATGGGACTTCAGAACCGGGAAATCTGGA
 TCAACAAGGGGTGTAGGGATGATGTAAGTTCTGAGAGTCGGTGTGAGATGAATGAGTCCCTTGTCCAAAATACTTTAAAGAAAGA
 TGAAAACATAAAAAAATTAATAACCACTTTTTTTATGATCAAATCACTACCCTGAGTGAGGTACCTCACAGTTTTGTCTTAAAGG
 ACGTCTTAAAGATCCCTGGGGGAGGACACGACCTCCCTCAAGAAGGGAACCTCTTAGTCTACATTAATAATCTTCCCCGCTCT
 ATATTTGTACTTTGGTCTAACTTTTATGTCCCCAAAGAGATCGGGACCGAGACCGGTGTCTGTCTGTTTCAATCGACCCCGACTC
 TCTAAAGGACCATATGTCGAGAGGAAGGGACTCAGGGTTCCGGGACGAGTCCCTGTACGTGGGGAGGAAGTGGTGGTTTGGTCAA
 GGAATGTATATATATATATATATAAAAAATTTCTTTAGGTTTCAAGATGAAGTTTTCGTGTGTAATCAGAGTTGATCCCT
 TAGTTGCTTTCGTCTTCGCTAAAAAAGGGGGAGGACTGTAGACCGAAGCTAACCAGACCTCCCCCAGTCCGACTGAAACAGTAA
 AACAGACAGGACCTAACCTCGGCAGGGATATTGGTAGATCAAGGCTCATGTTTACCTCTGTCTTTATTTATAATTTCTTTAGTAT
 CTGGCTGGTCCATTTCCGTTTCCCTACTTAAGGATGAAGTGATTGGGAAGGAATAGGACGGTGGAGCGGCCCCCGGTCTGCAGGA
 GGGGTTGCAGCGGGAGTTAAGGTGGCGGATACTAGGTCACTCCGTAAGAGCTGGATACCTCGCCGGCAACGGGTCTTGGCCTAGA
 TGAGCTGAGGGAAAAATAAGCGGTGTCTCTTACAGCAAAAGTCAAGGTTCGGCCCCCGGCATACTGATACCTAGATTAAAGGAAATG
 GTCCTCTTCTGTACGAGAGTTTGACGTCTGTTTTGTGGAATCCTGTATTGTGTGTCTGGAGTTAGCGAGTCTTAAATCAAGACT
 CGTCCCGTCTGACGCGGGTCTGGTCTTTCCGTCATAGGTCTAAATGGGGACCTACGTGCTTACTTAAGCGTGTCAACACTCA

CGAGAACAGUCCUCCUGUAAGAGCCUAACCAUUGCCAGGGAAACCUGCCUGGGCGUCCUUCAUAGCAGUAUUUUUUUAAA
UUAUUCUGAUUAAUAAUUAUUUUUCCCCAUUUAAUUUUUUUUUCCUCCAGGUGGAGUUGCCGAAGCUGGGGGCAGCUGGGGAGGG
UGGGGAUGGGAGGGGAGAGACAGAAGUUGAGGGCAUCUCUCUCCUCCGACCUCUGGCCCCCAAGGGGCAGGAGGAAUGCA
GGAGCAGAGAUUGAGCUUGGGAGCUGCAGAUGCCUCCGCCCCUCCUCUCUCCAGGCUCUCCUCCUGCCCCCUUCUUGCAACUCU
CCUAAUUUUUGUUUGGCUUUUGGAUGAUUAUAAUUAUUUUUAUUUUUGAAUUUAUAAAGUAUAUGUGUGUGUGUGUGGAGCUGA
GACAGGCUCGGCAGCGGCACUAAGAGGGAAGACGAGAAGAGAGUGGGAGAGAGAGAGGAGAGAGGAGAGGAGAGUGAC
AGCAGCGCUGCUAGUAGUUUCCUUAUUGGUUCAAUAUAGUAACUAAUUGUCUGCGGCUUACAGAAUAGGACAGCAGCGGCUU
UCUUGGUUAGGGAGAAAGAGGGGCGUGGCGAGAGGUGGGCGCGUAGUUAUUUAUGGGGAGGAAAUAUGCGGGGGGAGAGAAACG
GGGUAAACUGGAAUACCGGUAACUGGUAGUGGGGCGAUUGUGUUGUCCUGAUGUCACUGUGGCAAGUUGGGGUGGGGAGGCA
AGGGAACAGCACCUGUGGGGCUAAGCCGGCGGUCUAGGGCGCCAGCUCUCUGCUGCCCACUGUGUCAGUACCUGCUAAUUAUAGA
UCUUCGGUCUCUUGGCUAAUGGAACUGGGUCGAGCGGAACUGGAAGCUCAAGCUCACAGGCAGGAGAGAGAACUUAAGGCGUGGC
CCAUGCCCCCUAAGCAAAGUAGCCCCCAACCCCAAAACAAACUUUCCAGCUUCCCCUCCUGUCUGGCAGCCCCCUCUCCUCC
AGAGAGCGCGCAGCUGUAGAGCUCACACAUGCGGAGCAGUGUGGGCCAGGGCCGGGCGCCGAGCAGGAAGCGCGCAGUAGGCGG
CGCGGGGCGCUGUAAUUGGCAUUUAGGGGCGAGCGUGGUGGCGUGGCGCUGCAGGAGGAGAGCUGUCGCCACCCUCCUGC
UCCCGCCCCACUCGCGCGGAUGGAAGGGUGGGAGGUGCCUCGCUUGGGUGGAGGGUGGAGGUUGUAGGGUGGGGGUGGGGAUG
CUGUACUCAAAAGCCAUCUUGUGCUCAGAGAAAAGAGGCCUACCGGCUUUCCUCCGGGGUCCGGCGCCCCUACCCCCAGCCG
GGCAUCCAGCCGGGAUGCCCACUGGACCGGAUGCCCGCUCGCCACGCAUGGCUGCUCUGGGCUAGGACCUGCCUCGCCUCGGG
GGGAGGUGGAAAUGGGGAAGGCCUAGGCUGGGCGGGGACAGGCCAGUCCUGGUAGAGUCUCCUAGUGUCUGUGGAGGCGAGU
GAGGAUUCGGGGGCGUCUGGCUUAUGGGGUAGAGGCGUCAGCCUGGGGCGGCGGAGGGUCCUGGGCACUCGGGAUCUCUCUAAU
AAUUGGUGCUAUAUGGCCUGUUGUAUAGUCUUAACCCGACGCAUAUGGUUUCAUAAACACAGUGGCUUAUUUCCCAAUAGUACGU
GGCCUGAAGUUGGCGAGUUUUAUUAUAGAACCCGCCUCGUCUCCCGCCUGGUGCCAGGCGUCCAGGCGGCCCAUAGUGUG
AAUUUCUGUUUAUAAACGCGACGCACACGCAAAACGCCUACACACCCAGGCAGACACACCCGAGGCGUGGGGGAACAUAUCCAG
ACGAGGUCCGAGAGAACCUGGGGCCAGACUUAACCUAAUUCUAAGCCAAGACACCCCCUUAACCCCAACCCUCUCCAGAGGCGG
ACUGGGAUUCUGCGCCAAGUCCAGUGGGGGAGGGGAUCGCUCCACGAUUUCUGUUUAACCCUGACCCGUCGCCUGUGGGGGG
CGGGGAUGGGGGAAGAAACUGUUUUUAUCAAUUUUUAAAAUAACAGUUUGCAUUCAGAUCAAACCGUUUGGCCAGAGCUGACG
GUGAAUUUAUCUCCCCUCCCCCAACAACCCCUUCCCUUUUCUGCCCCGCUAAAUUCCCUUUCGCGACCCAUAUAAAGG
CCUCUGAGGUUAUUCUCCCGCGGAAGAAUAGGCAUUUUUCUCCUCCUCCCAACCCCAACCCCAACCCCAUAGGACUUGUGUGCG
CAGAGGCGUCAGCAGGGGGCUAAUUUCGUUUUCCUUGUUUACGACUAGCAAGAAAGCGUUGGCUUCCCCAAUUGGGGCCAGCCG
CCGCCUGCUCAGGCUCGCGCGCGCUCUCCCGCCGUCAGGCACGGCCAGUGGCUGAUGCCCGCGGAUCCAGCCUCGCGGAGGCC
GGAACCCGGAGUCUGUGGGACCUGAAAAGGGCAAAGGGAGUGCGGUGGGACACAAGAGGGGGCCGAGAGGCCCGAAAAAGGGAGCC
UUUCUCCUGGAGUAAUGUGGGAAUCGCCGACCGGUGAGGCCUCUAAUUUCUCUUUUCGCCUGCAUUUUGUCCGCCCCCGCCUUGU
GGAUGGCCCGCUGAUGGGGGGGAACAAGUGUCCUUGUCCGUCUCCCGAAAGAAAACCUUUUCAGAGAAGGCCAUUCAGGCCAGU
GGAGAUUCACCUAGAAUAUAGAAGGGCCCCUCCACCCGAGGGGCGGGCCUGGGCGAGGCUGAGGUAAUUUCUCUUAUUCUAGA
ACGACCCUACUGAGUAGUACCCAGCGAGAUUUGGAGGUGCUAGAGUUCGAGGUUGGCCGCUAACAGCACCACCUAGGAGGUCU
UUAAUAAUAAUUUAAUUGAAUUUCUGUCCUUCUGGCUACAGCCUCCAGCUCUCCCAACCCCAUCCUAGUCCCUUACCCCUA
GACCCAUCCUUAACAAGACAGAGGCAAGCAGAGAAGGCAUAGCAGCAGCAGCCGGCGCUCUGUUUUAUUUCCACUCUGGCCAGG

Secuencia traducida del gen:

AEQSSLEPNHCQGNLPWALPSLAVFFFLNSDLFFPHLIFFPFRWSCRWSGQLGRVGMGGERQKLRASLSSFPTLWPPR
GRRNAGAGVELGSCRCLRPSSSLPGSSSCPLLATLLNFVWLLDD
YNYFYFYIYIKYMCVCVELRQARQRHRMREDEKESGRERGREGERESDSSARVFPYWFKYVTNGPWRFQEDSSGFSWL
GRRGGWREGGRVNYGEENWRGERNGVKLEYRLVVGRLLCCPDV
TVCQVGGGEAREQHLWGSSRRSRAPALCCPLCQYLLIIDLRSLGWNWVGAELEAQAPGRRENLEAVPIAPAKLAPQP
PNKLSSFPSSLGSPPLPPRERDCSSHMRSVGPGRGRRAGSRRSA
AGGACLAIRGEAGGWCASAERRASAHPLLPPLGRMEGWVPCVGVWRVEVVGWGWGMLYSKAILCSEKRGLPAFPSPG
VRRPSPAAAIPAGMPTGPGCPLATHGCSGLGPASPRGEVEMG
KALGWAGTGPVLLVESPRCCGGSEDSGGCWLMGRRQPGAAGGSWALGISLIIGANGLLYSLNPTHMVSHSGLISSKC
TWPMCCSLIYDPALLPGLGPRLRGGPISVNFCLTATHTQTPH
PGRHTRRRGGTYSTRRGPREPGGQTYLNLSDQDTPSTPTLSQRRDWDSAPSPSGGGDRSTISCLNPDPSVGGRGWGKK
PVFIKFLKIQFAFRSNRLGQSRILSPLPPQPPSLPFLPLNSP
FAAIIKASEVFSRGGKWHFLSPSPTPPPVGLVCRQRRLTRGFRFPCLRSRKALGSPQMGPARRLLRLPRALPPLRH
GPVADAPRIQPAAGRNPESVGPEKKGSAVGHKRGPRGPKKA
FLLECGNRRPVRPLFLSFACILSAPAPVDGRWGGTQVSFVRPRKKTFSREGHSGQWRFTNIRRAPSHPRGGPGRGGI
SLNPRTTLSVPTARFGGARVRGWPLNSTILRCLYIIFLIFCP
FLATALQLPHPHMPFTPRPILTRQQAQAEKQORPALCFHFPLWPGDKLDPSGLQHQLSASQLLSIFFPFSFFLSLS
GIQTEAQEDPPVATLSLAGHIPCRGPPSCRRALLSNNSQGG
NLAQGLCTPGAETLNLWVRAWQHSLEPPPLSPLFGHPACLPLGALSALPSAVFHVLPVSGELGRANPFSPWSDSCD
PDRMRPCLEKATLLGWLCSETRLGSVIYLLYTGLNIQTQAYAI
YTDSYFLGGLYITMPIFAQTFFYVHAKFCSSLDESGLMEAPGALSRSVHALLEICAHWNRSRMRVPVTNGCHKGFLR
PYCFCCPSSWTGEILNPFCLSTPPGKLSQLRNYFTRETSCLE
IEFCGILFTGMVGLQAVRSSQGPAGSLTGAMKGSEWLIRWPLDSDLGQLLARVLGEACGKGKPKGVGMQISVSPPA
CPSRPLTPPQPDLPASVAADWEFSKVPQIPRSGYWMNCVLILV
AIARTRPGQGDGWSGQHGRGRWTRCLWPGALSMGYQGELMGLLSQPFSLPETQKDRAEGKYQCQWKTPKKAGTRAGL
EGFSATLSRSGSSQDSCPDRLHQEVCSFLPPSSRPSRRAALGF
LGLGHPRHWDITPGLCPSPSPRPLPRRRSRSCLLCLVVYLFCLLPCCVRNPPNQRTRPGEQSAVLSPDLSLFPATP
SCPHKLLHSPFRRGGSRLKGTGQLFIGSGELWGHKKPLGVPV
VDRLLSSPSLAFLPSPGDHLVKGATLIEKVEPTPINPVPHYPTPFCSRFPETQVVGFIRVMLVDSSIHVILLPHSL
SSAPQNCSEFKQILQNIAPLLKILYVSNPARCRVRLAFLGCC
LLNLKEYHSHKYSHFNFSPPSLSLSLFASTDMVLSDLQTIKDKIYREMRMLFYCWHIVPEHNNILFSSRRGTRGRKR
GKGKMMKGGQEPHTKTPAMYNWDIYNFYMTTHCDV IIPSSILF
STSQMGPPSSASC GGKPGGWSFDPVQSSTPPIAASSSCDILLRPYSGGKYSGYSAFFFSWAKQGHLLCPPQVSLG
IALSPFPDHLRELCPQALPKPPPQLILSCLERWEAWAPAPKKPE
MIFSRWSWKAQGAQGRDWDLFYVLTLSQRSELRSQTNANVRVCICWQQVKCWGKVRQKSGKREGGEEMHELAEPF
HIPFFSVLSKSSSSLSLTKRGITSAGLGGVWQGFSAHAQS
ALFLLSPSLLIPSSLQLVSPGPVKILAPPNDIHFCVRCVKFHFSPCLNFERRSVPFWNAVCVAGAGMCSDFSPHHI
YGKLEAPVSHALVFPNSIPCIQVLWQTRRLFAISFDEFHGDQR
SLFIHKTPGIWSRIYFLISKSLIIDQSGQLSSRSRTAKPPNSVIMRGAEQLLKLGLFLRKRDSPRPVAQKGPAGQP
FFSCSEKDLGVVGAGSGSCGRVQAAEHQALRLSTESGSSAW
VPEQPQEAAPSCSCSAAVRNGVCVSVCEGWVTNIRAEGLDGRERRQGEHNGGPLENIVPKKFQISSLSLQISLPRKL
IYFPELSGKQRPARGLPNTNFFRLGPVPPTLSPVQSSHAPSLSL
ARGVRLEWHLTCLGKSPSAQAAMPPLSLGTPGQLHPTQVQAGSKFLSYNRYQKHVFCIGEDWRSKWQFFPCILFLC
FLNSVFESTNSFSESENLLTPSPFRMPDLGGGARAVARRPAS
LMGLGGSVLFLARFIETKNKFSPLLPKTCAYLQSTTAVLIKHLFLLLLMSFHNCFTQNLTGKYKYIKNLSPSFPPID
SHSFAKKIKLTNGKKKISNVGKIGLIPALPPWSESFQQISAAL
KIEKEEGFFFPQEKVIGGGGGVENFCQRSLSSGGIYLYFCFPFSPSSISAFFTQQLVSVHYKSARPFSLCPLFRDV
IKHLTFWGGSHSASQVANLPEPTDFRPQRISPTLGPLSSPDMV
EEGLKETASLLGAVVHAAPAPSFFQGPSLPHPDIKEIFKYIENQHSLKVKPHDILSSIIQFVVHCNNLISRKYHHD
LSLQSLSEIGGGDRVTLSSGLFYLLHHYPGRKEAGNSERGYLC
PGVHLVPLSPQSTPALLSLPHLLPSRGPEASASCWKPLLHAMNSRTAGFGKKQLTLVSMFVYLPSPHPKNEITHLKF
FFLIKSSSTLSPSFPWGEESVIKEEPLLESFPPFSKYRQISVL
SGGLGLYFLKPTQIVWQEDSSTPRGSCINHVSTSSCTLSAFGFETLFFPPSDTPGKPRYRRINEFFLSIVPFMALL
LPIDYSALLHRRKRSNPATSGFLLPAFPPPNPPKNGGGGIAF
SAPRTLPTTLLCQLKAFSAYITISNVWTNQEKTPICTLSAVSVRLRLRPHKPRIQLPFRNRLFPFFSPKMGENGRC

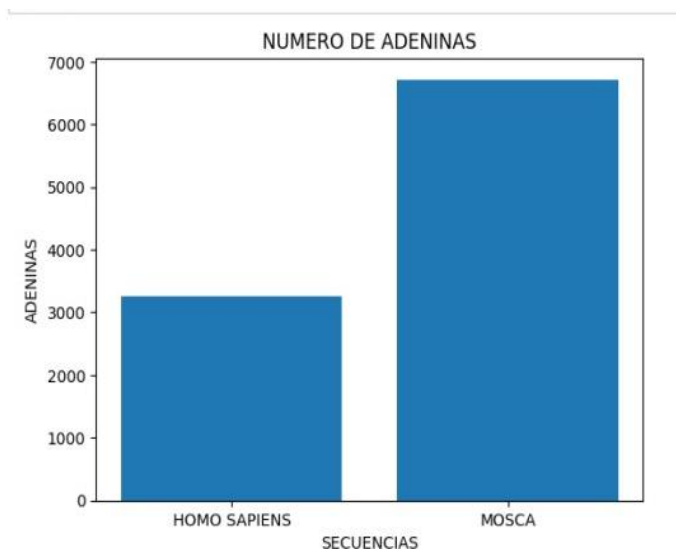


Figura 17. Comparación del número de adeninas del gen HOXC6 del Homo sapiens y abd-A de Drosophila melanogaster. (Fuente: el autor)

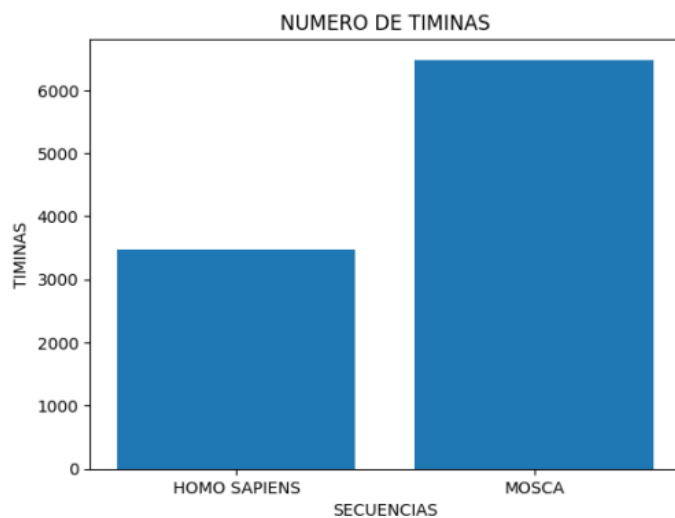


Figura 18. Comparación del número de timinas del gen HOXC6 del Homo sapiens y abd-A de Drosophila melanogaster. (Fuente: el autor)

Para el cumplimiento del segundo objetivo, se encontró que el porcentaje de guaninas/citocinas del gen HOXC6 fue de 51.68%, además el largo de la cadena del gen HOXC6 fue de 13936 nucleótidos, y del gen Abd-A, fue de 22886 nucleótidos.

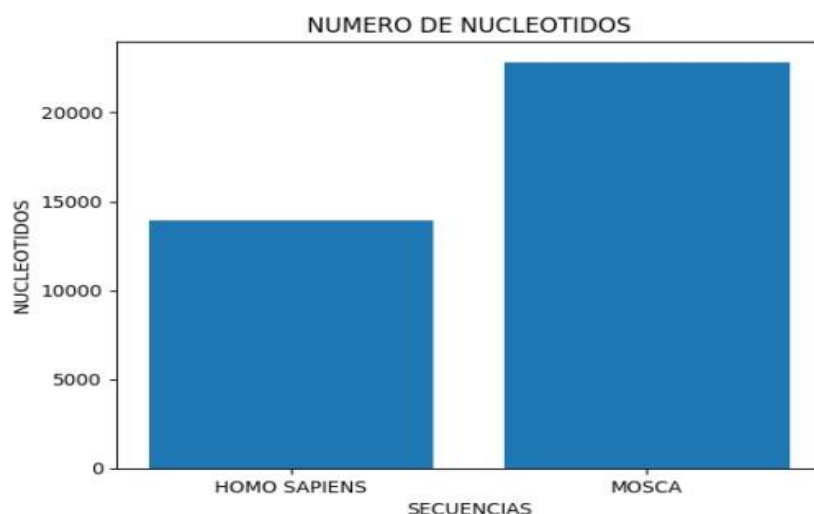


Figura 19. Comparación del número de nucleótidos del gen *HOXC6* del *Homo sapiens* y *abd-Ade Drosophila melanogaster*. (Elaborado por el autor)

El número de nucleótidos de *Drosophila melanogaster*, ($n=22886$) fue superior al del *Homo sapiens* ($n= 13936$)

Para el cumplimiento del tercer objetivo se encontró que el gen *HOXC6* está localizado en el cromosoma 12q13.13

(Cromosoma 12: 53,990,624-54,030,823. cadena delantera. GRCh38:CM000674.2) el gen tiene 4 transcripciones,

219 ortólogos y 42 parálogos . Se analizaron 2 transcripciones del gen.

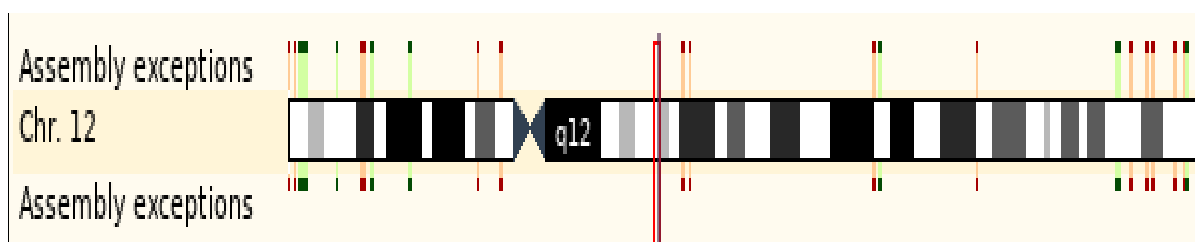


Figura 20. Ubicación del gen *HOXC6*. Ubicación del gen *HOXC6* en el cromosoma 12q13.13. Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). (Disponible en: <https://www.ensembl.org/index.html>)

Tabla 1. Transcritos del gen HOXC6

IDENTIFICADOR DE TRANSCRIPCIÓN.	NOMBRE	PB	PROTEÍNA	BIOTIPO TRANSCRIPCIÓN.	COINCIDENCIA UNIPROT
<u>ENST00000243108.5</u>	HOXC6-201	1651	<u>235aa</u>	Codificación de proteínas	<u>P09630</u>
<u>ENST00000394331.3</u>	HOXC6-202	2943	<u>153aa</u>	Codificación de proteínas	<u>P09630-2</u>
<u>ENST00000504315.1</u>	HOXC6-203	538	<u>51aa</u>	Codificación de proteínas	<u>D6RBH4</u>
<u>ENST00000509328.1</u>	HOXC6-204	358	<u>71aa</u>	Codificación de proteínas	<u>D6RC34</u>

Fuente: Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl).

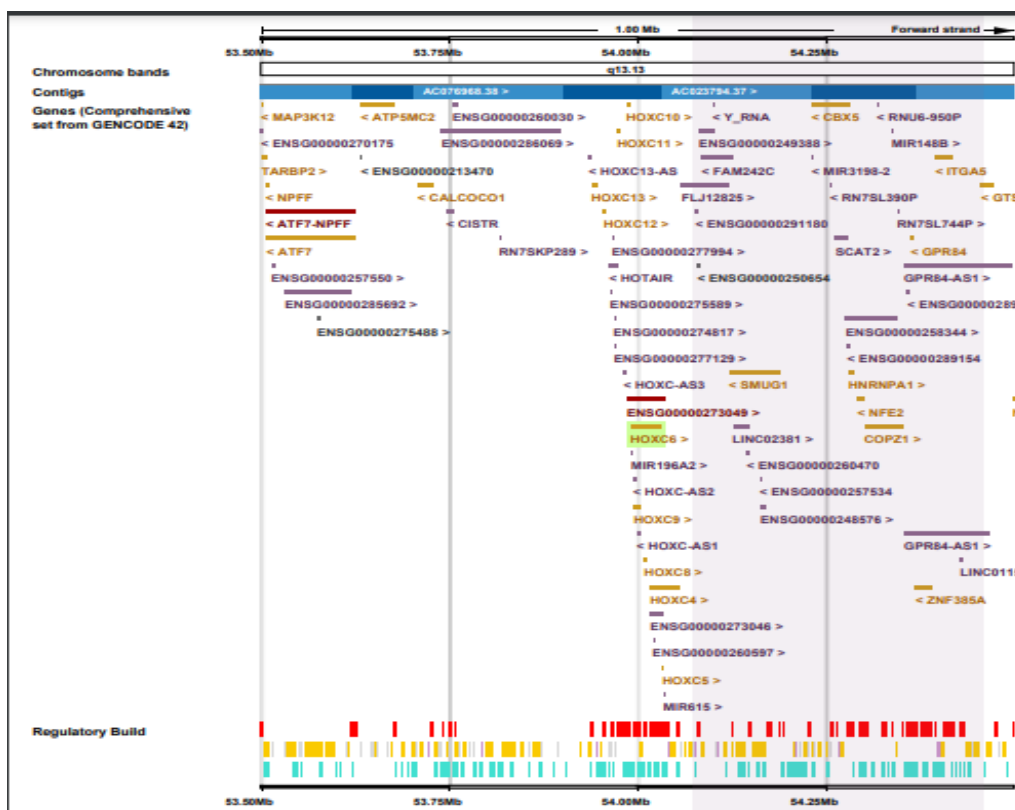


Figura 21. Análisis del transcrito HOXC6-201. Se identifica, la banda cromosómica, los contigs, la ubicación del gen HOXC6. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

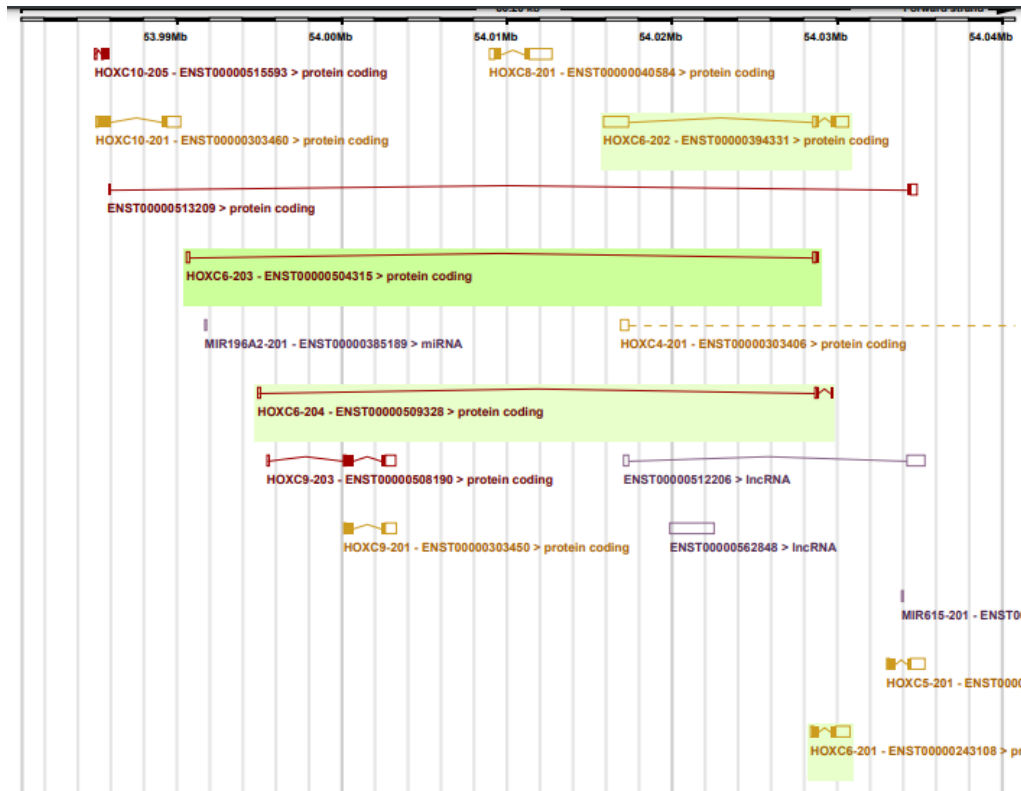


Figura 22. Ubicación del transcripto HOXC6-201. Se identifica la ubicación exacta del gen HOXC6. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

Mientras que el gen abd-A está localizado en el cromosoma 3 (Cromosoma 3.16,807,214-16,830,049) el gen tiene 4 transcripciones, 108 ortólogos y 12 parálogos. Se analizó 1 transcripción del gen donde se encuentre el homeodominio conservado.



Figura 23. Ubicación del gen abd-A en el cromosoma 3 de Drosophila melanogaster. Se identifica la ubicación del gen HOXC6 en el cromosoma 12q13.13. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

A continuación, la región en detalle:

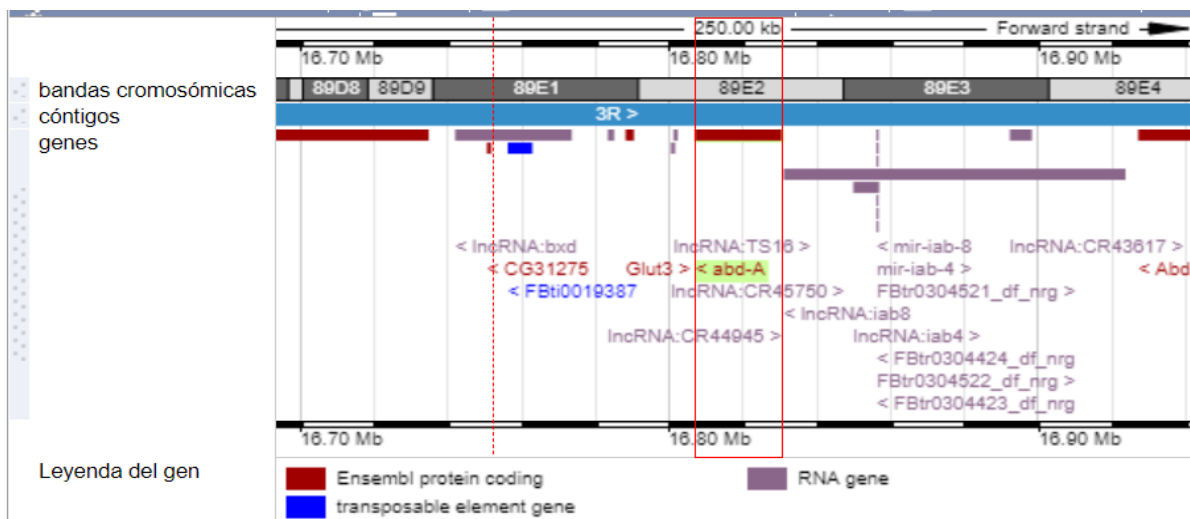


Figura 24. Región en detalle del gen *abd-A* de *Drosophila melanogaster*. Ubicación exacta del gen *abd-A*. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

Tabla 2. Transcritos del gen *abd-A*

IDENTIFICADOR DE TRANSCRIPCIÓN.	NOMBRE	PB	PROTEÍNA	BIOTIPO TRANSCRIPCIÓN.	COINCIDENCIA UNIPROT
FBtr0306337	abd-A-RD	5269	330aa	Codificación de proteínas	E1JIM9
FBtr0300485	abd-A-RC	4559	330aa	Codificación de proteínas	E1JIM9
FBtr0083387	abd-A-RA	4458	330aa	Codificación de proteínas	E1JIM9
FBtr0083388	abd-A-RB	3935	590aa	Codificación de proteínas	P29555

Fuente: Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl).

Como se detalla a continuación el análisis del transcripto *abd-A*- RB permitió obtener la proteína homeobox, tipo antenapedia, sitio conservado, que constituye un factor de transcripción específico de secuencia que forma parte de un sistema regulador del desarrollo que otorga a las células identidades posicionales específicas en el eje anterior-posterior. Se muestra características para motivo, unión al ADN, región. El motivo hace referencia a una secuencia conservada corta de no más de 20 aminoácidos de importancia biológica. (Secuencia: RYPWMT). La unión al ADN, determina la posición y el tipo de cada dominio de unión al ADN presente en la proteína, en este caso, el dominio

Para el cumplimiento del cuarto objetivo se encontró que el gen HOXC6, contiene 2 exones de codificación con una longitud de transcripción de 1651 pbs. separados por un intrón de 733pb, codifica una proteína de 235 aminoácidos. La secuencia ascendente 5" y descendente 3" se muestra a continuación.

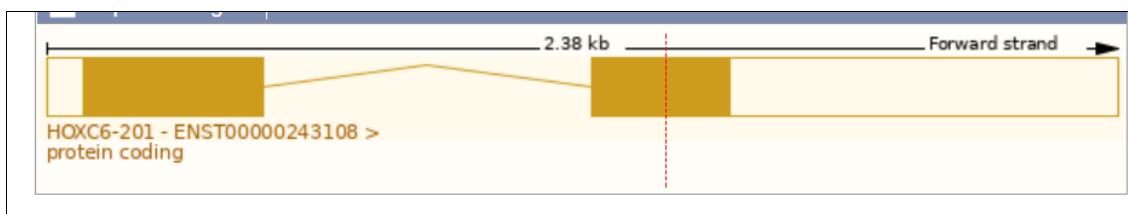


Figura 27. Segmento codificador del Gen HOXC6. Se puede identificar los 2 exones, separados por un intrón correspondientes al gen HOXC6. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

EXON 1. Longitud 482 pb

```
>HOXC6-201 ENSE00000838340 exón: proteína codificación
ATAACCATCTAGTTCGAGTACAACTGGAGACAGAAATAATTTAAAGAAATCATAGA
CCGACCAGGTAAGGCAAAGGGATGAATTCCTACTTCACTAACCCCTTCCTTATCCTGCCA
CCTCGCCGGGGGCCAGGACGTCCTCCCAACGTCGCGCTCAATTCACCGCCTATGATCC
AGTGAGGCATTTCTCGACCTATGGAGCGGCGCTTGCCAGAACCGGATCTACTCGACTCC
CTTTTATTCGCCACAGGAGAATGTCGTGTTCACTCCAGCCGGGGCGGTATGACTATGG
ATCTAATTCCTTTTACCAGGAGAAAGACATGCTCTCAAACTGCAGACAAAACACCTTAGG
ACATAACACACAGACCTCAATCGCTCAGGATTTTAGTTCTGAGCAGGGCAGGACTGCGCC
CCAGGACCAGAAAGCCAGTATCCAGATTTACCCCTGGATGCAGCGAATGAATTCGCACAG
TG
```

EXON 2. Longitud 1169 pb

```
>HOXC6-201 ENSE00000838341 exón: proteína codificación
GGGTCGGCTACGGAGCGGACCGGAGGCGCGGCCCGCCAGATCTACTCGCGGTACCAGACCC
TGGAACTGGAGAAGGAATTCACCTTCAATCGCTACCTAACGCGGCGCGCGCATCGAGA
TCGCCAACGCGCTTTGCGCTGACCGAGCGACAGATCAAAATCTGGTTCCAGAACCGCGGA
TGAAGTGGAATAAAAAATCTAATCTCACATCCACTCTCTCTCGGGGGCGCGGAGGGGCCCA
CCGCCGACAGCCTGGGCGGAAAAGAGGAAAAGCGGGAAGAGACAGAAGAGGAGAAGCAGA
AAGAGTGACCAGGACTGTCCCTGCCACCCCTCTCTCCCTTCTCCCTCGCTCCCAACCA
CTCTCCCTAATCACACACTCTGTATTTTACTTGGCACAATTGATGTGTTTGTATCCC
TAAACAAAATTAGGGAGTCAACGCTGGACCTGAAAGTCAGCTCTGGACCCCTCCCTCA
CCGCACAACCTCTCTTTTCAACCACGCGCCTCCTCCTCCTCGCTCCCTTGTAGCTCGTTCTC
GGCTTGCTACAGGCCCTTTTCCCGCTCAGGCCTTGGGGGCTCGGACCTGAACTCAGA
CTCTACAGATTGCGCTCCAAAGTGAGACTTGGCTCCCCCACTCCTTCGACGCCCCACCC
CGCCCCCGGTGCAGAGAGCGCGGCTCCTGGGCTGCTGGGGCTCTGCTCCAGGGCTCA
GGGCCCGGCTGCGAGCGGGGAGGGCCGAGGCCAAGGAGGCGCGCTTGGCCCCAC
ACCAACCCCGAGGCGCTCCCGCAGTCCCTGCCTAGCCCTCTGCCCCAGCAATGCCCA
GCCCAGGCAAAATTGTATTTAAAGAAATCCTGGGGTCAATATGGCATTTTACAACTGTGA
CCGTTTCTGTGTGAAGATTTTGTAGTGTATTTGTGGTCTCTGTATTTTATATTTATGTTTA
GCACCGTCAGTGTTCCTATCCAATTTCAAAAAAGGAAAAAGAGGGGAAATTAACAAA
AGAGAGAAAAAAGTGAATGACGTTTGTAGCCAGTAGGAGAAAAATAAATAAATAAATA
AATCCCTTCGTGTACCCCTCTGTATAAATCCAACCTCTGGGTCCGTTTTCGAATATTTA
ATAAACTGATATTATTTTAAACTTTA
```

Intrón 1 733pb

```
> HOXC6-201 intrón 1: codificación proteína
GTGAGTTTACAGCTCCGAGATATAAATAAAATAAACTGAACCTGGCTTTATGACCGGCTT
CCCTAGAGAAACGGGCGGAGAGAGTTTATGAGGCTTAAATAAATAACAGCTCGTCT
CCTCACCAGACAGGGCCCCCTCTCTGCCCCCTCAGCTCGCTCTCGCTCGCTTGCAGG
GGCTTGGCCCCCTTGGCCCTGACGGATTGGAGGGCCCCAAGGCGGGCTGAGGGGCGAGG
AACAGGGCAGGGGATGAGGGGAGGAGCAGAAGATAGGGGAGCCCCCAAGTGGAGTCA
GTCTAGGAGAGAGGGAGGGGAGAGGAGAGGCAGATAAGTGGGAGAAAGTTTCAAGGAA
CTAGGATGCTTGGGGAGATGGGGGAGCCAGTCTTGAAGAAAGGGGCGCTCATCCCCAT
TATCTGGGATTTTCAAAAAATCTGCTTTTAGACTTGACAGCTTCTGTTTGGCTTTTGCC
CGCCCCCGCCCCCCCCCACCACACACCTTCTTTGGGACACACAGTCCACCTTA
TAGGAGGTCTGAAGTTGGGGTCTCGCTGTACCCCAAGTGGGGTGGGGCAAGGCAAAAG
GGAGAAGGCTAGAGCCCTAAGGAGGCTGTGAGCTGTGGCCAGGTTGGGAGGTCAGGAC
```

```

TTTGCTAGGCGAAACAGTCTGCAGAGGACGCTTTGCTGATTGCTTTTAGTGTGTTTTGTG
CCCGGATCTTTAG

secuencia ascendente 5
>HOXC6-201 utr5: proteína_codificación
ATAACCATCTAGTTCGAGTACAAACTGGAGACAGAAATAAATATTAAAGAAATCATAGA
CCGACCAGGTAAAGGCAAAGGG

secuencia descendente 3'
>HOXC6-201 utr3: proteína_codificación
CCAGGACTGTCCCTGCCACCCCTCTCTCCCTTTCTCCCTCGCTCCCCACCAACTCTCCCC
TAATCACACACTCTGTATTATCACTGGCACAAATTGATGTGTTTGTATTCCCTAAAACAA
AATTAGGGAGTCAAACGTGGACCTGAAAGTCAGCTCTGGACCCCTCCCTCACCGCACAA
CTCTCTTTACCAACGCGCCTCCTCCTCGCTCCCTTGCTAGCTCGTTCTCGGCTTGTC
TACAGGCCCTTTTCCCGTCCAGGCCTTGGGGGCTCGGACCTGAAGTCAGACTCTACAG
ATTGCCCTCCAAGTGAAGACTTGGCTCCCCACTCCTTCGACGCCCCACCCCGCCCCC
CGTGACAGAGCCGGCTCCTGGGCTGCTGGGGCTTGTCTCAGGGCTCAGGGCCCCGG
CCTGGCAGCCGGGGAGGGCCGAGGCCAAGGAGGGCGCGCTTGGCCCCACCAACCC
CCAGGGCTCCCCGCGAGTCCCTGCCTAGCCCCCTTGCCCCAGCAATGCCAGCCAGGC

```

Figura 28. Estructura nucleotídica del gen HOXC6. Registro de una secuencia de ADN almacenada en formato FASTA con sus exones, intrones, secuencia ascendente, descendente, con el número de nucleótidos correspondientes. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl).)

Se obtuvo la proteína Homeobox HOX-C6, que constituye un factor de transcripción específico de secuencia que forma parte de un sistema regulador del desarrollo que proporciona a las células identidades posicionales específicas en el eje anterior-posterior. Se muestra características para motivo, unión al ADN, región.

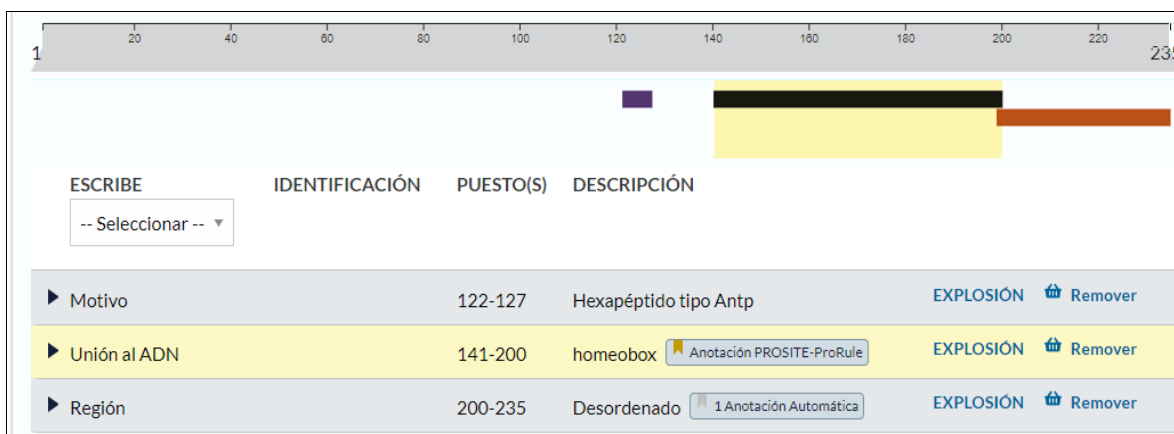


Figura 29. Características del gen HOXC6: motivo, unión al ADN, región. Se identifica el Hexapéptido tipo Antp, homeodominio y la región. (Tomado de base de conocimiento universal de proteínas en 2021. (Uni-Prot). Disponible en <https://www.uniprot.org/help/publications>)

El motivo hace referencia a una secuencia conservada corta que normalmente no contiene más de 20 aminoácidos, de gran importancia biológica. (Secuencia: IYPWMQ). El homeobox, también denominado homeodominio, consta de unos 60 aminoácidos y está implicado en la unión al ADN. La secuencia es la siguiente:

RRRGRQIYSRYQTLELEKEFHFNRYLTRRRRIEIANALCLTERQIKIWFQNRRMKWKES.

La función unión al ADN especifica la posición y el tipo de cada dominio de unión al ADN presente en la proteína, en

este caso, el dominio homeobox se une al ADN a través de una estructura de tipo hélice-giro-hélice.

A continuación, la proteína Homeobox codificada por el gen HOXC6 y la secuencia de aminoácidos correspondiente.

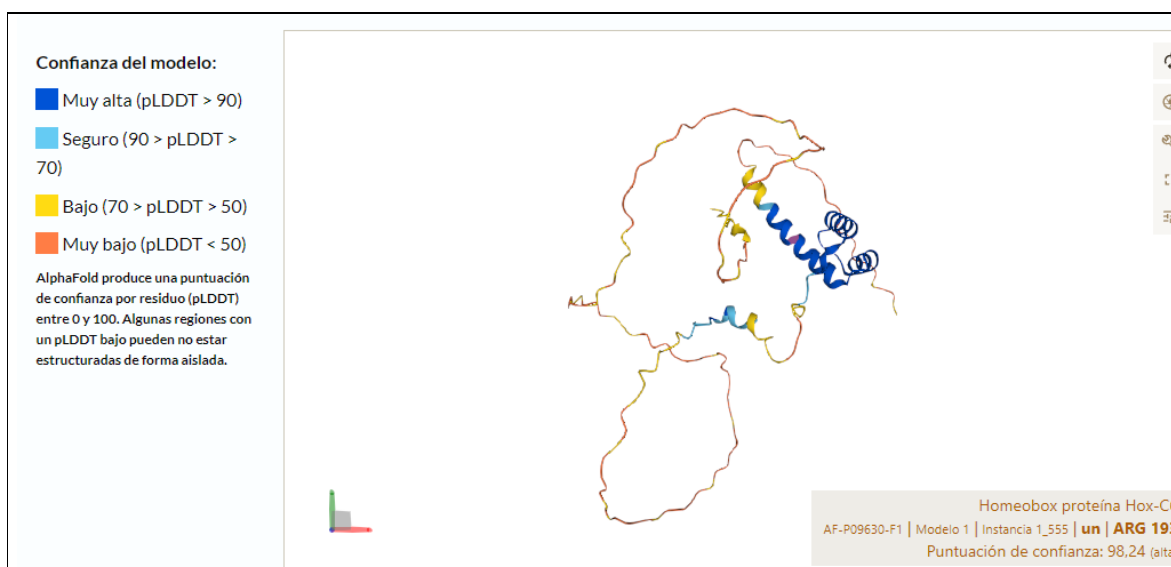


Figura 30. Estructura de la proteína Homeobox. HOXC6. (Tomado de UniProt: la base de conocimiento universal de proteínas en 2021. Disponible en <https://www.uniprot.org/help/publications>)

Instrumentos Descargar Agregar Destacar copiado

Longitud 235

Masa (Da) 26,915

Última actualización 2006-10-03 v3

Suma de verificaciónⁱ BD695C7347CB71C0

102030405060708090

MNSYFTNPSLSCHLAGGQDVLPNVALNSTAYDPVRHFSTYGAAVAQNRIVSTPFYSPQESVVFSSSRGPYDYGSNSFYQEKDMLSNCRQN

100110120130140150160170180

TLGHNTQTSIAQDFSSEQGR TAPQDQKASI QIYPWMQRMN SHSGVGYGAD RRRGRQIYSR YQTLELEKEF HFNRYLTRRR RIEIANALCL

190200210220230

TERQIKIWFQNRRMKWKKEKNLTSTLSGGGGGATADSLGGKEEKREETEE EKQKE

Figura 31. Secuencia de aminoácidos de la proteína Homeobox. Secuencia de aminoácidos codificado por la proteína Homeobox. (Tomado de UniProt: la base de conocimiento universal de proteínas en 2021. Disponible en <https://www.uniprot.org/help/publications>)

El transcripto HOXC6-203, revela 2 exones de codificación, 1 longitud de transcripción de 538pb y una longitud de traducción de 51 aminoácidos.

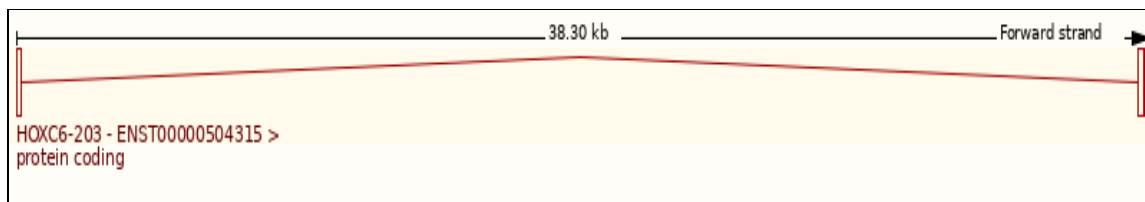


Figura 32. Estructura del Gen HOXC6/203. Se observa 2 exones de codificación. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

```
>HOXC6-203 ENSE000002085054 exón: proteína_codificación
CGGATCTGATTGCTGGGGTGCCTGGGGCTGGGCGCACCCCACTGCTGCTTGGGGGTCACCCC
AAGAGCCCCCACCATCCAAGAAAGGAATAAACTTAACCGCTGGAGGGGCATGAGGAG
CTCAGCCCCCACCCTCGACAGCAGCTAACAGAAACGCCAGGAGCAGCAGGAGGAGGAG
GCAGCGACAAG

>HOXC6-203 ENSE000002049261 exón: proteína_codificación
GACGTCCCTCCCCAACGTGCGCCCTCAATTCCACCGCCTATGATCCAGTGAGGCATTTCTCG
ACCTATGGAGCGGCCGTTGCCAGAACCGGATCTACTCGACTCCCTTTTATTCGCCACAG
GAGAATTGTCGTGTTTCAGTTCAGCCGGGGGCGGTATGACTATGGATCTAATTCCTTTTAC
CAGGAGAAAAGACATGCTCTCAAACCTGCAGACAAAACACCTTAGGACATAACACACAGACC
TCAATCGCTCAGGATTTTAGTTCTGAGCAGGGCAGGACTGCGCCCCAGGACCAGAAAGCC
AGTATCCAGATTACCCCTGGATGCAGCGAATGAATTCGCACAGTGG
```

Figura 33. Nucleótidos de los exones del gen HOXC6-203. Registro de una secuencia de ADN almacenada en formato FASTA, (Tomada de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

La proteína codificada por el gen HOXC6 transcrita HOXC6-203, es la proteína homeobox, tipo antenapedia, con su sitio conservado. El homeobox es un motivo de 60 residuos identificado por primera vez en varias proteínas homeóticas y de segmentación de *Drosophila*, pero ahora se sabe que está bien conservado en muchos otros animales, incluidos los vertebrados.

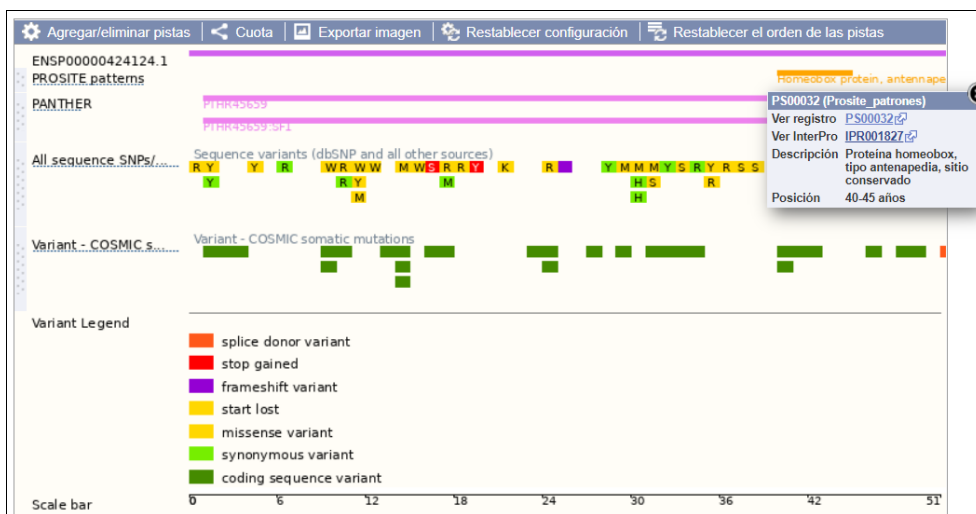


Figura 34. Resumen gráfico de la localización de la proteína homeobox. Proteína tipo antenapedia, con su sitio conservado. (Tomada de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl). Disponible en: <https://www.ensembl.org/index.html>)

A continuación, la proteína Homeobox codificada por el gen HOXC6 y su secuencia de 51 aminoácidos.

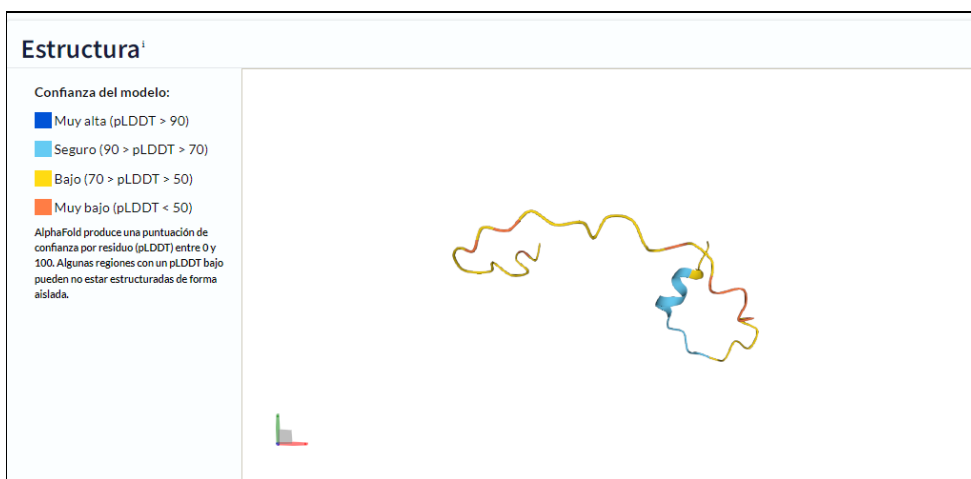


Figura 35. Estructura de la proteína homeobox HOXC6. (Tomado de UniProt: la base de conocimiento universal de proteínas en 2021. Disponible en <https://www.uniprot.org/help/publications>)

UniProt Instrumentos ▾ SPARQL UniProtKB ▾ Avanzado | Lista Búsqueda

Función

Nombres y taxonomía

Ubicación subcelular

Enfermedad y variantes

PTM/Procesamiento

Expresión

Interacción

Estructura

Secuenciaⁱ

Estado de la secuenciaⁱ Fragmento

Instrumentos ▾ Descargar Agregar Destacar ▾ Copiar secuencia

Longitud 51 Última actualización 2016-10-05 v8

Masa (Da) 5,824 Suma de verificaciónⁱ 53F790A0D8E672CF

10 20 30 40 50

MLSNCCRQNTL GHNTQTSIAQ DFSSEQGRTA PQDQKASIQI YPWMQRMNSH S

Figura 36. Secuencia de aminoácidos de la proteína Homeobox. (Tomado de UniProt: la base de conocimiento universal de proteínas en 2021. Disponible en <https://www.uniprot.org/help/publications>)

El gen abd-A, contiene 5 exones de codificación con una longitud de transcripción de 3935 pbs. separados por 4 intrones y que codifica una proteína de 590 aminoácidos. Además, se muestra la secuencia ascendente 5" y descendente 3".

```

EXON 1. 971 PB
>abd-A-RB FBtr0083388-E1 exón: codificación proteína
AATCATTCGAACCTCAGAAATCTCCGAACTAAATAGAACACAAGTGGACTAAACGAAA
TTCGGAAGATCCGATCCTCGGCCAGAGAGCAAAATATGTGCA ATGTCAAAATTTGTTTTC
GATAGTATGCTGCCAAAGTATCCACAGTTCAGCCGTTTATCAGTTCGCACCACTTAACC
ACCAACCCCGCGAATTCGTCATCTGCAGCAGTTGCCGCGCCCTAGCCGCGCCGCGCCG
TCTGCATCCGCCAGCGTTTCAGCCAGCAGCAGCAGCAACAACAACAGCAGCAACACGATC
GCTGGAAGCAACCAAGCAACCAACAATAGCAGCAGCAGTCCAAGCAGCAGCAGCAAC
AACACAGTAATCTCAATTTGAGCGGAGGATCCCTATCGCCGAGCCATTTGAGCCAACAT
CTGGGCCAATCGCCCCACTCACCGCTCTCCCTCATCGTCGCCCTTCCAGCAGCACCACCA
CAGGTGCAGCAGCAACCTGAACCACCAGCAGCAGCAACATCTGCACCACCAGCAGCAG
CAACACCACCACCAATATCTCTCGCTCTCGGCCGCTCTTCAGCTACAGCAGCAACAACAC

```

CACATTAGCAAAGTACGCCGCCGCCGAGTTGCCTCGCACGGACACGCCCCACCAGCAACTG
TTGCTAACGCCCCCGTCAGCGGGCAATTCTCAAGCGGGCGACAGCAGCTGCTCCCCATCG
CCCTCGGGATCCGGTAGCTCCTCGCTACACCGCAGCCTCAACGACAACCTCTCCCGGGTCC
GCATCCGCGTCCGCTCCGCATCCGCGGCCAGTTCCGTCGCTGCTGCTGCAGCCGCCGCC
GCCGCTGCGGCTAGTCTCTCGTTTGCCATCCCCACCAGCAAGATGTATCCGTACGTGTCT
AACCACCTAGCAGCCATGGAGGGCTCTCCGGAATGGCCGGCTTCACCGGACTGGAGGAT
AAGTCGTGCAG

EXON 2 257 PB

>abd-A-RB FBtr0306337-E3 exón: codificación_proteína

CAGGTACACGGACACTGTCATGAACAGCTACCAGTCGATGAGCGTACCTGCCTCGGCATC
TGACAGTTCGCTCAGTTCTATCAACATGCCACAGCCGCCGCATCCGCGGTATCCGCGGC
CAGTGCCGGCGCGATCGGCGTGGACTCGCTGGGAAATGCCTGCACACAGCCCGCTCCGG
CGTGATGCCAGGGGCGAGGGGAGCGGGCGGAGCCGGTATCGCCGATCTGCCAGGTATCC
CTGGATGACGCTTACAG

Exon 3 48pb

>abd-A-RB FBtr0306337-E4 exón: proteína_codificación

ACTGGATGGGAAGCCCCCTCGAGCGTGTCTTTGTGGCGATTTCACG

EXON 4. 224

>abd-A-RB FBtr0306337-E4 exón:proteína_codificación

GCCCCAAGCGGTGCCAGAGCGCGGTGCCAGACCTACACTCGTTCAGACCCCTCG
AACTGGGAAGGAGTTTCACTTCAACCACTACTTAACTCGGCGAAGCGCATCGAGATCG
CACATGCCCTCTGCCTGACCGAGCGACAGATCAAGATCTGGTTTCAGAACCCTCGCATGA
AGCTGAAGAAGGAGTTACGAGCCGTCAAGGAAATAAATGAACAG

EXON 5 2435 pb

>abd-A-RB FBtr0083388-E5 exón: codificación_proteína

GCGCGACGCGATCGAGAGGAGCAGGGAATGAAGGCCAGGAGACGATGAAATCCGCC
CAGCAGAAACAAGCAAGTGAACAGCAGCAGCAACAGCAACAGCAGCAGCAACAGCAGCAG
CAGCAACAGCACCAACAGCAGCAACAACAGCCGAGGACCACCACTCGATCATCGCACAC
AATCCAGGCCACTTGCACCACTCGGTGGTAGGTCAAAACGATCTCAAGCTCGGCCTTGGC
ATGGGCGTGGGCGTCCGAGTGGGCGGCATCGGGCCGGGCATCGGTGGCGGCTTGGGCGGC
AATCTGGGCATGATGAGCGCCCTGGACAAGAGCAATCACGACCTGCTAAAGGCGGTGAGC
AAAGTCAACTCCTAA GCTGACAACCAACCCATCGCCAGCCGAGCCAAACGAGCCAATCTT
CACCAGCCGACCACTAGCACCACCACCGCAGCCATCGCCCGGATCAGCGTCCGGCGTTA
TTTTTCGTCTCCTCGTCTGTGTAGCGCGCAGATAGCAAGCAGAAGATAAATCGATAAC
TAAATACGTAAAGTTTACATAGTAAATAGTTGCATAAGTCTACAAAAACATACATAC
ACGCAATGGCTGGGTGCGATGGCTTAGAGGCAATTAAGTGGACCGGGTTTTATTGCTGT
GGATCGACGCTTTCTGAGGTTTCCCAATTGTAGTTGTCTATTTGATAGAGCAACGCATG
CGATTCCTTCCAGATCTCCAGTGGCAATTTGTGTATTCGACATTTCTTCCATACTTTCG
CTTTTGCAACCCCTAAATCGGCAATTCAACCCCCAGTGTAGAGTCTCTGGGCCAACGGGC
CCAATTCGGTAATATTAATGGATAACTTGAACCTACTTAGAATAGCGGCAAGGCCATAAGG
AAAAGTTAGCAGGAGACGCGATCTATATAAATGAATATTGACGAAATGAGCGATTTTGT
ATAGTCAGAGAATCCGTTGCCACCACCACTCAACCAACTTCAATGGAAGTGGTTTGGGGA
AGGCAACTCGACGGATTTCGAGTGTAATTTGCGACTGCTATAGGATAACCTAGTTC
GATATTTAGAGCCAGTTTCGACTATGTTAAACGGTATATGAATACATAAATATATGGATA
TGCTTACAGATTAAAGATACTTACCTATAGAGTTATAGAGCGTACGTTTTGTATCCACCT
ATTATTGAAATTTATGTTTATGTTTTCAAAAGAAAAACCCCTCTCCAATGTTTTGTT
AAGTGATGTGTTTACTGTGTGTTTGTGTTTTCGTTGTATGCCAACGCTTGAATACAATG
ATTTCTTCAATGCCATATTAATCTTAAGTTCGAGAGCAAGTAAACAATTAAGCAGCTAAAG
TAAAGTCGCAAAAACAATTGCGCAAGTATTTCCATAAATACCAGTTGTAAGAACAAACGG
TATGCAACATACCTAAAAGCATAATCACGATTTCAAATACGTAATATTAATAATGAGATA
TGAACAGAAATGGAGTAAGACGGAGCTGAGAAATTAATGTGCTTTGCCAGTGACTAAA
ATTTTAAATTAATACCAATAAGCAATAAATAAATAGGCATCGATTGAAAGGCCACAA
AGCTACAAAAAACAACAGAAAAGAACTCTGAAATACGTATTAGCTTAAAGTTG
CAACCACAAACAACTGTCCAATGAAGGCCAAGAAAGGAGAGTTATAAAAAATGAAATAT
TGGCAGAACCATATACGTAGATTTCAAACCTCCCAAAATAAATTTACCTAGCGTAGTCT

TCTCGCGGATTGCAAGTTGAATTAAAGTCAGAAACTCTATTTATTATACAAAACCGAAGT
GGATGGCTTTCACAGCGCATATTTACATAAACAATTATTGACAAATATATATGAATATAT
ATTACAAATATACAACTATGATCTAACATGTGTATGTATAAACACCTAAATGTATGTG
TATCTAGTATGGTTAACACAAAAAGCAGTCGAGTCAAAGTAAAAGTCACCTTTTGCCCAA
AACCAGACACCACAAAAATAAGCATCCCAACCAGCGCTCATGTTTTTCTCGATGGGA
TCCCTACTATCCTACTGTCTTTTGTGTTTTCAAGTCCCCTGGTCGAGTGATATTCGCCGG
TGGCAGCAACCTTTTGCCTTGTTCACAACCGGATTGAAGGAACTGAGTGGCTTCTGTT
CCTGTTCCCGCTTCAGTTTTGTTTTGTTGCGTTTTAGTTTTAAGTTAAATGCTT
AGTCTTATTTGGACGGCTAGCAAATTAGATCGTAAATCAATATTTGTTGAACCTTAC
TGCTAAACTTCTGTAAATTTTACTTTAAGTTTAGTTGTGACGAATTGTGTATTATATAA
TTTTATGCGATTTTAAATGAATGCGAAAAAAT

Intron 1-2 827 pb

>abd-A-RB intrón 1: codificación proteína

GTAAGATAGATTGCCCTGTTTAAAGTATGGGATAATGCAGATTTTTATAAATCGGGATGGG
 TTTTATGTCCAATCGTAGTAAATGATCTTAAATGACCTTAATAGGAAGTGCAAAATATAA
 AGTCAATGACAAATGATCTATACTATCTAATATGATATCAAATATTGGAGTGCATTGA
 TAGAACTTCTCAGACAGTATCATATTTTAAATATATGCAAATTTAGTATGACCATTAGC
 TTACCTGGCTTCAAACCTTTAAGAAGATGTTTCCGGGGTTGAAAATACAAAAAAG
 AAACACTCAGCTCTTCCAAGAACTAATTTTCGCGTGCTCTCAATTTGGCGAGCGGAAATGC
 CAGTCGAGACCTTCAGTAAATCTGCAAACTGCAAAACCGCAAAATCCGTTTCGAATGC
 GCTCTAATTATCGGGTCTCGGATCCTGGGACTCTGTCCATTTTCCTCTCGGCCACCGGCC
 ATTCATATTGAATGGCCATAAAGTCGGAAATTAGGCAGCCTGTGGTGGCCGCTGGCAA
 TTTGTGCGCGGCCACAAGATTTCGCACGCACGCTGACCTTCCGCTGCGGCCATTTGCATAT
 TAGTACTACTGTATGGGCACTCCATGTGCATGTACAGTGAACCTGCTGTGCGTGGGGCGT
 GTTCGGGATTAGTGTGAATTAGAGCTAACCGCAGTTTGGTGAAAGGGGCATGCAAAATGG
 CCACCTATTAGCATTTTGACTTTAGTACGGGCGTGGAGTTGGGAACGTACTATGAGAGC
 AGGGATACCAACATAATGGCTACTTCTCTCTCCACACCCACAG

INTRON 2-3 3581

>abd-A-RB intrón 2: codificación proteína

GTGAGTCAAACTCGATACTCGATAGTGGGAGATATATTATACATAAGTAAAAAATAA
 CTGCCAGATTTCGGACTCTTTGTGAAAGACTTCGGATCTGAATGTTGAGTGCCAAATTTGAT
 AAATCTTGAAAGTTATCCGATTTTGAATTGAGAATCACACATATCCCAATACAAGTTT
 ATCAGACTGAGCCAATCAATCAGTCTTTTAAATTAAGAGTCTGTCTATCGTATCTACTGT
 GTTCTATATATCATGCCAATCATTATAAAATCAGAAATAATCCGTGCACTTCAGGGAATG
 TATATCGTCTTAAATCAAGAATAAATTCGCTTCAAGATCTACAAAGCGATCTCTTAAAT
 TTTTGTGTACATATAAATCATGGCCACTTACACTGAAATCAAAAATCAAAATAAAAAT
 TGTAAAGAACTTATATTGGATTGAATTATTTAGTCGTATGTTAGGTAAATACTATAATT
 AAAAAATTTAAACAGCTAAGTTGTTTAAACACCCGTAGACTTAAACCAGGTTTCGAACATGA
 TTTACAAATTTGTATGACCTAAACAAAAGCGTCCAGTCAATTTATCGAATTAAATTTCCCA
 AGTGTAGTATTTGGAATGGTTGCGTGCCTACCTTTTATAACAATAGATAAACGTGGCGCA
 AACGTTCCCAATAGCATTCCCATTTGCACCGCCATTTGCGCTACAAATCCAAATTCGTAA
 TGAACACTTTGGCGAGCACTTCCAAGTCGTTTATTTCGCATAAGTCTTCCAGGCGATTTCG
 AATCCGTAAACGAAATGAACGAAGAGCTAAAAAGTTGGCAGTAAATAAAAATGAAATAG
 TTTGAGAAATGCTTCTCTGTGTGCTGTTATTGGCGTCGATTTGTAATTTAAATAGC
 CTCAAGGTTGCCCAACCCGACAAACGTATAGCCCAAAATGCCGACCAAAAAACACTGCG
 ATGTGCCCCGCGCATACAAAATCCACATATCCGTGAGAGATACAGATACAGATACTTTT
 TGGCCGCGCTCTTTTATGCGGCTAACGTAGTACGGGCGAACTCGCTTCAAATTTATGT
 GGGTCAGTTTAAATTTACGCGATCCGCTCGCTTTTCCGCGCATTTTGACACAGCAGATTTA
 TTTATCATTTCTCATCAGGCCACTCGATTTCCACGCTGCACAGAAGAAATGAACGAAT
 TCAATTTTCCATTCTGCCCGCGGGTTGAATGCCTCCTGATTAAATGGCAGACACCATATG
 AATAGACTGTCAATTTGGCTGACTTATGGACTGCACATCGGTGCTGTAATAGAAAAAGA
 TCGCGCGACGGATTTACGCGGAATTTACGGCATAGGCATGTTAATTAACAGTATAAGTA
 ATTGTTAATTGACGATCTAGCAGACAGAGTTTAAATAATGGCATTAAATACGAGATAACAA
 GGTAGTGTGCTTAAGGTTCCAATGTTTATCAAGTGGTATATTTTGTCAAATCATTAGAAC
 TTAATAATCATCGAGAAAAGAGGTTCAAAAGTTTCTCGGTCTTTTCTTTTGGCCAGCATA
 ATTTACATAATTTGACAGATTTGGGTGTCAATAAATCGATAAACCAGCGCTCCTATTAACT
 TATGGCAGATTAATAATCCATTCTCTATTTTATGTTGGTTTCTGGTTCGTGATACCCCGA
 AAGATATTAATATGTGGCTAACACACCTATGTATTTGGCCGGCCAGCTAATAGATACACA
 CATCTGCGAGCAAAACCTGATGGTGAAAAGAAAACCCGAAGCATCGGCCAGTTGGGAAG
 ATAAACGGTATTTTCGCACAATGAAATATTTTGCATATCTCCCGCTCACAGCGCCAAAGCG
 ATAAATTTGTCAACGAGATCATAGAGACGCGACTGGCGGCGATGAAAGTAAATACGAGCA
 TGACGCCATTATGGCAGCCAAGTGCAGCACAGCTCGACGCGAGCCCTTTATCCACCACT
 CCTTAATGACATGCTGTACATAAAGCTGTATATACATACGTACGAAGTATATATGATG
 GTGTTTTGGCACTATATCCATGCACTTGCATGTGGATTGCACAAGAATCGCTTCGCCTCC
 TTCGGAATAAGCTTTTTCGTTTGCTTTCACTTTGCCAAGCCTGCGAACCCACCCAAAAA
 CTCACAGACGAAGACCAAGACCCAGACCCGACCCGCAACATGAACAACTGCAACA
 GCAACTGCAACAACCATTTGCTGCAACAACGGCAACGTTTACAGCTGAATGCTTGCCAAA
 TGCAATTTAACTAATGAATTTTAAATATTTTACGCCAAAGCGTCTGAATGAATTTTCATTT
 TCACATCGTTGCACCGACTCCGACAACAACATCAACAATGTTGCCAAGGGGGCAGCAGCG
 GGAAATCGCAGGAAAAGCTGGAAAATGAGTAGAGCTGCGCGCTGCACCTGATTATGCTG
 ATAAGAAGCAAGAGGTTAACTTTTCTTAATATCTGGAAGAGCGCAGCCGACAAAGCA
 TGAAAATGGCAAGTTTCCCATCGCTATCCACATTGAGAATTCGCAAACTCAGCAGCTCC
 GTGGAGAAAGGAGGAGGTGGACACAAAAGGTTAAGAAGTAGGCACACTGACAGAAAACG
 TGCAGTGGGTTCTAAAAACCCGTGGATATATTTACTGACTCAAAATAATTAAATCTTGT
 TGAAGAAACATTTTTTTCAGAAATGTAAACCTATAAAATACATTCCTAGTTATTTCAGCAC
 AATAGACTTTGGAATTTTATATTGATATTAGAAGTTTGTATGGGTCGAGTATCTATTA
 ACGTATTTTCGTCCATGCTGCAGTCAAGAACCTGACTTGGCTAGATAAGAACGCTGTCAC
 CATAAATTTGCAATCATCAGTGCACAAAGAAGTCTCGGCCTGCAAAATCTTTCACAT
 ATTAGAGATTGTCTGTGCCCCGTCCGTCCGTCTTTCCGTGTGTCAATAAAGTGGCAACA
 ACGACTGGCGAATGACCGAATTTGTGTATTTTAAATTAATAATGCACCAACGATGCCGA
 GCAGCTTCATCTGCTTATTAATTAGGAGGAGTCCGTCTGGTCTGGCTGCATCATCAT
 TCTCCCCCTCATCTTGCAGTGCCTTGACTAGTATTTCTCTATATCTCCGTGATATCGC
 CGACATCTTCTCTCTCCAGGCCGACCATCGTGGTCTACTTAAAGTGGAGCCGTCCTCAA
 TGGCCGAGGTTTCACTCGGCCGGTGGTCCGCGACTGCCGACTGCATTGGGGCCGCTAA
 TGCTGCTTTTTGTGGCTGGCGTGCACAACTAAATGAAAACGGCCCGATTGAGTGGTTTA
 TGAAAGCTGCCATGTGGCACAATGTCCGAGGGGCGAGTGGACAGATATGATTTGAGGGA

TCGACGCTCCATTACTCAGCGGGATGCTCTCAAGGCTGTTTGGCGCGGAATTTTCGCATTA
GCAGCACATATGAGAAGCACAACTCCTTAGTTCTTAGAAAAAACTCAAAGAAAAACCG
AAAAATTAATTGCTTTTCTCTTCTCTTTTAAAAATTTTTCAG

INTRON 3-4 12086

>abd-A-RB intrón 3: codificación proteína

GTAGTACCGCAATTAAACCGGAAACCCCTTGAAGTTAATATCCCGACTGACATTGAAGCAA
TGGGTTATTTTCTACTAAATCGGCCGTGGAATATGCCGGAATAATGCTGTTTTTGAAGCGCTGC
CTAGATGGCCTTGAATAATATGATTTACCTGCTCAACCCAGCGACTTTCACCCGCTGAAG
GAAACCCCTTCTTTGTAAATCGAGCCTGCAATCTGCCGTGCACATGACTCCTGCTGCTCCTG
TCATAAATCCCTCAACCTTTGTTCTCACAATCCTTGTCTTATTACATTTGATATAATTTTCATTG
TTTTTCGGAGTCACTCTTCTGGCTAATATCCAAGCAACCAATGGTTTTGAAGCCAAAGGT
TTGGACGCTTGGCAAGTGTTCGGCTTCTATGTTGATTTAGCAAACCTCGTGAAAGTTGCA
CCGATAACTCAAAATGCCGGTCAAGTACCCGACCAAGATCGTAAATCACAATTTCCAATG
CTGGGGCCCCCAAACTGCAAGCTGCTGTTGTGCATACAAATTTGGCGAATACCTAATGA
GTGCATCTCGATGTGCAGAAACTACTCAAAAGCGGCTCGATGATGTCGATTTGTGAT
GATCGCCGATCTGCGATGGGTCCGTTGTTCTGGGCTCGTTAAACGTCCTTTTTATGGGCG
TTTTATGATCTATTACGATTTTACGCAATGGTCTTGTGCCGTGCATACAGCATCGTAATGT
TGCGGCCACTGTGCAACGATCCCCGTTTTATGGTCCCGAATATGGCGTTGATTTGTTGGC
GAATCACCATTGCGTTTTAATAGTTCGACAGTCGACAGGGCAGCCGCAACACGAGTAG
CCACTGTCAATTTGCTGGTTACAAAAGGCGGTCAATTAATCGGTATTCACATTCACATTTCCC
AATCCCAATCCCATCTATGTATAGCCCAATTTCACTTGGATTAAATGGCGCTACCCATTAG
TCCGCGGCTCGAATTAGGGCGCATGTATGCAATGAATATGAACCAAAGTGTTTATT
ATTTCACTGCTTGGCTTATTTCCGTTGATTAAGTGGGCCAGAAATTTCTAATAGTAATTA
ACAATATAGTTTGACCGTAATGACTAATTAACCCAGCTGGATGCTTGCATATGCAATAAATA
TTGTTATGTACTTTGTAATAATTTGTTCTTCCCTTATTTATTTATTTCTATTATTATTATA
TGTTTCGTATTTTCAACAATAATCCAAGCCTGGCTCACAACCTAAAAAATCGTAATGATGTT
TTATCGATTGGAATAATGAATCCAACCATCAACTCCCTACATTGCGCATACGCGTAGTGG
TCCCCTAGTAGACATTAACAGCCTACCCCTGTTGTTGGTGTGGCATGTAATTTGCGAACC
GTTTAATTTGGGCTCGACAGCTGGGCAAAACATTCGGGCACATGGCCCAACATTCGCTTGT
CTGGTACTATGGTGTGGAGGCCCTCTGGTCTCTGTGCTTTGGTATTTCTTATGTGGTCTC
TGGACCTGGAATGAAATCGGAGCGCTTCTGCCAGAGTTTGGCCACCGCAAGTACGCTGCC
ACGGCTCAGCTTTTGGCGTCAGAGATCGTTTGAATTTGACGGGTTTTCAAGTCCCTGAA
CTTCCCCTTTGGCACTGGCGATCCTCTGGTGTGTTGTTTTGTAATAAATCACTGGCATTG
GCGTTGAGCGGAGTCAATTCGGGCAATTTACAGCATAACAGCCGCTGTTGCAAGTAAT
TAACTCACCAGGTTCCAGTTGCCATTTTCGATAGATTTGCTGTGGTCGAGTTTGGTGGG
TATAAGATTACAATCAGAAGATCGGACGACAAAAGCGTTTTCTTTAAGCTGAATTTTACAA
GTTTTTCATCAGCTTAGTTTTGTTTCACTTTTCCCTATGAAATTGAAACAGCAACCCTAG
TGTGAGTGTAAATAATGTAATTTCCCAAGGTGCGCTCCCAAGACAATTAATAATTTGTAA
AGCAATAACATAAAAAATCTGTAAGACCGCAATATGTGTAATTTTAATATATTATAGT
CATCATATTTTCATTACAGAGCTCAAGAATTGCTAATTCATTTATTTCAAACCTACATAC
AAGTCTCTCAATCAAAACGCAACATGTGTAAGTTTGCATACGATAAAAAACATTTCTTT
TCGGCAGTTTTTAATTAACATTTACAGTTTTGAAATTTGTTGTTACTGCAATTTGTAAGCTG
CTCTACAATTTATGGTCAGTTTTTAAAGCTCTTGATTTTTTTTTCTGAAACAAATCACTCA
TTTTCTTTTCAAAGTCCACAAAGCTATTATCTATCTTCATGGCAGATTTATTTGCATTT
TTGCACAGCGAAATCGGCGGCAATGGAATAATAAATAATATTTTATTAATTTTACA
AGCCATAATTTGGCGTCGCGATAATCTCGTCAGTCGGGGAATGAAATTTATGCCACAAAG
ACCTCTCTATCTCTGGGCCAGTCTCTTTTCTGCGCCCGCCGCGCAAGAGTCTCTTTCTGAA
TTGAATGCTTTGTCTTGGGATTTGAACTGGTCTTGTGCTAGAAATCAATGCGAATATCAATGT
TTAAATCTCAATTTGCTGGCGCAATCAATGTTCATTGGCAGGGATCTGCGGGCGCGGAAGA
TCCGCTCTGCGCAGATAATCGTTGATCCCGCACTGGGGTTAAGATCCCTCAGCACCCCCC
ACCCACTAGCATCTCTCTGTCTGCATACGTGGTGTCTATCTCGCTCGGAATCGGATGGAG
ACGGCTGGCAAGGTTAATTTAGCGCGGATAACGCAATTCATTTCTTGTTCGGTTTTT
GTTTAATTTCCGATTCTCTGTGTTTTCTGTTTTCTCATTTTTTATTTGGTCTTGTGTTAT
TTACATTTTGGTACTCTTTCTCTTTTCAACCAAGCTATCTCTCTCCGTAGGGGA
GTGTGCGCAGCCGTGTTATCGGCACGTTTTTCGGTCTGTGATATTTTTTTGTTTTTTTT
TTTTTATTTATTTCTGCTCGCGTGGCGGAAATTCGAATGACAGTTTCTCCGAGTTTTT
GTGTTGTTTTCTCGCGTTTTTAACTACCCCGTAGCGTTTAATCTCGCGTTTTAATGCC
CAAGTGGCCGCAATTTGGCGAAAGGGTTTTGGATTCAAGGATCCGGTGGCGGAGCACCCAC
ACAGATGACAGAACACAGACAGACGAAATGCGACAGTACGCGGCTGATTTTTCGTT
CAGTGGTAGTGGGCGGCACACATGTAAATTTGGCCAGACAAAATCAAATTAAGGCAC
ACACGTTCCACAGACAATTCGGCATCGATCCCAAGTCATCGCTGGCGCTTAATCAAAGT
TATTTCCCTTCATGCATGACACAGACAGACGAAAGTTGATAAAGCTTGCTATTATAGTTT
TCAAAAAAGTAATTAATAAGTTTTATGAAATTTTCAATTTGTTAAAGATACATCTTC
TATCTTCTGTTCTAAATTAGCGTTATTTATTTTGGTAGCACTCTTTTTCACAGTGCAC
ATTTGTGAGTCCAAAACAGCACATATTAGGCAAACTTACCATTATCCCATTAAGTTAAA
TATTTAATTTACAAAAGTTTTATTTAAACCTTGACTATGGCGTGGCATTATTTTCGAG
TTTTCCATTTTCTGTATCGCTGCAATGCGTGACTGTGGCATTTTCCGACCATCCTGTT
GCGATAAATTTGAAGACAGAGTAGTAAATTAATTAATTCGCGGAATTAACAAAGTT
GTGCGCGTGTGCGTTAATGACCCCTGTTGTTCTATGCGCGCTAAATTTATTTCTT
CTGATTAATTTGCGAGATTAGAGTGTTTATGATGCACATCAGTACTCTTTGATGTGCCATAA
TCGTTGGGTAATGTAAGTGGCAACGAGGTACCCGATTTCTTGAGGATGAGAATGTATT
AGGATAGGAAAGGAATTTTATATAGTAGAATTTTGTATGTTCTAAGTAGATCTTCTTA
TAATTTGAAATTTAATTTTGTGAGCTCTATTTTAAATATTTGTCATATGACTATATA
AGTATATCTGTGACAAATTTTAAACGTTACAAATCTGAAGAACCTTCCCTCTCAATTTCACTT
AGGGAATGAAGGAAACGATGTGCCATTTGTTGTTTTTGGTATCTTCCCTCGAATAACTGT
TTTTTCCAGTGTGTTGCTGATGTGAGAATCTTTCTTACCCTCAGTCACTCACTCAATG

AAGCAATTCAAGTGTATTGGAGTTCCTCCGCCATAGCTTCGTTGTAATAGTTGTGAAAAGA
 ACAAACTTGAAGGGTCGAACACTGGGGGAAATAAAAACCTGTAGGTGACAAAAACCTT
 TCATCTGGGATCAATGCAACAGCAACAGCCACCATTTCTCAAATTACTCTCTGGCAGCC
 ATTAACAAAAGGCAAACTGTAGCTGTGTGTGTGAGCCATTAATCAGCTTTGTGAGCCAAA
 GAGAATTTTCGAGGGAAGTTCAAGATTGGAACACTAATAGGTGAGGTTTCGTTTCTGCGG
 TCTGCGGTGCATCGACATTTTCTCAGGAGCCACAGGACCACACCTTGGCATTTGCGGGG
 GGGAACTAAACAGGGGTTTTCTGTGGGCATTATAATGAAATTTCCAGACGAGGGGGATA
 AGCATGAGTTATGTGCACGCCTAGGTAATATAGAACCAGCATCGATCGATCGATGCTTT
 TTATACGAATTTAAGTGTACGTTCTCGTTTGTGGAAGTTGTGCGGGGATTAGAGAGG
 GGACACCGATATGGTGGGGGTGCTTCACTGAATGGCGTTTTAATACGTCATCGTTTCAA
 GTCGGCCACGCAATCTAAGTTATGACCGACATCTGCCCCCAAGTTCTGCATTCGAATGG
 GAATAGGAATCCGGAGTTCGTGTTGTTTTTACCAATAGGAATATGTATATATAGGTAT
 ATACACGAGTGCCTTGGATTCGTGCACTGGGCCATGAACAGAGTTGGGAAAAATGAGACG
 AAAAAGGCAAGGAAAACCAACCAAGCAAAACACCCGAAGAAACAATACCAAAATG
 CGTTCTCAGTCGAAGCAGTTGGCTCAATTTATATGATTTTCAATAAAAGTCGATATAA
 TCGTTGCCAAGGTTCAATAAGACGACCAATGAACCTTGGAAACGGCCACAGAAAACAAAAA
 GGAGAAAAAGAAAAACCGCGGTGGCAGCCAATGAAGCAATCAATTTCAATTCAGTTT
 CCTCGCCTTTCCAAAACGCCAAACCAATGAGCCGTCCAAAGGGCTAGACGAGGCCAAT
 TGCAGATGGGTTACGGCTTCAGAACCTTTTTTGGTACCCTTATCTGTTTAAATGGCTGT
 ATTGAATTTCTAGACAGGCTTACAGGTATTGGCCAAGCAAGGTTGTGAAAAAGATAAAA
 ATTTAATGCTCAACTGAATATTTAATCATATCATTTACAATATATATATTCATAATCGAA
 ATTGCATTACAAATAAGCTTTATATAAAGTAAAACTCTTTAAAAATAAATTTTGTAGTAATG
 GTTAAATGATTGCAACTATTTTCAAGAGTACCCAGTATGGTACGTCCCAACGAAACTT
 TTGCTATACAGTACCCTTTGTCTTCATTTATCGTCCGAAATGAACCCCTAACAAAGGATT
 CAAAAAGTGTGATGAACCCGACACCGACCGAATGTGAGCTATGTGAGATACAGATACT
 TTTTCAGATACAGATACGGACAACCTAATTCACCTACATACAGATGGAGTCTCTTTGGCG
 CCACCTTCATTTGTGATCAATTTGAATGGCGGTTGATTGTTGTGCGATCGATTGCGCGGATTT
 AATCGCAGGGCGGCGACATTAAACGCCTTTAATTTGCAATTAATTTGCGGCGAACAGACA
 GAGATAGAAAAGCGGCCAGAAACAGATAGGACAAGGGTGTCTGAGTTGCGGTGAGTTGGC
 GACACGTGTTCCCTTGGAAACAGCGCTTGAACAGACGCGGTTGCAAAAAATCTTACGGATT
 ACCAATTTCTAGACGCTTAATGCCATTTATAGGCCATCGCTACAAATGCTCCACAGCTGTC
 CGTCCTTCCAAAAATAGTTGGGCTCCAGCTCGCCACCGGAACTCCTCTCTCTCTCTCTCC
 TTTCTCTCCGGGTGGGCCAACAAATTTATTGGAACCACTTCGCGCGGATGGTTTAAAGAA
 GGGGGCCAGAAAAAGTCTGATTTTCCGGTTTCTTTTGGCACGGCTTTCTGCCAACCCC
 ACACATACATCTCTGCGGTGGTTTCTTCCGTTTTCCCTTTTCACTTTTGGCCATATATTT
 TTCACATTTATGTCGGGTATTTTGTCTTATCACACCCCTCGTGTGTCAGTGCGTGCAAAA
 AATGCGAGGGACGGTGGGGAGTTGTGCGATGGGCGGTATGGTTGTGCGCTCCACACCC
 ACAAGATTAAAAATGCCATAAAGACAGTAAGGGTGAATGGCTCTAATGGCTGCCGTTGCC
 GTTATTTGCCATCCCACTCGCTTGCCTCTCTCTACCTCTCTCACTCCCTCTTTTTCCGT
 CTCGTTTTCTACTCTGTGCAAAAGAAAACCTCGGGCTGTTACCATGTTGGCTTGAGTTACC
 ATGGCTTCTGGTTTGTAGTCAGAGCTGAGGGCTGTCTGCACTTTTTTGGGATACAGGGCGCA
 TTTCGAATATTTGGAACCTGGCAATGATTATCTGGCGAGATAATGGGAAAAACATATGAG
 AACATTCCATACATTCGATAATGCATCGGTAAGGTATTTTTCGAAGGAGATTGGCTAGCA
 TTAGACATATGCTGCAAAAAATTTGATAAGACTCAAAAAACAAACGAAATTCCTAAGATT
 TCCCAATGAAATCATTTGTTTTTGTAGTTGATAAAGCTTTAATACGTATATGGTTTTTAA
 ACTTATCGTTAATATATAAAATCCAAATTTGGAGGAACACTTATCTTTTCCATTTTAAAT
 TCGAATTTGAATACACATTTCTTTTCAACCCCAATACTTCTCTCTTGGCTTAAACAGAA
 AACGAGGAGTAATTTCAATAATGGCGTGAAATATTTGATGAGAATATATTAATCCCCCAG
 GAGCCTGTGTATTTGCAACATTTGCCAGGAATTTTCTCCACTTTTGGCCGCGGAACAGAA
 CGAAACACAACCTGACAGCGCAGCTTGCCATATATTTTCCAAGATTAAGCACACAATTAG
 CAGGGGAAATGGGACGCTCGGAGTTAGGGGCGTGGCAGATCTGATAGCACAAACAGCAACA
 TTGGCAACGCGCTGCACATAAATGCGAGAGAAGGAAGACTCCCCGTAAGCAGGCTTTC
 CGCTGCCAGGTTTTCCGCAAAACAGACACCAACCCACTTTTCTCTCTGCGGAACGAAATG
 GCTGTGAGACTGATTAAATTTAAATAAATTTAATAAACATAAATTTATTTGCAATGGCAAC
 CAGGAGCTGCTGTTTTAGCCACCCCAAAAAGTCGAGAATGAGCGCAATTTAATGGCGGCG
 CTTACCTGCCACTCCCCAAGTATCCTGTGTACCTGTGTCTCGTGAACATCCTGCCCAT
 CCTCCCCCTTTTCGTAAAGGCCACAAAAAGCGCCAGATGTTGCCCGTGAGTCTTGTGCG
 GAATAGCGTTTAAAGCGGCTTTACTTTTCAAAGTTTAAATAGTCTTCTGTCGCTCAAAGA
 ATTCCACGCATTTGCATATCCTTGAAGTTGGGGCCGGGAATGGGCGATAGAAACAGCTC
 CCTGGGGATCTCATTTATCCCGAGGTACATTGCAATTTGCGCTAAATGCCGTAAGATCG
 TCTCCGTAACAGCATGTACTTCTTAACAGCGGAGAGAGATTTGTTGAGCCATCACT
 AAATTTCTCGCAGCCCCAAAAATATCCTCATTTTCAAGCATGTTGCTATATCTGCGTCTG
 CCTTTGTTTATATATCTCTCAACATGCCATCCGGGAAGTTAGTTTCGAAAAATTCATACC
 AAATATAAACAAATCAGATAAAACATATGCCAGCCAGTGAATTTCTTCTCTCTGAAAT
 AAAAAACGAGATAAATATACGACGGCATAAATATGCGTACACATTGGTGGAGCCTTAT
 TCCTTATATGAATGTATTTTCTTCAAAAAATTTGAACATTTTAAACCAAAAAATGCGTGAT
 TTCGATAAAGATGTCTACAAAAGACATCAATATCTTCACTTAAGCAGTACAATTTTGTG
 CTTACGTACAAAAATCTGTAAAAATGTCTACAAAATGTGCTTTTACTGACAAAAATTCGTT
 AAAATATATTTGTACAAATAAATGCCTTACATGAGTATAAATTCGCTATATAGTAGAGTA
 CGAAATTATACTTAAATTTTGTGCTAAATCAGTAATATATTAATTTTACCCACCCAGTC
 CCATAAAGTTTACGAAAACTCAGCAATTCGCTCAGAGCAAAATACACAAACGGAATTTG
 GCGCCAAAAGTCAACAAACAAGTCAAAGCGGGGAAAAGTTCTCCAATTCACAAAACTTG
 AATGGTCAACAGCAGCAAAATCATGAATTTATTTGCCGGCAACAAATTTGAACAGATA
 GGAATAGGAATTCAAATTCAACTAATTTGCCGCACTGAACAAACAACAAATGCGCTGAA
 GCAAAAGCAGCAACAAATGTTTCAAAGTTCTTTGTCAAGCGATATGTAATTTGTTTAT
 ATTACCGAAACGAAAGCAAGACGCTAGAAAGGATGGGTGCATAGAAAAAATAGGAG
 CCAACACTAAACACGAAATACTTGTGCGGCAAAATGTTTGTATTGTACATTTTTTTG

TCGCTTGATAAAACGAAGACTCGAAGACTTGTGACGCTCCCCACGGAACCACTTTTTCGAATGC
 ACACATAGATACTCACACCACACACACAGTGCCAGATACTTGGGCAAACATATGGCGCT
 CGCCCCCTTTGAGTGTGTGTGAGTATCTTTATGTACGCTTGAATTTGTTTTATGAT
 TTTTGTAGTCATCATAAGTTTCAGTCTGCGCTGTTTTTTTTTTCAGCTTCTCGTTTTTGT
 TGTGTTTTGTGCTTATTTATACGCGCGCTGTTTTATTGAACAATTACGTTTATGATTATTA
 ATATTTTTAATTGTTAATGTCAATTGTTATAGACGCTGTCCCGGGTCTGTGCCATAACCAT
 TCCCATTCTCATTCCCATAAACATTCCCAGGCCATTTCGACCGTTGACAGCGGCGCTT
 GGGACGGCTTTCCTGGCCTTTTGTACCTCTCGGATTTCCTCAGCCAATCCGTAGAGTA
 TACAGTGGAAATGCCCCATTGTGGGCAGCCCCATTTGGACGCCAACGCCACCGCTCTGG
 AATTTCTTCGATTCCGAAACAGATTCTGCCCTCAGCTCGATTTTCCGCTGGCTGTTT
 TGTTTACCGCCCGCTCGCTTCACTCAATTGCAGCCACGATTATGGCGCGCTAATTGATT
 TTTAATTGAAAAATGGTTTTCTTAGTAATTGTGCAACGGACCGTTTCGACCGGCCACCT
 CAACTCGAAAATATAAACTTAAGATAGCCGCTAAGCGCTTGTCAAATCTACCGAAGCTCT
 CAAGTTAATTTAACGATGAGTGTGTAATTGAAATTCACCAATAACCCACTTCTTATCCA
 TGAATTTAAGTCAATTGCTTACATCAGAAAACCTCAAAGGTATTTTTGCTTTATAAAATAC
 TCTGTGAGATTAATCTAATCTAATGACTCACCAGGGGGCAAATTGACGCATGTTATTAAT
 GAATGCAATTAATAATATATTTATTTTCATTGTTTTTACTAATTAACGCTTCAATTAAT
 ATTTTGTGTTTAAAGCAGTATATAAACATAGAAAACCTTACCCCTAATCAACACTTAAC
 GCTTTTTGAAATACCTTTTCTGTAATATTGTACTACCTTACCTATCTTCTTCTG
 CTCCTTTGGCTCTCACTGAGCTCCTCCATACTCTGCTTATAATAAAGATTTCATAAATTT
 TAGAAAGATCAATGAACCTCGAAAGTCCCCACAACGCGCTCGCTTAAACACTCAAT
 CCGAAACCCGAATTGAAACCGCTTAGACTCTAGACTTTTCCCAAGCGGATTTAAGGCTT
 TCCGAATAGTTTTTTCATGATGAGTGAAGTGCAGTTGGCATTGGAATAATAACAAAGT
 TGGGCGGTGACTCAAAGTGAAGCTCTCCTTTTCGCTCAGATCCACATTACATCCACGT
 GTGAGTGCTCTCTTCGCAATTCGGTTTGGCTGCGCGGACTCACCCTCACACTAAGAAA
 AAACCTCAGGCGTTGCCATGCCGACGTTTTGCTCTTCCCTCTCTCTATGACAATGTGTGT
 GTGTGTGTCTACCTGTGCGAAGGCAGACACTGCGGCAGCCATGTTGGATTCCATCACTCC
 GTCTGCGAGCCAGCAGCATCTCGCTCGCACCTGCGCTTAGCCATTGGCGGCCATTGCG
 GGGGTCGTAATCTTGTTTAGGGGCGGTGTGTCTGGGAGTCGTGATTTATGGGACGAA
 TTGTAGCTGCTCATTTGTAGCGGTAAATAAATTTAGATTAGTTTGGTATTCCTGGCCACAG
 GATAGCAGGGGGAACCGGAAAGGGGGTTGATTAGAATTGAAATTGCTATTGCCCCCT
 TCCCTTTACACTGTTTCTGTTCACTCCAGTAACAATGTTTTTCGCAATTTATTGCTGCT
 CTCTTTAAATTAAGCAAAGTATAGTGTGTTTAAACAAGAAATTAATTTAAAGATTA
 ATTGAATTTAATTATCGGATAATGAAATTAATAAAACGATTCTTTTCCATTTTAAGAC
 ATTTGGGACCATTCGGCTCCATACATATAAATTTTTTCTTAACATCACTGTTCCGTTT
 ATCCGAAAAACAGGAGCTGGGGTTTTTGTATTTTCCACCAGTGAATCACCAATCTTCT
 TATCGGTTGAGGTAAATTCAAATGTACGGCCTTATTCGCCGTAATATTTTGCCAACT
 GTTCAGTCTCATAAAAAATGCAGCGGAAACTGAAAAAATCACAATGATTACAACCTGCA
 GCGGAGCAACATATACAAATTCGACCTTTCATAATATTACGTAAAAATGTCCCTTTTT
 TATTTGTCTGCTTGAAACAGTTACCGTTGGAAAGTCGCAACTTGCAACCATAAAAAAT
 GCCGTTGACATGAAACCTTTTGCCAGAAACTCGGCACAAAGGCAGCAGTGCAAAAAAG
 AAGAGACATGAAAGCCATAAAACGGCAAGCGGGTATAAGTGAGAGGCTGTAGGCTGGGA
 CTGGGACAGCAATTTGTAGATAGCGAATTTTGTATTTCCGTAACATAAACACGGAGCTG
 TGCCATTCCGAGCGCATCTCCGCACACCGTCGCGCAAAAAATTCGCACAGGTAGCATGGA
 AAATTGCTAAAAATAGAAACAAAAAAGTAGACAAAGAAAGTAGAGAAAGGAATGCGGGGA
 AAATCGGGGAAAAGCGCTGGCTGCTGGCCACCAGGGCGAGACCTGCCGCTGGGCGAGTCG
 GTGTGCGTATTGGTAGCTTTACCCCTTTCACGCACTGCGTGCGATGCCAATCCGCAATTGA
 ATCAGGTCCGCAATGACCCCGGGGTGTGACGTCCTGGGGCGCCGCTGCTGACGTCAG
 CTCTAGCATCAGATCACTCCCTTCTCTCTCTCTGTTAAAAACAATCTGTTGTTTTT
 CCCGTAATTTGCAGTCAGACATTTCGAGTGCAGCTCAGAAATGTACGCCCCAACCTCCA
 AACCCCCATACCCCGCAAAATCATGTGTCGACGCTGACATGTGTGACAGTGGCCAA
 AGTACCATTTTCGACCGTCTGACACCGTCTTAGCAGCTCATTAGGCGTTGCCGAAAAGG
 CAAAAACATCTTCTTAAACATCGTTGCAACATCTGACGCAACTTTCTCCGTTTTCTCCT
 CCGCAG

INTRON 4-5 70

>abd-A-RB intrón 4: codificación proteína

GTGAGCTGGTCTTCCGAGCACCTTAAACTCTTAACTAAGTGTCTTCCGTTTTTAATT
 TACCTTTAAG

SECUENCIA DESCENDENTE 3

>abd-A-RB utr3:proteína_codificación

GCTGACAACCAACCATCGCCAGCCGAGCCAACGAGCCAATCTTCCACGAGCGCACCAC
 TAGCACCACCAACCGCAGCCATCGCCGGATCAGCGTCCCGGGTTATTTTTCTGCTCTCTCG
 TCTGTGTAGCGCGGAGATAGCAAGCAGAAGATAAACTCGATAACTAAAAACGTAAGT
 TTTACATAGTAAATAGTTGCATAAGTCTACAAAAACATACATACACGCAATAGGCTGGG
 TCGATGGCTTAGAGGCAATTAACCTGGACCGGGTTTTATTGCTGTGGATCGACGCTTTCT
 GAGGTTTCCCCAATTGTAGTTGTCTATTTGATAGAGCAACGATGCGATTCCCTTCCAGA
 TCTCCAGTGGCAATTGTGTATTCGACATTCTTCCATACTTTCGCTTTTGCAACCCCTA
 AATCGGCAATTCAACCCCAAGTGTAGAGTCTTGGGCCAACGGGCCAATTCGGTAATAT
 TAATGGATAACTTGAACCTACTTAGAATAGCGGCAAGGCCAAGGAAGAAGTTAGCAGGA
 GACGCGATCTATATAAATGAATATTGACGAAATGAGCGATTTTGTAGTCAGAGAACTC
 CGTTGCCACCATCAACCAACTTTCAATGGAAGTGGTTTTGGGGAAGGCACTCGACGGA
 TTCGAGTGTAATAATTGCGACTGCTATAGGATAACCTAGTTCGATATTTAGAGCCAC
 GTTCGACTATGTTAAACGGTATATGAATACATAAATATATGGATATGCTTACAGATTAAA
 GATACTTACCTATAGAGTTATAGAGCGTACGTTTTGTATCCACCTATTTATGGAAATTT
 ATGTTTATGTTTTCAAAGAAAAACCCCTCTCCAATGTTTTGTTAAGTGATGTGTTTAC

```

TGTGTGTTTGCTGTTTTCGTTGTATGCCAACGCTTGAATACAATGATTTCTTCAATGCCA
TATTAATCTTAACTCGAGAGCAAGTAACAATTAAGCAGCTAAAGTAAAGTACGCAAAAC
AATTGCGCAAGTATTTCCATAAATACCAGTTGTAAGAACAAACGGTATGCAACATACCTA
AAAGCATAATCAGGATTTCAAATACGTAATTAATAATGAGATATGAAACAGAAATGGA
GTAAGACGGAGCTGAGAAATTAATGTCGTTTGCCAGTGACTAAAATTTTAAATTAACCTA
CCAATAAGCAAATAACTAAATAGGCATCGATTGAAAGGCTACAAAGCTACAAAAA
AACAAACAGAAAAGAACTCTGAAATACGTATTAGCTTAAGGTGCAACCACAAACAAAC
TGTCGAATGAAGGCCAAGAGGAGAGTTATAAAATGGAAATATTGGCAGAACCCATAT
ACGTAGATTTCAAACCTCCCAAATAAAATTACCTAGCGTAGTCTCTCGCGGATTGCAA
GTTGAATTAAGTCAGAACTCTATTTATTATACAAAACCGAAGTGGATGGCTTTCACAG
CGCATATTTACATAAACAATTATTGACAAATATATATGAATATATATTACAAATATATAC
AACTATGATCTAACATGTGTATGTATAAACACCTAAATGTATGTGTATCTAGTATGGTTA
ACACAAAAGCAGTCGAGTCAAAGTAAAGTCACCTTTTGCCCAAAACAGACACCA
AAAATAAGCATCCCCAACAGCGCTCATGTTTTTCTCGATGGGATCCCTACTATCCTAC
TGTCCTTTGTTTTCAAGTCCCAGTGGTCGAGTGATATTTCCCGGGTGGCGACAACTTTT
GCCTTGTTTTCACAACCGGATTGAAGGAACTGAGTGGCTTCTGTTCTCTGTTCCCGCTTCA
GTTTTGGTTTTGTTGCGTTTCAGTTTGATTAAAGTAAATGCTTAGTCTTATTGGACG
GCCTAGCAAATTAGATCGTAAATCAATATTTGTTTGAACCTACTGCTAAAACCTTCTGT
TAATTTTACTTTAAGTTTGTGTGACGAATTGTGTATTATATAATTTTATGCGATTTTA
AAATGAATGCCAAAAAATT

```

SECUENCIA ASCENDENTE 5

>abd-A-RB utr5: proteína_codificación

```

AATCATTCGAACTCAGAAATCTCCGAACTAAAATAGAACACAAGTGGACTAAACGAAA
TTCCGAAGATCCGATCCTCGGCCAGAGAGCAAAATATGTGCA

```

Figura 37. Secuencias de nucleótidos de exones e intrones del gen *abd-A*. Registro de una secuencia de ADN almacenada en formato FASTA. La secuencia en azul representa la región traducida y en rojo corresponden a la zona UTR. (Tomado de Instituto Europeo de Bioinformática, una división del Laboratorio Europeo de Biología Molecular. (Ensembl).)

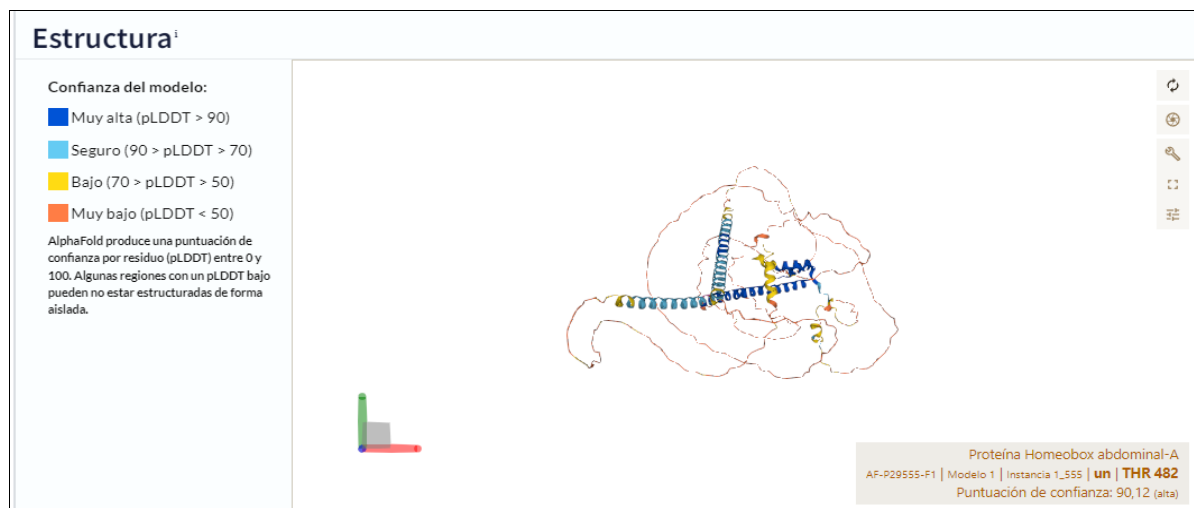


Figura 38. Estructura de la proteína homeobox abdominal A. (Tomado de UniProt: la base de conocimiento universal de proteínas en 2021. Disponible en <https://www.uniprot.org/help/publications>)

A continuación, resumen del gen *abd-A* y sus ortólogos, se observa que el gen *HOXC6* es ortólogo del gen *abd-A*

Ortólogos humanos (a través de DIOPT v8.0)									
Homo sapiens (humano) (31)									
Especie/Símbolo genético	Puntaje	Mejor puntuación	Mejor puntaje inverso	Fuente	Almuerzo				
				comparar punto de hueso intermedio homólogo intraspecie OXA OrthoDB Orthomancer Pantene RefSeq Transcript	¿Complementación? ¿Transgén?				
HsapiHOXA6	3 de 15	Si	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ● ×	=				
HsapiHOXB6	3 de 15	Si	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ● ×	=				
HsapiHOXC6	3 de 15	Si	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ● ×	=				
HsapiHOXA5	2 de 15	No	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ○ ×	= 1				
HsapiHOXA7	2 de 15	No	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ○ ×	=				
HsapiHOXB5	2 de 15	No	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ○ ×	= 1				
HsapiHOXB7	2 de 15	No	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ● ○ ○ ○ ×	=				

Figura 40. Ortólogos humanos del gen *abd-A*. Entre los ortólogos humanos del gen *abd-A*, se encuentra el gen *HOXC6*. (Tomado de Consorcio FlyBase. (2022). Uso de FlyBase: una base de datos de genes y genética de *Drosophila*.)

Según Jeff Chang, las secuencias biológicas son el objeto central de la bioinformática, Biopython presenta un mecanismo para tratar con secuencias “el objeto Seq”; estas secuencias son esencialmente cadenas de letras como AGTACACTGGT, propias de formatos de archivos biológicos, con las cuales se puede aplicar métodos biológicamente relevantes. (Jeff Chang, 2022). Es el caso del presente estudio, Biopython y el uso de librerías complementarias permitieron acceder a la base de datos en línea de información biológica NCBI, para recuperar

las secuencias nucleotídicas de los genes HOXC6 y abd-A, además se aplicaron funciones adicionales como: número de nucleótidos, número de adeninas y timinas, número de aminoácidos, complemento, transcribe, y translate.

Se han desarrollado diversos programas bioinformáticos que, a partir de una secuencia genética, predicen el número, la localización del gen objeto de la investigación, estructura proteica, localización exacta de exones e intrones especialmente en eucariotas; en el presente trabajo mediante el uso de bases de datos como Ensembl, UniProt, se encontró el gen HOXC6 ubicado en el cromosoma 12q13.13, mientras que el gen abd-A de localización en el cromosoma 3, además se logró determinar la presencia del homeodominio en ambas estructuras genéticas, el motivo de unión al ADN y la proteína homeobox codificada por ambos genes.

En esta investigación la caracterización estructural de los genes HOXC6 y abd-A, resultó similar a la planteado por: Eduardo E Robertis (Robertis,2014) que describe la caja homeótica en todas las especies estudiadas, cuyos genes HOX codifican factores de transcripción específicos, la misma posee 180pb que codifican para 60 aminoácidos denominados homeodominio, esta descripción se la realizó inicialmente en la mosca de la fruta *Drosophila melanogaster* y posteriormente en los vertebrados. Según Morata, durante los años 1984-85, Walter Gehring y Matt Scott observaron que varios de los genes homeóticos clonados tenían secuencias en común, específicamente, todos tenían una secuencia de 180 pares de bases, que se llamó homeobox, demostrándose que la homeobox codifica para un polipéptido básico de 60 aminoácidos, el homeodominio, que es capaz de unirse al DNA. (Morata, 2001)

En el genoma humano, según Paco (2020), de acuerdo a la semejanza de secuencia y la similitud de ubicación en el cromosoma, los 39 genes pertenecientes a la familia HOX estarían distribuidos en cuatro grupos: HOXA, HOXB, HOXC y HOXD, ubicados en distintos cromosomas:7p15, 17q21.2, 12q13, 2q31, respectivamente, lo que coincide con nuestra búsqueda de localización del gen HOXC6, que reveló la ubicación en el cromosoma 12q13. Nuestro estudio también concuerda con lo descrito por Grier (2005), quien asegura que los genes HOX son pequeños y están formados por dos exones separados por un intrón, el exón está localizado cerca del extremo 3' y contiene el homeobox de 180 pb que codifica el homeodominio de la proteína HOX, que constituye un motivo de 60 residuos y se presenta conservado tanto en la mosca de la fruta *Drosophila melanogaster* como en los vertebrados.

El hexapéptido obtenido cerca del homeodominio en este análisis estructural de los genes HOXC6 y abd-A,

coincide con lo planteado por Erselius (1990), quien sostuvo que la mayor parte de las proteínas HOX, contienen una secuencia de seis aminoácidos cerca del homeodominio, el mismo que funciona como sitio de unión para cofactores que incluyen proteínas TALE como PBX y MEIS.

Busturia (1989), menciona que el gen *abd-A* de la familia *Drosophila* presenta una región codificadora de proteínas y dos regiones reguladoras en cis; también contiene el “homeobox” codificador de una secuencia de 60-aminoácidos de largo, segmento polipeptídico que fue denominado como “homeodominio”, lo que concuerda con la proteína Homeobox y sus dominios, encontrada en la *Drosophila Melanogaster*, obtenida desde la base de datos Ensembl, en la que se evidenció además las posiciones de aminoácidos correspondientes y el sitio conservado homeodominio, que han sido motivo de múltiples proyectos mapeados.

En el presente estudio asimismo se determinó la ubicación del gen *abd-A* en el cromosoma 3, compatible a lo planteado por Bofarull (2005) quien sostuvo que el complejo de genes HOX, se encuentran en el cromosoma 3 de *Drosophila melanogaster*.

Finalmente, la comparación de la longitud del gen HOXC6 y *abd-A* reveló una diferencia significativa en la cadena de nucleótidos: el gen HOXC6 posee 13936 nucleótidos, mientras que el gen *Abd-A* contiene 22886 nucleótidos; lo que se ajusta a lo planteado por Bofarull (2005), quien sostiene que el elevado tamaño de los genes HOX y del HOM-C de *Drosophila melanogaster* se debe a la complejidad de las regiones reguladoras, que incluyen enhancers que promueven la expresión de los transcritos en dominios específicos, y regiones “barrera” que aíslan las diferentes regiones reguladoras. (Bofarull, 2005)

9. CONCLUSIONES

Con la metodología empleada para analizar las bases de datos y lograr resultados concretos que permitieron discutirlos con investigaciones previas, es factible concluir que:

1. El estudio y análisis completo de las estructuras genéticas permite tener una visión integral de su composición, determinar alteraciones presentes, predecir diferencias en su expresión, funcionalidad, y con ello entender el comportamiento, modificaciones y mutaciones a lo largo del tiempo, para tomar decisiones acertadas en la

prevención, diagnóstico temprano y tratamiento de diferentes patologías.

2. El uso de herramientas bioinformáticas potentes y bases de datos sólidas y actualizadas permiten la derivación al conocimiento biológico, principalmente al estudio de secuencias genéticas de ADN, ARN y proteínas que ya están identificadas, constituyendo además una fuente importante de información sobre la localización cromosómica, propiedades, expresión, función, relaciones filogenéticas.
3. El estudio detallado del gen HOXC6 y abd-A mediante Biopython, permitió diferenciar los nucleótidos que forman parte de la estructura del gen en el que se determinó una longitud y concentraciones muy superiores en el gen abd-A.
4. El análisis comparativo del largo de la secuencia del gen HOXC 6 y abd-A reveló que éste último presenta 8950 más nucleótidos que el gen HOXC6, puesto que HOXC6 presenta 13936 nucleótidos, y el gen abd-A, está formado por 22886 nucleótidos.
5. La localización de los genes HOX: abd-A y HOXC6 mediante bases de datos primarias identificaron que el gen HOXC6 está localizado en el cromosoma 12q13.13, en tanto que el gen abd-A está localizado en el cromosoma 3 (Cromosoma 3.16,807,214-16,830,049), presentando ambos genes el homeodominio que codifica la proteína homeobox, tipo antenapedia, con su sitio conservado.
6. El análisis de la estructura de los genes HOXC6 y abd-A reveló diferencias en el número de exones, intrones, longitud de transcripción y número de aminoácidos resultantes de la proteína codificada, con predominancia del gen abd-A, que codifica una proteína de 590 aminoácidos en comparación con la proteína de 235 aminoácidos codificada por el gen HOXC6.

10. RECOMENDACIONES:

- Como perspectiva a futuro, se plantea la posibilidad de estudiar a profundidad el ordenamiento de las secuencias de nucleótidos del gen HOXC6, y los aminoácidos codificados, para así conocer las alteraciones que derivan en patologías, especialmente el cáncer, abriendo las puertas a una futura terapia génica.

11. REFERENCIAS BIBLIOGRÁFICAS

- Betti, M. I. L. (2016). Genómica poblacional del comportamiento de oviposición en *Drosophila melanogaster*. Universidad de Buenos Aires. Facultad de ciencias exactas.
- Busturia, A., Casanova, J., Sanchez-Herrero, E., Gonzalez, R., & Morata, G.(1989.). Genetic structure of the abd-A gene of *Drosophila*. 107, 575-583.
- Bofarull, B. N. (2005). Caracterización Genómica y Funcional de las Reorganizaciones del Complejo de Genes Hox en *Drosophila*. Departamento de Genética y de Microbiología, Universidad Autònoma de Barcelona. Barcelona: Facultad de Ciencias.
- Cantile, M., Franco, R., Tschan, A., Baumhoer, D., Zlobec, I., Schiavo, G., Forte, I., Bihl, M., Liguori, G., Botti, G., Tornillo, L., Karamitopoulou- Diamantis, E., Terracciano, L., & Cillo, C. (2009). HOX D13 expression across 79 tumor tissue types. International Journal of Cancer, 125(7), 1532-1541. <https://doi.org/10.1002/ijc.24438>
- Castagnino, J. (2004). La *Drosophila melanogaster*, el genoma y el cáncer. Acta bioquímica clínica latinoamericana, 38(2), 147-150.
- Centro Nacional de Información Biotecnológica (NCBI): Biblioteca Nacional de Medicina (EE. UU.), Centro Nacional de Información Biotecnológica; [2004]. Disponible en: <https://www.ncbi.nlm.nih.gov/>
- Cortés, C. G. (2020). Estudio de la función del gen Hox Ultrabithorax en la morfogénesis del halterio de *Drosophila melanogaster*. Universidad Autónoma de Madrid. 129.
- Del Saz Soler, D. (2018). Estudio de la regulación de la proliferación y la formación de patrón en el desarrollo del abdomen por los genes Hox Ultrabithorax y abdominal A de *Drosophila Melanogaster*. Universidad Autónoma de Madrid.
- De Robertis. (2012). Biología Celular y Molecular. Edición 16ª. Ed Promed. Buenos Aires.
- Erselius, J. R., Goulding, M. D., & Gruss, P. (s. f.). Structure and expression pattern of the murine Hox-3.2 gene. Departamento de Biología Molecular. Vol 110, 629-642
- Feng, Y., Zhang, T., Wang, Y., Xie, M., Ji, X., Luo, X., Huang, W., & Xia, L. (2021). Homeobox Genes in Cancers: From Carcinogenesis to Recent Therapeutic Intervention. Frontiers in Oncology, 11, 770428.
- Fernández, M. (2012). Herramientas computacionales para la comparación de genomas y detección de genes ortólogos con un enfoque de grafo bipartido. T. Universidad Central “Marta Abreu” de Las Villas.
- Gaunt, S. J., & Gaunt, A. L. (2016). Possible rules for the ancestral origin of Hox gene collinearity. Journal of theoretical biology, 410, 1–8. <https://doi.org/10.1016/j.jtbi.2016.09.009>
- Gehring, W.J., Yan G, Qian, t. Martin B, iilieter, t Katsuo Furukubo-Tokunaga, '5 Alexander F. Schier, 'Il Diana Resendez-Perez; Markus Affolter; Gottfried Otting, t. (1994). *Homeodomain-DNA Recognition*. 78, 211-223.
- Gilbert, P. M., Mouw, J. K., Unger, M. A., Lakins, J. N., Gbegnon, M. K., Clemmer, V. B., Benezra, M., Licht, J. D., Boudreau, N. J., Tsai, K. K. C., Welm, A. L., Feldman, M. D., Weber, B. L., & Weaver, V. M. (2010). HOXA9 regulates BRCA1 expression to modulate human breast tumor phenotype. Journal of Clinical Investigation, 120(5), 1535-1550. <https://doi.org/10.1172/JCI39534>
- González, I. (2007). Función del gen dRYBP en *Drosophila melanogaster*. Facultad de Ciencias. Universidad Autónoma de Madrid. <https://digital.csic.es/handle/10261/8342>

- González-Astorga, J. (2021). *Drosophila melanogaster*: La mosca del vinagre- *Drosophila-melanogaster*. Portal Comunicación Veracruzana. Instituto de ecología. <https://www.inecol.mx/inecol/index.php/es/2013-06-05-10-34-10/17-ciencia-hoy/1291Drosophila-melanogaster-la-mosca-del-vinagre>
- Grier, D., Thompson, A., Kwasniewska, A., McGonigle, G., Halliday, H., & Lappin, T. (2005). The pathophysiology of HOX genes and their role in cancer. *The Journal of Pathology*, 205(2), 154-171. <https://doi.org/10.1002/path.1710>
- Guerrero-Preston, R., Michailidi, C., Marchionni, L., Pickering, C. R., Frederick, M. J., Myers, J. N., Yegnasubramanian, S., Hadar, T., Noordhuis, M. G., Zizkova, V., Fertig, E., Agrawal, N., Westra, W., Koch, W., Califano, J., Velculescu, V. E., & Sidransky, D. (2014). Key tumor suppressor genes inactivated by “greater promoter” methylation and somatic mutations in head and neck cancer. *Epigenetics*, 9(7), 1031- 1046. <https://doi.org/10.4161/epi.29025>
- Jeff Chang, Brad Chapman, Ido Friedberg, Thomas Hamelryck, Michiel de Hoon, Peter Cock, Tiago Antao, Eric Talevich, Bartek Wilczyński. (2022). Tutorial.pdf. Biopython 1.80.
- Keller, A. (2007). *Drosophila melanogaster's* history as a human commensal. *Current Biology*, 17(3), R77-R81. <https://doi.org/10.1016/j.cub.2006.12.031>
- Levine, M, Tjian, R. (2003). Transcription regulation and animal diversity. *nature* 424: 147-151.
- Lin, J., He, J., He, X., Wang, L., Xue, M., Zhuo, W., Si, J., Wang, K., & Chen, S. (2020). HoxC6 Functions as an Oncogene and Isoform HoxC6-2 May Play the Primary Role in Gastric Carcinogenesis. *Digestive Diseases and Sciences*, 65(10), 2896-2906. <https://doi.org/10.1007/s10620-019-06013-7>
- López-Romero, R., Marrero-Rodríguez, D., Romero-Morelos, P., Villegas, V.,Valdivia, A., Arreola, H., Huerta-Padilla, V., & Salcedo, M. (2015). El papel de los genes del desarrollo tipo HOX en el cáncer cervicouterino. *Rev Med Inst Mex Seguro Soc.*, 53(2), 188-193.
- Luo, Z., Rhie, S., & Farnham, P. (2019). The Enigmatic HOX Genes: Can We Crack Their Code? *Cancers*, 11(3), 323. <https://doi.org/10.3390/cancers11030323>
- Moon, S.-M., Kim, S.-A., Yoon, J.-H., & Ahn, S.-G. (2012). HOXC6 Is Dereglated in Human Head and Neck Squamous Cell Carcinoma and Modulates Bcl-2 Expression. *Journal of Biological Chemistry*, 287(42), 35678-35688. <https://doi.org/10.1074/jbc.M112.361675>
- Morata, G. (2001). La historia de los genes homeóticos. Orígenes de la Biología del Desarrollo. Consejo superior de investigaciones científicos., *Arbor*, 662, 229-246. <http://arbor.revistas.csic.es>
- Paço, A., Aparecida de Bessa Garcia, S., Leitão Castro, J., Costa-Pinto, A. R., & Freitas, R. (2020). Roles of the HOX Proteins in Cancer Invasion and Metastasis. *Cancers*, 13(1), 10. <https://doi.org/10.3390/cancers13010010>
- Parra-Soto, S., Petermann-Rocha, F., Martínez-Sanguinetti, M. A., Leiva- Ordeñez, A. M., Troncoso-Pantoja, C., Ulloa, N., Diaz-Martínez, X., & Celis-Morales, C. (2020). Cáncer en Chile y en el mundo: Una mirada actual y su futuro escenario epidemiológico. *Revista médica de Chile*, 148(10), 1489-1495. <https://doi.org/10.4067/S0034-98872020001001489>
- Perera, O. H., Callicó, A. M., & Pérez, J. C. R. (2006). ¿Qué son los genes Hox? Su importancia en la enfermedad vascular y renal. 11. *Revista Nefrología*. Volumen 26. Número 2.
- Quinonez, S. C., & Innis, J. W. (2014). Human HOX gene disorders. *Molecular Genetics and Metabolism*, 111(1), 4-15. <https://doi.org/10.1016/j.ymgme.2013.10.012>

- Rezsohazy, R., Saurin, A. J., Maurel-Zaffran, C., & Graba, Y. (2015). Cellular and molecular insights into Hox protein action. *Development*, 142(7), 1212-1227. <https://doi.org/10.1242/dev.109785>
- Rux, D. R., & Wellik, D. M. (2017). Hox genes in the adult skeleton: Novel functions beyond embryonic development: Hox Genes in the Adult Skeleton. *Developmental Dynamics*, 246(4), 310-317. <https://doi.org/10.1002/dvdy.24482>
- Saha, S. S., Chowdhury, R. R., Mondal, N. R., Roy, S., & Sengupta, S. (2017). Expression signatures of HOX cluster genes in cervical cancer pathogenesis: Impact of human papillomavirus type 16 oncoprotein E7. *Oncotarget*, 8(22), 36591-36602. <https://doi.org/10.18632/oncotarget.16619>
- Shenoy, U. S., Adiga, D., Kabekkodu, S. P., Hunter, K. D., & Radhakrishnan, R. (2022). Molecular implications of HOX genes targeting multiple signaling pathways in cancer. *Cell Biology and Toxicology*, 38(1), 1-30. <https://doi.org/10.1007/s10565-021-09657-2>
- Solís, Calero, C. (2008). Identificación in silico de un grupo de secuencias ortólogas conservadas (COS) de *Ipomea batatas*. *Revista Peruana de Biología*, 15(1), 79-84. <https://doi.org/10.15381/rpb.v15i1.1679>
- Soliz, Dénise. (2011). *Mecanismos Moleculares del Cáncer*. Universidad de Cuenca: 2011.
- Tang, L., Cao, Y., Song, X., Wang, X., Li, Y., Yu, M., Li, M., Liu, X., Huang, F., Chen, F., & Wan, H. (2019). HOXC6 promotes migration, invasion and proliferation of esophageal squamous cell carcinoma cells via modulating expression of genes involved in malignant phenotypes. *PeerJ*, 7, e6607. <https://doi.org/10.7717/peerj.6607>
- Tolosa, A. (2015). *Drosophila Melanogaster* y cáncer. *Genética Médica News*. Genotipia. https://genotipia.com/genetica_medica_news/drosophila-y-cancer/
- Uribe, R. A. J. (2019). Estudio de la expresión de genes Hox y de la función de la proteína "Homeodomain interacting protein kinase" en procesos de regeneración en *Drosophila melanogaster*. [Doctoral Thesis]. <https://repositorio.uam.es/handle/10486/691192>
- Verena, M. (2017). HOX genes in pathogenesis of Ewing sarcoma (Doktors der Medizin). Webber, C., & Ponting, C.P. (2004). Genes and homology. *Current Biology*, 14(9), R332-R333. <https://doi.org/10.1016/j.cub.2004.04.016>
- Yahagi, N., Kosaki, R., Ito, T., Mitsuhashi, T., Shimada, H., Tomita, M., Takahashi, T., & Kosaki, K. (2004). Position-specific expression of Hox genes along the gastrointestinal tract. *Congenital Anomalies*, 44(1), 18-26. <https://doi.org/10.1111/j.1741-4520.2003.00004.x>
- Yamamoto, S., Jaiswal, M., Charng, W.-L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N. A., Xiong, B., Zhang, K., Bayat, V., David, G., Li, T., Chen, K., Gala, U., Harel, T., Pehlivan, D., Penney, S., Bellen, H. J. (2014). A *Drosophila* Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases. *Cell*, 159(1), 200-214. <https://doi.org/10.1016/j.cell.2014.09.002>
- Yang, Y., Tang, X., Song, X., Tang, L., Cao, Y., Liu, X., Wang, X., Li, Y., Yu, M., Wan, H., & Chen, F. (2019). Evidence for an oncogenic role of HOXC6 in human non-small cell lung cancer. *PeerJ*, 7, e6629. <https://doi.org/10.7717/peerj.6629>
- Zhang, F., Ren, C.-C., Liu, L., Chen, Y.-N., Yang, L., & Zhang, X.-A. (2018). HOXC6 gene silencing inhibits epithelial-mesenchymal transition and cell viability through the TGF- β /smad signaling pathway in cervical carcinoma cells. *Cancer Cell International*, 18(1), 204. <https://doi.org/10.1186/s12935-018-0680-2>

12. ANEXOS:

Análisis de Secuencias con Biopython:

Una vez obtenida la secuencia de los genes HOXC6 de *Homo sapiens* y abd-A de *Drosophila melanogaster* se trató los objetos Seq como si fueran cadenas normales de Python, y se pudo obtener la longitud o iterar sobre los elementos de la cadena.

Se agregaron las librerías necesarias para realizar el análisis de la secuencia del gen; e importamos la secuencia del gen HOXC6 y abd-A del NCBI

```
In [7]: from Bio import SeqIO
        from Bio.SeqUtils import GC
        for seq_record in SeqIO.parse("/usr/local/bioinf/sampled/data/t-coffee/sequence.fasta", "fasta"):
            print(seq_record.id)
            print(repr(seq_record.seq))
            print(len(seq_record))
```

```
NC_000012.12:54016888-54030823
Seq('GCAGAACAGTCCCTCCCTGTAAGAGCCTAACCATTGCCAGGGAAACCTGCCCTGGGCGCTCCCTTCATTAGCAGTATTTTTTTTAA
13936
```

Se guardó la secuencia para su manejo de manera más sencilla

```
In [7]: seq_1 = seq_record.seq
```

Se imprimió la secuencia para comprobar la integridad de la misma.

```
In [8]: print(seq_1)
```

```
GCAGAACAGTCCCTCCCTGTAAGAGCCTAACCATTGCCAGGGAAACCTGCCCTGGGCGCTCCCTTCATTAGCAGTATTTTTTTTAA
TTAATCTGATTAATAATTATTTTTTCCCCATTTAATTTTTTTTCTCCAGGTGGAGTTGCCGAAGCTGGGGGCAGCTGGGGAGGG
TGGGGATGGGAGGGGAGAGACAGAAGTTGAGGGCATCTCTCTTCCCTCCCGACCCCTCTGGCCCCCAAGGGGCAGGAGGAATGCA
GGAGCAGGAGTTGAGCTTGGGAGCTGCAGATGCCCTCCGCCCTCCTCTCTCCAGGCTCTTCTCTGCCCCCTTCTTGCAACTCT
CCTTAATTTTGTGGCTTTTGGATGATTATAATTATTTTATTTTGAATTTATATAAAAGTATATGTGTGTGTGTGTGGAGCTGA
GACAGGCTCGGCAGCGGCACAGAATGAGGGAAGACGAGAAAGAGAGTGGGAGAGAGAGAGGCAGAGAGGGAGAGAGGGAGAGTGAC
AGCAGCGCTCGGTAAGTGTTCCTTATTGGTTCAAATATGTAATAATGGTCCGTGGCGGTTTCAAGAATAGGACAGCAGCGGCTT
CTCTTGGTTAGGGAGAAAGAGGGGGCTGGCGAGAGGGTGGGCGCGTAGTTAATTATGGGAGGAAAATTTGGCGGGGGAGAGAAACG
GGGTAAACTGGAATACCGGTAACCTGGTAGTGGGGCGATTGTGTGTGCTGATGTCACTGTGTGCCAAGTTGGGGGTGGGGAGGCA
AGGGAACAGCACCTGTGGGGCTCAAGCCGGCGGTCTAGGGCGCCAGCTCTCTGCTGCCCACTGTGTGAGTACCTGCTTATTATAGA
TCTTCGGTCTCTTGGCTAATGGAACCTGGGTCGGAGCGGAACCTGGAAGCTCAAGCTCCAGGCAGGAGAGAGAACTTAGAGGCTGTGC
CCATTGCCCCCTAAGCAAAGTTAGCCCCCAACCCCCAAACAACTTTCCAGCTTCCCTCCCTGTCTGGCAGCCCCCTCCCTCCC
AGAGAGCGCGACTGCTAGAGCTCACACATGCGCAGTGTGGGCCCAGGGCCGGGCCGCGAGCAGGAAGCCGGCGCAGCTAGGCGGC
CGGCGGGGCTGTAAATTGGCAATTAGGGGGGAGGCTGGTGGCTGGTGGCGTCCAGCCGAGAGGAGAGCGTCTGCCACCCCCTGC
TCCCGCCCCCACTCGGGCGGATGGAAGGTTGGGAGGTGCCCTGCGTTGGGTGGAGGTTGGAGGTTGTAGGGTGGGGGTGGGGGATG
CTGTACTCAAAAGCCATCTTGTGCTCAGAGAAAAGAGGCTACCGGCTTTCCCTTCCGGGGTCCGGCGCCCCCTACCCCCAGCCGC
GGCCATCCAGCCGGGATGCCCCTGAGCCGGGATGCCCGCTCGCCACGCATGGCTGCTCTGGGCTAGGACCTGCCTCGCCTCGGG
```

GGGAGGTGGAATGGGGAAGGCCCTAGGCTGGGCGGGGACAGGCCAGTCTCTGGTAGAGTCTCTAGGTGCTGTGGAGGCAGT
 GAGGATTCCGGGGGCTGCTGGCTTATGGGGTAGAGGCGTCAGCCTGGGGCGGCGGGAGGGTCTGGGCACTCGGGATCTCTCTAAT
 AATTGGTGCTAATGGCCTGTTGTATAGTCTTAACCCGACGCATATGGTTTCATAACACAGTGGCTTAATTTCTTCCAAATGTACGT
 GGCCCTGAATGTGTGAGTTTGATATATGACCCCGCCTGCTCCCGGCTGGGTCCGAGGCTCCGAGGCGCCCATAGTGTG
 AATTTCTGTTTATAAACGGCGACGCACACGCAACGCCTACACACCCAGGCAGACACACCCGGAGGCGTGGGGGAACATATTCAG
 ACGAGGTCCGAGAGAACCTGGGGGCCAGACTTACCTTAATCTAAGCCAAGACACCCCTCTACCCCCACCCTCTCCAGAGGCGCG
 ACTGGGATTCTGCGCCAAGTCCCAGTGGGGGAGGGGATCGCTCCACGATTTCTGTTTAAACCCTGACCCGTCGCTGTGGGGGGG
 CGGGGATGGGGGAAGAAACCTGTTTTTATCAAATTTTAAAAATACAGTTTGCATTAGATCAAACCGTTTGGGCCAGAGCTGACG
 GTGAATTTTATCTCCCTCCCCCACAAACCCCTTCCCTTCTTTCTGCCCCCGCTAAATTTCCCTTTTCGCAGCCATCATAAAGG
 CCTCTGAGGTATTCTCCCGCGGGAAGAAATGGCATTCTCTCCCTCCCCACCCCCACCCCACTAGGACTTGTGTGTCTCGG
 CAGAGGCGTCTGACGAGGGGCTAATTTTCGCTTTCTTGTACGATCAGAGAAAGCGCTTGGCTCTCCCCAAATGGGCCCAGCCCG
 CCGCTGCTCAGGCTGCCGCGCTCTCCCGCGCTCAGGCACGGCCAGTGGCTGATGCCCGCGGATCCAGCCTGCGGCAGGCC
 GGAACCCGGAGTCTGTGGGACCTGAAAAGGGCAAAGGAGTGCCTGGGACACAAGAGGGGCCGAGAGGCCGAAAAAGGGAGCC
 TTTCTCTGGAGTAATGTGGGAATCGCCGACCGGTGAGGCTCTATTTCTCTCTTTCGCTGCAATTTGTCCGCCCCGCCCCCTGT
 GGATGGCCGCTGATGGGGGGGAACACAGTGTCTTTGTCCGTCCCGGAAAGAAACCTTTTCCAGAGAAGGCCATTTCAGGCCAGT
 GGAGATTCACCTAGAATATTAGAAGGGCCCCCTCCACCCGAGGGGCGGGCTGGGCGAGGCTGAGGTATTTCTCTTAATCCTAGA
 ACGACCTCACTGAGTGTACCCACAGCGAGATTTGGAGGTGCTAGAGTTCGAGGTTGGCCGCTTAACAGCACCATTCTGAGTGTCT
 TTATATAATATTTTAATTGATTTTCTGCTTCCCTTGGTTACAGCCCTGAGAGTCCCTTCCCTTCTGCTTCTGCTTCTGCTTCT
 GACCCATCCTTACAAGACAGAGGCAAGCAGAGAAGGCATAGCAGCAGCAGCGCGCTCTGTTTTCATTTTCCACTCTGGCCAGGG
 GATAAACTGGACCCAGTGGACTCCAACACCAACTCTCTGCTTCCAGTTACTTTTCGATTTTCTTTCCCTTCTCTTCTTCTCTC
 CCTCTCTGGCATTCAAACAGAAGCACAGAGGACCCACAGTGGCAACTCTAAGCCTCCTTGCAGGCCACATTCCCTGCTGACGTG
 GTCCTTCTGTCAGGAGAGCCCTTCTGTGATTATCTAATAATAGCCAGGGTGGCTGAGAATAGAATCTGGCTCAAGGACTCTGCACA
 CCGGGGACAGACACTGAATCTGTGGGTGAGGCTGGCAACATCTTGGAGCCTCCCCCTCTGAGCCCACTCTTTGGGCAACC
 TTAGGCTTGCTGCCCCCTTATAAGGGGCTTTGTGACCCCTGCCATCCGCTGTATTTTCATGTGCTGTTTCTAGCTAAGGAGAAT
 TGGGACAGGACAAACCTTTCTCACCTTGGTCTGACAGTTGTGATCCCGACAGAATGCGACCTTGTTTGGAGAAGGCCACCTTGTG
 GGTGGCTGTGTTCTGAAACAGGCTTGGGAGTGTATATATTTATATATACAGGCTTAAATATACAGACTCAGGCTTATGCATG
 AATATATACAGACTCATATTTCTTGGTGGTTAATTGTATATTACCATGCCTATATTCGCTCAAACACCATTATTTATGTTTCATGCAA
 AATTTTGTAGTCTCTACTAGATGAGAGTGGGCTCATGGAGGCACCGGTGCACTGAGCAGGCCCTCTGTGCATGCTCTGCTGGAA
 ATATGTGCTCACTGGAATCGGAGTATGCGGGTTCCAGTAACCAATGGATGCCATAAAGGCTTTCTCAGACCTTATTGCTTTTGTG
 CCCAAGCAGCTGGACTGGGGAATACTAAATCCCTTCTGCTTTCTCTCCACCTCTTGCAAACTCTCCCACTTAGAAATTTAT
 TTACAAGGAGAGACAGTTGTCTTGAATTTGAGTTCTGCGGATTTCTTTCACGGGCATGGTGGGGCTGCAAGCAGTTCTGGAGTAGT
 CAGGGCCCTGGCGCTGGAAGCCTCACTGGAGCGATGAAGGGATCAGAGTGGTTAATAAGGTGGCTCTGGACAGTGATCTGGGGCA
 GCTGCTGGCCAGGGTGTGTTGGAGAAGCTTGTGGAAAAGGCAAGCCAAAAGGGGTAGGGATGCAGTGAATATCTGTTTCTCCCCCTG
 CTTGCCCCAGCCGGCGCTCACCCCGCCCCAGCCGATTTGCTGCTATCCGTGGCGGCAGATTGGGAGTTCTCTAAGTTCGCCCAA
 ATACCCAGATCTGGTTACTGGATGAAGTGCCTTTTAACTCCTGCTGCTATCTGAGCCAGGACAAGACCTGGTCAGGGAGACGGCTG
 GTCCGGGACAGCAGGGAGAGGACGCTGGACCCGCTGCTGTGGCTGGAGCACTGAGTATGGGCTACCAAGGCTGAGAATTGATGG
 GCCTCCTTTCCAGCCTTTTCCCTTCCAGAGACAAAAAGACAGAGCAGAGGGGAAATATCAATGTCAAGTGGAAAACCCCATAA
 AAGAAAGCAGGACAGCGGGGCTGGAGGGTTTTGCTGCTACTCTCTCAGGTGATCGGGTTCCCTCCCAAGACTTCTGTCCAGA
 CAGGCTGCACCAAGAGGTGTGCAGCTTTCTTCCACCTTCTCCCGCCCTTCAAGGAGGCGAGCCTTAGGCTTTCTGGGCTTGGTC
 ACCCCAGGCACTGGGGTGACATTACTCCAGGATTGTGCCCTCGCCAGCCACGGCTCTGCCACGGCGCCGACGGCTCTCTGT
 TTATTGTGCTGTGAGTTGTTTACTTGTGTTGTGCTGCTGCCACCTGCGTCAGAAACCCGCCAAACACGCAACTCGCCAGGAGA
 ACAAGTGCTGTGCTATCCCCGACTCCCTCCCTTTCCAGCAACCCCTCTGCTCTCACAAGCTCCTGCATTCTCCCTTTTCGGA
 GGGGAGGAAGCAGAAATAGTCTTAAACAGGGCAGCTGTTTATAGGGTCAAGGGGAATTATGGGGACACCATAAACCTCTGGGTGTC
 CCTGTGTGAGCAGACTTCTGAGCAGCCCTTCCCTTCTAGGCTTTTCTTCTCTTCCCTGGTGACACTTGGTTAAGGGGCCAC
 ACTCATTTAGAAAGTTGAGCCACCCCATCAACCCAGTCCCCACTACCCCACTCTTTCTGCAGAGTTTCCCTGAGACACAGG
 TTGTTGGGTTTTAAATCAGGGTAATGTTGGTGGATTCTAAGCTAGATATAGCATGTCATTCTTCTCCCACTCCCTCTCTCA
 GCCCCCAAAATTCAGTGAATTTAAGCAAATCTTACAAAATTAATTTGCCCCCTTGTCTAAAGATCTTATATGTCTCCAATCTGCTC
 CAGGTGTAGAGTCAGATAGTTGGCTTTTGGCCTTTAATTTTGTGCTGCTAGCTATTAAATTTAAAGAGTATCATTTGAAGTAAGTAAT
 ATTCTCATTTCTGAAATTAATTTTCCCTCCCAGTCTCTCTCTCTCTTTTGGCCAGCACAGATTAATGGTTCTTTCAGATCTG
 CAAACAATAAAGGATAAAATATTTATAGGGAAATGAGGATGCTGTTTTATTGTTGGCATATTGTTCCAGAACATAACAACATTTT
 ATTTTCAAGCAGGAGAGTTAGACCAGGGGGAGAAAAGGGGGAGAAAAGGAAAAAACATGAAAGGAGGGCAGGAACCCACACAAAA
 CTCCTGCCATGTATAACTGGTAGGACATTTACAATTTCTATATGACAACTCACTGTGATGTATAATACCGAGTTCTTGAATTTTA
 TAATTTCTCATGAACCTCTCAAATGGGTCCAAGCAGTGTTCATGTGGAGGGGTATAGAAGAAGCCAGGAGGCTGGGGATCCTTTGA
 CCCAGTGACAGCAGCACACCCCTTATGCGGCTTCTTCTCATGTGACATCCTCTTGGGCGCTATCTGGGGGCTGAAAATATT
 CTGGATATTCTGCATTTTAATTTTTCTCTGGGCCAAGCAGGGCCACTTGCTTTGCCCCCTCAAGTGTCCCTAGGCATCTAAGCT
 CTCTCCCCATTCCCAGATCACAGAGAGCTGTGTCCCCAGGCTCTCCCCAAGCCCCCTCTCAGCTAATACTCAGTGTCTTGAATA
 GAGGTGGGAGGCTTGGGCACAGCCCCCAAAAAGCCAGAAATGATTTTTTCCAGATGGAGCTGGAAGGGGGCTCAGGCTGAGCGT
 GATGGGGTAGAGATGTTTGGGACCTATTTACGTTTTTAACTTAAGTCAAGAGAAGTGAATTGAGAAGCCAAACAAATGCCAATGTG
 AGAGTTTGTATCTGCTGATGGTAGCAGCAATAAGTGAAGTGTGGTAAGGAAAGGTCAGGCAGAAATCTGGAAAGAGAGAGGGAGG
 AGAGGAAATGTGACATGAGTTAGCAGAGCCCTTCCATATACCTTTCTCTCCGTTTAAATGTAATAAAGCAAAAGCAGCTCGGGTA
 GCCTCTCAACCTCTTAGCTCAAAAGAGGAATCACCAGCGCAGGCCCTTGAGGAGTGTGTTGGGGCAGTTCTCAGCAGCTTAACATGCC
 CAATCTGCCCTCTTCTCTCTTTCCCCCTCCCTGCTGATCCCCAGCTCCCTTCAACTGGTCAGCCCGGGCCCTGTCCCCAAAATATT
 AGCACCCCAAAATGATATCCATTTCTGTGTTAGGTGTGTCAAATTTCACTTCTCTCCATGCCTAACTTTGAAAGGAGGAGTGTAC
 CTTTTTGAATGCGGTGTGTGTAGCGGGGCGAGGTATGTGCTCAGACAGCTTCCCCCATCATATCTATGGGAACTGGAAGCCCCCT

TAGGTGAGCCATGCTTTGGTGTTCCTCCAAACAGCATTCCCTGCATTTCAGGTTCTCTGGCAGACCAGGAGGCTCTTTGCGATCTGATC
 CTTTGGACTAGGAGTTCCATGGAGATCAAAGGAGCTTATTTATACACAAAACACCAGGCATTTGGAGTAGAATTTATTTTTTAATTT
 CTAATAAAAAAGTCTAATCATAGACCAATCGGGCCAACTGAGCAGTCGCTCCAGAACAGCAAAGCCTTAACCAAATAGTGTGATT
 ATGAGGGGAGCTGAGCAATTACTGAACTCGGCTTTCTCCGTAAGAGGATTTTAGCCCTCGCCAGTGGCACAAAAGGGGCCAGC
 GGGGCAACCGTTCTTCTTCTGCTCAGAAAAGTGAGACTTGGGGGTGGTGGGGGCAGGGTCGGGCAGCTGCGGTGCGCTCCAGGCAG
 CTGAGCACCCAGGCCAGGCGCTGAGGCTGAGCACTGAGTCTGAGGCTCCAGTGCTTGGGTCCCCGAGCAGCCTCAGGAGGCACCG
 AGCTGTAGTTGCAGCGCGGCAGTCAGGAATGGAGTATGTGTATCTGTGTGCGAGGGTTGGTGAGTTACAAACATCCGCGCAGAGTT
 AGGGGATCTGGGGAGAGAGAGGAGACAGGGAGAACACAATAAGGAGGACCTCTGGAAAATATTGTCCCCAAATAAAAAATTCCAGA
 TCTCATCCCTCTCACTTCAAATATCCCTCCCCAGGAAATTAATATATTTCCCAGAGCTGAGTGGAAGCAAAGGCCAGCAAGGTGA
 GGCCTGCCACGAATTTCTCAGGCTGGGTCCAGTCCCAACCCACCCCTTTCCCAGTTCAGAGCTCACATGCCCCAGCCTTTCCCT
 GGCCTAAAGGGGAGTAAGGCTGGAGTGGTAGCATTTTAACTTGCTTAGGAAATCCCAAGTGCCAGGCTGCAATGCCACCTTAT
 CTTGGGGACTCCAGGCCAGTGACTGCACCCACTCAGGTCCAAGGCGCAGGCAGTAAATTTCTCTCCTATAATTAGCGCTATCAA
 AAGCATGTGTTTTGCATTGGAGAGGACTGGAGAAGCAAGTGGCAATTTTCCCTTGATTTCTTTTTCTGTGTTTTTTGAATTTCTGT
 TTTTGAATCCACTAATTAAAGCTTCTCTGAAAGTGAAATCTGCTCTTGACACCAAGCCCTTCCGAATGCCTGATTTGGGAGGCT
 AAGGGGCCAGAGCAGTAGCAAGGAGACCTGCAAGCCTGATGGGGTTGGGGGGTAGTGTCTCTGATTTCTTTAGGCTCGTTTTATA
 GAAACAAAGAACAAGTTCTCTCCCTCCTTCCCAAGACTTGTGCATATTTACAGAGCACAACCTGCAGTTTTAATAAAACATCTGTT
 CTTGCTTCTGCTGATGAGTTTCCATAATTGTTTTAGACAAATTTAACAGGAAATATAAATATATTTAGAAAAATCTAAGTCCCA
 GCTTTCCCTTATAGACAGCCATTCTTTGCGAAAAAATAAATTAACATGAAATGGGAAAAAGAAAATATAATGCCAATGTTAGG
 AAGATCGGGCTCATCTGACCTGCTCTCCCCCTTGGAGCGAATCCTTTCAGCAAATTCAGCTGCATAATTAAAGATCTAAGAGAA
 AGAAGAGGGCTTCTTCTTCTCAGGAATGAAAGGTAATTGGGGGGGGGGAGGTGTGAAAATTAATTTTGCCAAAGGTCTTTAA
 GCAGTGAGGAATTTATTTGTATTTCTGTTTCCCTTCTCTCCCCCTTCCAGCATTTCTGCCTTTTTTACCCAGCAGCTGGTTTTCT
 GTTCATTATAAATCGGCGAGACCTTTAGTCTCTGCCCTCTGTTTAGGGACGTCATAAAACACTTAACTTTCTGGGGCTGAGGCTC
 ACATTTGCTCTCAGTTCGCCAACCTCCCTGAGCCCCACCGACTTTAGGCCCCAATAGAGAATCTCCCCAAGTTTGGGGCCCCCTT
 CCTCTCCAGACATGGTTGAGGAGGGGCTCAAGGAAACCTTAGGCATAGCATCTGCTGGGGGCTGTAGTCTGCTGCCCAAGAGCCCT
 TCTTTCTTCCAAGGTCTTCCCTCCCCACCCAGACATAAAAGAGATTTTAAATACATAGAAAATCAACACTCATTGAAAGTGAA
 ACCTCATTAAGACATCCTATCTTCCATCATTCAATTTGTGTACATTGCAACAATTTAATATCTTGAAGGAAATATCACTGACATG
 ATTTATCCCTCTAGCAATCTCTCTCTGAAATTGGAGGGGGGACAGAGTTACTCTCTCTTCCGGCCTCTTCTACTTGTACACCAC
 TACCTTAGGGCAGAAAGGAGGCTGGGAACCTCTGAGAGAGGCTACTTGTGTCCAGGGGTGCACTGACTTGTTCCTAGCTCAGTCC
 CCAGAGTACTCCTGCCTTGTGTCTCTCCCCACCTCTTGCCCTCCAGGGGGCTGAAGCCTCAGCCAGCTGTTGGAACCTCTGC
 TCAGGCATGCCATGAATTCGAGGACAGCTGGATTCTAGGGCAAAAAGCAACTCACATTGGTGTCTATGTTTGTGTACCTGCCAAGT
 CACCCCAAAAATGAGATAAGTATCATCTAAATTTTTTCTTAATTTAAAAATCCCTCCTCTACCTTTTCCCTCATTTCTCTGGGG
 AGAAGAAAGTGTGATCTAGAAGGAGGAACCTTTATAATTATAGGAGTCTTTCCCTTCCCTTCCAAATACCGACAAATTTCTGTCC
 TTTCTGGAGGATTGGGGCTGTATTTTCTAAACCAACACAAATTTGTCTGGCAGGAAGACTCTTCAACCCCTTAACGAGGTTCATGT
 ATTAACATAACATGTTGTCTCCACCAGCTCCTGCACACTGTCTGCTTTTGGGTTTGAACCTTTATTTTTTCCCCCTTAGCGACACTCC
 AGGGAAACCTTAACGTTACAGAAGATGAATAAACGAGTTTTTTCTCAGCATAGTCCGTTTATGGCTTTATTGCTACCCATTGATT
 ACTCCGCTTTGTACACAGAAGGAAAAGATCATAAAATCCGGCGACATCCGGGTCTTGTATAGCCAGCTTTCCCCCCCCCAAC
 CCACCAAAAATGGTGGGGGGGGGATATGAGCTTTCTCTGCTCCCCGACACTCCCCACCACCTTCTTTGTCAACTCAAGGC
 ATTTTCAGCCTATATCAGCATATCAAAATTAAGTTTGGACTAATCAGGAGAAAACCCCATCTGCACACTTTCCGCTGTTTCTGTTT
 GCCTACTGCGGCTCCCTCAAGCCTCGGATTGAGTGCCTTTTCGTAACCGTCTCTTCCCTTTTTTTAGCCCCAAGATGGGGGAA
 TAAATGGGCGTTGAGGTTGCTGAGAGGCCATCCCCACCACGGGTGACTGGCACTTGGGTGAGAGCTCCTTAGAACCCCTTTCTG
 AAGTCTCCAAGCCTCATAGCTCAGGTCCAGGGGCGCTCCACTGGCGGTGCCCTGAGTCTGGAGGGGGCAGGGGGGCTTCTGACCA
 AGGCAGGGGCTTGAGCTGCAGGAAGCTAACTGTCCATCAGAGCCCCACCAGCCTTCTCTTTCTGGCTGGATCCTATGTGAATTAC
 CCCACAGGCCCTCACTGGCTCCAAATAATAAATTTCTACCAACATAAACAGGAGTTGATGTCTCTAGAAAGATTCTCAGAGTTTTC
 GTTCTCTCTCTCTCTCTTCTCTCTGCTTTTCCCTCCGGTGCACCTCCCTACATTTTACTTCTTTTCTCTTTTCTATT
 GCATTTTTTATACATTTTCCCTTTTAGGCAAAAATGCTGGAAGAGTTTTCCTAACAGAGGGGCTCTGAGAAGGAGGCAGAAGGGAG
 CCAGGAGGCAAGGAGAGCAGATGCGTGGCCCATTTGCCACTCCAGCTCTGCCACTGTACCCTGAAGTCTTGCCCCCTTAGACCT
 AGTTGTTCCCCACATCCCTACTACATTCAAGACTCTCAGCCACACTCTACTTACTCAGGGGAACAGGTTTTATGAAATTTCTTTCT
 ACTTTTGATTTTTTTTTTAATTATTGGTGAAAAAATACTAGTTTAGTGATGGGACTCACTCCATGGAGTGTCAAAACAGGATTTCC
 TGCAGGATTTCTAGGGACCCCCCTCTGTGCTGGGAGGGAGTTCTTCCCTTGAGAAATCAGGATGTAATTTTAGAAGGGGGCAGA
 TATAAAACATGAAACAGATTGAAAATACAGGGGTTTCTCTAGCCCTGGCTCTGGCCACAGCAGCAAAGTTAGCTGGGGCTGAG
 AGATTTCCCTGGTAGTACAGCTCTCCTTCCCTGAGTCCCAAGGCCCTGCTCAGGGACATGCACCCCTCCTTACCACCAAACAGTT
 CCCTTACATATATATATATATATATATATTTTTTTAAAGAAATCCAAGTCTTACTTTCAAAGCACACACTTAGTCTCAACTAGGGA
 ATCAACAGAAGCAGAAGCGATTTTTTTTCCCTTCCCTGACATCTGGCTTGGGATTGGCTGGGAGGGGGTCACTGACTTTGTCTATT
 TTGTCTGCTTGGATTGGAGCCGTCCCTATAACCATCTAGTTCCGAGTACAACTGGAGACAGAAATAAATATTAAGAAATCATA
 GACCGACCAGGTAAAGGCAAGGGATGAATTCCTACTTCACTAACCCCTTCCCTATCTGCCACCTCGCCGGGGGCCAGGACGTCT
 CCCCACGTCGCCCTCAATTCCACCGCCTATGATCCAGTGAGGCATTTCTCGACCTATGGAGCGGCCGTGCCCAGAACCCGATCT
 ACTGACTCCCTTTTATTCGCCACAGGAGATGTCTGTTCAGTTCAGCGGGGGCGTATGACTATGGATCTAATTCCTTTTAC
 CAGGAGAAAGACATGCTCTCAAACCTGCAGACAAAACACTTAGGACATAACACACAGACCTCAATCGCTCAGGATTTTGAATTTGA
 GCAGGGCAGGACTGCGCCCCAGGACCAGAAAGCCAGTATCCAGATTTACCCCTGGATGCAGCGAATGAATTCGCACAGTGGTGAGT
 TTTACAGCTCCGAGATATAAATTAATAAACTGAAGTGGCTTTATGACCGCTTCCCTAGAAGAACGGGCGGAGAGAGTTTTTTAT
 GGCCCCATAAAAACAATCACGCTCGTCTCCTCACCAGACAGGGCCCCCTCTTCTGCCCCCTCAGCTCGCCTCTCGCTCGCTTGCA
 GGGGCTTGGCCCCCTTGGCCCCCTGACGGATTGGAGGGCCCCAAGGCGGGCTGAGGGGGCAGGAACAGGGCAGGGGATGAGGGGAGG
 GAGCAGAAGATAGGGGAGCCCCCAAAGTGGAGTCAGTCTGAGGAGAAGGGAGGGGGAGAGGAGAGGCAGATAAGTGGGGAGAAGT
 TTCAAAGGAAGTAGGATGCTTGGGGAGATGGGGGAGCCAGTCTTGAGAAAGGGGGCCCTCATCCCCATTATCTGGGATTTTTCA

```

AAAATCTGCTTTTAGACTTGCAGCTTCTGTTTGCCTTTTGCCCCGCCCCCGCCCCCCCCCACCACACACCTTTCTTTGGGA
CACACAGGTCCCACCTTATAGGAGGTCTGAAGTTGGGGTCCTCGCTGTACCCCCAGTGGGGGTGGGGCAAGGCAAAAGGGAGAAGG
CTAGAGCCCTAAGGAGGCTGTGAGCTGCTGGCCAGGTGGGAGGGTCAGGACTTTGCTAGGCGAAACAGTCTGCAGAGGACGCTTT
GCTGATTGCTTTTAGTGTGTTTGTGCCCCGGATCTTTAGGGGTCGGCTACGGAGCGGACCGGAGGCGCGGCCCGCCAGATCTACTCG
CGGTACCAGACCCTGGAACTGGAGAAGGAATTTCACTTCAATCGCTACCTAACGCGGCGCCGGCGCATCGAGATCGCCAACGCGCT
TTGCCGTGACCGAGCGACAGATCAAAATCTGGTTCCAGAACCGCCGGATGAAGTGGAAAAAGAATCTAATCTCACATCCACTCTCT
CGGGGGGCGGCGGAGGGGCCACCGCCGACAGCCTGGGCGGAAAAGAGGAAAAGCGGGAAGAGACAGAAGAGGAGAAGCAGAAAGAG
TGACCAGGACTGTCCCTGCCACCCCTCTCTCCCTTTCTCCCTCGCTCCCCACCAACTCTCCCTAATCACACACTCTGTATTTATC
ACTGGCACAATTGATGTGTTTGTATTCCCTAAAACAAAATTAGGGAGTCAAACGTGGACCTGAAAGTCAGCTCTGGACCCCTCCCT
TCACCGCACAACCTCTCTTTCACACGCGCTCCTCCTCCTCGCTCCCTTGCTAGCTCGTTCTCGGCTTGCTCTACAGGCCCTTTTCC
CCGTCCAGGCCTTGGGGGCTCGGACCCTGAACCTCAGACTCTACAGATTGGCCTCCAAGTGAGGACTTGGCTCCCCCACTCCTTCGA
CGCCCCCACCCTCGCCCCCGTGCAGAGAGCGGCTCCTGGGCCTGCTGGGGCTCTGCTCCAGGGCCTCAGGGCCCGGCTGGCA
GCCGGGGAGGGCCGGAGGCCCAAGGAGGGCGCGCCTTGGCCCCACACCAACCCCCAGGGCCTCCCCGCAGTCCCTGCCTAGCCCT
CTGCCCCAGCAAATGCCAGCCAGCCAGGCAAATTGTATTTAAAGAATCCTGGGGGTCATTATGGCATTTTACAAACTGTGACCGTTTC
TGTGTGAAGATTTTTTAGCTGTATTTGTGGTCTCTGTATTTATATTTATGTTTAGCACCGTCAGTGTTCCTATCCAATTTCAAAAAA
GGAAAAAAGAGGGAAAAATTACAAAAAGAGAGAAAAAAGTGAATGACGTTTGTTTAGCCAGTAGGAGAAAAATAAATAAATAAAT
AAATCCCTTCGTGTTACCCTCCTGTATAAATCCAACCTCTGGGTCCGTCTCGAATATTTAATAAAACTGATATTATTTTAAAC
TTTA

```

Se guardó el número de nucleótidos del *homo sapiens* en una variable para su comparación

```

In [9]: nuc_homo=len(seq_record)
        print(nuc_homo)

```

13936

Se obtuvo el número de Adeninas y se guardó en una variable.

```

In [10]: a_seq1=seq_1.count('A')
         print(a_seq1)

```

3266

Se obtuvo el número de Timinas y se guardó en una variable.

```

In [11]: t_seq1=seq_1.count('T')
         print(t_seq1)

```

3469

Obtuvimos el porcentaje Guaninas/Citocinas

```
In [4]: GC(seq_record.seq)
```

```
Out[4]: 51.6719288174512
```

Se obtuvo el complemento de la secuencia de ADN

```
In [11]: Seq1_C=seq_record.seq.complement()
          print(Seq1_C)
```

```
CGTCTTGTCAGGAGGGACATTCTCGGATTGGTAACGGTCCCTTTGGACGGGACCCGCGAGGGAAGTAATCGTCATAA
AAAAAATTTAATTAGACTAATTATTAATAAAAAGGGGGTAAATTAAGGAGGGTCCACCTCAACGGCTTCGA
CCCCCGTCGACCCCTCCACCCCTACCCCTCCCTCTCTGTCTTCAACTCCCGTAGAGAGAGAAGGAAGGGCTGGGAG
ACCGGGGGTTCCCCGTCCTCCTTACGTCTCTCGTCTCAACTCGAACCCCTCGACGTCTACGGAGGCGGGGAGGAGAGA
GGGTCCGAGAAGGAGGACGGGGGAAGAAGCTTGAGAGGAATTAACAAACCGAAAACCTACTAATATTAATAAAAA
TAAAACTTAAATATATTTTATATACACACACACACCTCGACTCTGTCCGAGCCGTCGCCGTGTCTTACTCCCTT
CTGCTCTTTCTCTCACCCCTCTCTCTCTCCGTCTCTCCCTCTCTCCCTCTCACTGTCTCGTCGCGAGCCATTACAAAGG
AATAACCAAGTTTATACATTGATTACCAGGCACCGCCAAAGTTCTTATCCTGTCTGTCTCGCGAAGAGAACCAATCCCT
CTTCTCCCCCGACCGCTCTCCACCCGCGCATCAATTAATACCCCTCCTTTTAACCGCCCCCTCTCTTTGCCCAA
TTTGACCTTATGGCCATTGACCATCACCCCGCTAACACAACAGGACTACAGTGACACACGGTTCAACCCCCACCCCT
CCGTTCCCTTGTCGTGGACACCCCGAGTTCGGCCGCCAGATCCCGCGGTGAGAGACGACGGGTGACACAGTCATGG
ACGAATAATATCTAGAAGCCAGAGAACCGATTACCTTGACCCAGCCTCGCCTTGACCTTCGAGTTCGAGGTCCGTCC
TCTCTCTTGAATCTCCGACACGGGTAACGGGGGATTTCGTTTCAATCGGGGGGTTGGGGGTTTGTGTTGAAAGGTCGAA
GGGAGGGACAGACCGTCGGGGGAGGGAGGGTCTCTCGCGCTGACGATCTCGAGTGTGTACGCGTCACACCCGGGTC
CCGGCCCCGGCGCTCGTCTTCGGCCGCGTCGATCCGCCGGCCGCCCGGACAATTAACCGTTAATCCCCCTCCGA
CCACCGACCACGCGCAGTCGGCTCTCCTCTCGCAGACGGGTGGGGGACGAGGGCGGGGGTGAGCCCGCTACCTTCC
CACCTCCACGGGACGCAACCCACCTCCACCTCCAACATCCACCCCCACCCCTACGACATGAGTTTTTCGGTAGA
ACACGAGTCTCTTTTCTCCGGATGGCCGAAAGGGAAGGCCCCAGGCCGCGGGGAGTGGGGGTGCGGCGCCGGTAGGGT
CGGCCCTACGGGTGACCTGGCCCTACGGGCGAGCGGTGCGTACCGACGAGACCCGATCCTGGACGGAGCGGAGCCCC
CCTCCACCTTTACCCCTTCCGGGATCCGACCCGCCCTGTCCGGGTGAGGAGGACCATCTCAGAGGATCCACGACAC
CTCCGTCACTCCTAAGGCCCCCGACGACCGAATACCCCATCTCCGAGTCGGACCCCGCCGCCCTCCAGGACCCGT
GAGCCCTAGAGAGATTATTAACCACGATTACCGGACAACATATCAGAATTGGGCTGCGTATACCAAAGTATTGTGTC
ACCGAATTAAGAAGGTTTACATGCACCGGACTTACACAACGTCAAACCTATATACTGGGGCGGGACGAGGGGCCGG
ACCCAGGCTCCGAGGCTCCGCCGGGGTATTACACTTAAAGACAAATATTTGCCGTGCGTGTGCGTTTTCGGATGT
GTGGGTCCGTCTGTGTGGGCCCTCCGACCCCTTGTATAAGGTCTGCTCCAGGCTCTCTTGACCCCGGCTGTAAT
GGAATTAGATTTCGGTTCTGTGGGGGAGATGGGGGTGGGAGAGGGTCTCCGCGCTGACCTAAGACGCGGTTTCAGGGT
CACCCCTCCCTAGCGAGGTGCTAAAGGACAAATTTGGGACTGGGCAGCGGACACCCCCCGCCCTACCCCTTTC
TTTGGACAAAATAGTTTAAAAATTTTATGTCAAACGTAAGTCTAGTTTGGCAAACCCGGTCTCGACTGCCACTTA
AAATAGAGGGGAGGGGGGTGTTGGGGGAAGGGAAGGAAAAGACGGGGGCGATTTAAGGGGAAGCGTCGGTAGTATT
TCCGGAGACTCCATAAGAGGGCGCCCTTCTTTACCGTAAAAGAGAGGGGGAGGGGGTGGGGGGTGGGGGTGATCCT
GAACACACAGCCGTCTCCGACAGTGTCTCCCGATTAAAGCGAAAGGAACAAATGCTAGCTCTTTTCGCGAACCGAG
AGGGGTTTACCCGGGTGCGGCGGCGGACGAGTCCGACGGCGCGGAGAGGGCGGCGAGTCCGTGCCGGGTACCGAC
TACGGGGCGCCTAGGTGCGACGCCGTCCGGCCTTGGGCCTCAGACACCCCTGGACTTTTCCCGTTTCCCTCACGCCAC
CCTGTGTTCTCCCCGGGCTCTCCGGGCTTTTTCCCTCGGAAAGAGGACCTCATTACACCTTAGCGGCTGGCCACTC
CGGAGATAAAGAGAGAAAGCGGACGTAAAACAGGCGGGGGCGGGGACACCTACCGGCGACTACCCCCCTTGTGTTTC
ACAGGAAACAGGCAGGGGCTTTCTTTTGGAAAAGGTCTCTTCCGGTAAGTCCGGTCACCTCTAAGTGGATCTTATAA
TCTTCCCGGGGGAGGGTGGGCTCCCCGCCCCGACCCGCTCCGACTCCATAAAGAGAATTAGGATCTTGCTGGAGTGA
CTCACATGGGTGTCTGCTCTAAACCTCCACGATCTCAAGTCCAACCGGCGAATTGTCGTGGTAGGACTCCACAGAAA
TATATTATAAAATTAATAAAAGACAGGGAAGGACCGATGTCCGGACGTGAGGGGGTGGGGGTAGGATACGGGAAG
TGGGGATCTGGGTAGGAATGTTCTGTCTCCGTTTCGTCTCTTCCGTATCGTCGTCTGTCGCGGAGACAAAAGTAAA
AGGTGAGACCGGTCCCTATTTGACCTGGGGTACCTGAGGTTGTGGTTGAGAGACGAAGGGTCAATGAAAGCTAAA
AGAAAGGGAAGAGGAAGAAAGAGAGGGAGAGACCGTAAGTTTGTCTTCTGTTCTCCTGGGTGGTCACCGTTGAGAT
TCGGAGGAACGTCCGGTGTAAAGGACGACTGCACCAGGAAGGACGTCTCTCGGGAAGACACTAATAGATTATTATC
```

GGTCCACCGACTCTTATCTTAGACCGAGTTCCTGAGACGTGTGGGCCCCGTCTCTGTGACTTAGACACCCAGTCCC
GGACCGTTGTAAGGAACCTCGGAGGGGGAGACTCGGGTGAGAAACCCGTGGGAATCCGAACGGACGGGGGGAATATT
CCCCGAAACAGTCGGGACGGTAGGCGACATAAAGTACACGACCAAGGATCGATTCTCTTAACCCGTCTCGTTTGGG
AAAGAGTGGGACCAGACTGTCAACACTAGGGCTGTCTTACGCTGGAACAAACCTCTTCCGCTGGAACAAACCAACCG
ACACAAGACTTTGGTCCGAACCTCACAATATATAAATAATATATGTCCGAATTTATATGTCTGAGTCCGAATACGT
ACTTATATATGTCTGAGTATAAAGGAACCACCAATTAACATATAATGGTACGGATATAAGCGAGTTTGTGGTAAAAT
ACAAGTACGTTTTAAACAATCAGGAGTGATCTACTCTCACCCGAGTACCTCCGTGGCCACGTGACTCGTCCGGGA
GACACGTACGAGACGACCTTTATACACGAGTGACCTTAGCCTCATACGCCCAAGGTCATTGGTTACCTACGGTATTT
CCGAAAGAGTCTGGAATAACGAAAACAACGGGTTTCGTGACCTGACCCCTTTATGATTTAGGGAAGACGAAAGAGAG
GTGTGGAGGACCGTTTGAGAGGGTTGAATCTTTAATAAAATGTTCCCTCTGGTCAACAGAACTTTAACTCAAGACGC
CCTAAGAAAAGTGCCCGTACCACCCCGACGTTTCGTCAAGCCTCATCAGTCCCGGGACCGGACCTTCGGAGTGACCT
CGCTACTTCCCTAGTCTCACCAATTATTCCACCGGAGACCTGTCACTAGACCCCGTCGACGACCGGTCCACGAACC
TCTTCGAACACCTTTTCCGTTTCGGTTTTTCCCATCCCTACGTCACTTATAGACAAAGAGGGGGACGAACGGGGTCGG
CCGGCGAGTGGGGCGGGGTTCGGGCTAAACGGACGTAGGCACCGCCGTCTAACCCCTCAAGAGATTCCAGGGGGTTTAT
GGGTCTAGACCAATGACCTACTTGACGCAAAATTAGGAGCAGCGATAGACTCGGTCTGTCTGGACCAGTCCCTCT
GCCGACCAGGCCCGTCGTGCCCTCTCCTGCGACCTGGGCGACGGACACCGGACCTCGTGACTCATACCCGATGGTTC
CGACTCTTAACACCCGGAGGAAAGGGTCGGA AAAAGGGAAGGTCTCTGTGTTTTCTGTCTCGTCTCCCTTTTATA
GTTACAGTCACCTTTTGGGGTATTTTCTTTCGTCCCTGGTCTCGCCCCGACCTCCCCAAAAGGCGATGAGAGAGATC
CACTAGCCCCAAGGAGGGTCTGAGAACAGGTCTGTCCGACGTGGTTCTCCACACGTGAAAGAAGGTGGAAGGAGGG
CGGGAAGTTCCTCCCGTCGGAATCCGAAAGACCCGGAACCAAGTGGGGTCCGTGACCCCACTGTAATGAGGTCTAAC
ACGGGGAGCGGGTTCGGGTGCCGGAGACGGTGCCGCGGGCGTCGGCGAGGACAAATAACACAGACACTCAACAAATGAA
CAAAACAGACGACGGTGGGACGCAGTCTTTGGGCGGTTTGGTTCGCTTGAGCGGGTCTCTTGTTCACGACACGATA
GGGGCCTGAGGGAGGGGAAAGGTTCGTTGGGGGAGGACGGGAGTGTTTCGAGGACGTAAGAGGGAAAGCCTCCCTCCT
TCGTCTTTATCAGAATTTTGTCCCGTCGACAAATATCCCAGTCCCTTAATACCCCTGTGGTATTTGGAGACCCACA
GGGACACCACCTGTCTGAAGACTCGTCGGGAAGGGAACGGATCCGAAAAGAAGGAAGGGGACCACTGGTGAACCAAT
TCCCCCGGTGTGAGTAACCTCTTCAACTCGGGTGGGGGTAGTTGGGTGAGGGGGTGATGGGGTGAGGAAAGACGTCC
TCAAAGGGACTCTGTGTCCAACAACCCAAAATTTAGTCCCATTACAACCACCTAAGTATTTTCATCTATATCGTACA
GTAAGAAGAGGGTGTGAGGGAGAGGAGTCCGGGGGTTTAAACGTCACCTTAATTCGTTTAGGATGTTTTAATTTAAC
GGGAAAACGATTTCTAGGATATACAGAGGTAGGACGGTCCACATCTCAGTCTATCAACCCGAAAACCGGAAATTA
ACGACGATCGATAATTTAAATTTTCTCATAGTAACCTTCATTCTATTATAAGAGTAAAGACTTTAATTTAAAGGGGAGG
GTCAGAGAGAGAGAGAGAAAAACGGTCGTGTCTAATTTACCAAGAAAGTCTAGACGTTTGTATTTCCTATTTTAAT
AAATATCCCTTTACTCCTACGACAAAATAACAACCGTATAACAAGGTCTTGTATTGTTGTAAAATAAAAGTTTCGTCC
TCTCCAATCTGGTCCCCCTCTTTCTCCCCCTTTCTTTTTTGTACTTTCTCCTCCCGTCCTGGGTGTGTGTTTTGAGG
ACGGTACATATTGACCATCCTGTAAATGTTAAAGATATACTGTTGAGTGACACTACAGTATTATGGCTCAAGAACCTT
AAAATATTAAGAGTACTTGGAGAGTTTACCCAGGTTTCGTACGAAGTACACCTCCCCATATCTTCTTCGGTCTCCG
ACCCCTAGGAACTGGGTACGTCTCGTCGTGTGGGGGATAACGCCGAAGAAGGAGTACACTGTAGGAGAACTCCGG
GATAAGACCCCCGACTTTTATAAGACCTATAAGACGTA AAATTA AAAAGAGGACCCGGTTTCGTCCCGGTGAACGAAA
CGGGGGGAGTTCACAGGGATCCGTAGATTGAGAGAGGGGTAAGGGTCTAGTGTCTCTCGACACAGGGGTCCGAGAG
GGGTTTCGGGGGAGGAGTCGATTATGAGTCGACGAACCTTATCTCCACCCCTCCGAACCCGTGGTTCGGGGGTTTTTCGG
TCTTTACTAAAAAAGGTCTACCTCGACCTTCCCCCGAGTCCCGACTCGCACTACCCCATCTCTACAAACCTGGATA
AAATGCAAAATTGGAATTCAGTCTCTTCACTTAACTCTTCGGTTTGTTTACGGTTACACTCTCAAACATAGACGACT
ACCATCGTCGTTATTCACTTCACGACCATTCCTTTCCAGTCCGTCTTTAGACCTTTCTCTCTCCCTCCTCTCCTTTA
CACTGTACTCAATCGTCTCGGGAAGGTATATGGAAGAAGAGGCAAAATTAACATTATTTTCGTTTTTCGTGAGCCCAT
CGGAGAGTTGGAGAATCGAGTTTTCTCCTTAGTGGTCGCGTCCGGAACCTCCACAAACCCCGTCAAGAGTCGTGCA
ATTGTACGGGTTAGACGGGAGAAGGAGGAAAGGGGGAGGGACGACTAGGGGTGAGGGGAAGTTGACCAGTCGGGCCC
GGACAGGGGTTTTATAATCGTGGGGGTTTACTATAGGTAAAGACACAATCCACACAGTTTAAAGTGAAGAGAGGTA
CGGATTTGAAACTTTCTCCTCACATGGAAAAACCTTACGCCACACACATCGCCCCCGTCCATACACGAGTCTGTGCG
AAGGGGGTAGTATAGATAACCTTTGACCTTCGGGGAATCCACTCGGTACGAAACCACAAGGGTTTTGTGTAAGGGAC
GTAAGTCCAAGAGACCGTCTGGTCTCCGAGAAACGCTAGACTAGGAAACTGATCCTCAAGGTACCTCTAGTTTCCT
CGAATAAATATGTGTTTTGTGGTCCGTAAACCTCATCTTAAATAAAAAATTAAGATTTATTTTTTCAGATTAGTAT
CTGGTTAGCCCGGTTGACTCGTCAGCGAGGTCTTGTGTTTTCGGAATTGGTTTTATCACACTAATACTCCCTCGACT
CGTTAATGACTTTGAGCCGAAAGAGGCATTCTCCCTAAAATCGGGAGCGGGTACCGTGTTTTTCCCGGTGCCCCG
TTGGCAAGAAGAGAACGAGTCTTTTCACTCTGAACCCCCACCACCCCGTCCCAGCCCGTCGACGCCAGCGCAGGTC
CGTCGACTCGTGGGTCCGGTCCGCGACTCCGACTCGTGACTCAGGACTCCGAGGTCACGAACCCAGGGGCTCGTCGG
AGTCTCCGTGGCTCGACATCAACGTGCGCCGTGAGTCCTTACCTCATACACATAGACACACGCTCCCAACCACTC
AATGTTTGTAGGCGCGTCTCAATCCCTAGACCCCTCTCTCTCTCTGTCCCTCTTGTGTTGATTCTCTGGAGAC
CTTTTATAACAGGGGTTTATTTTTTAAGGTCTAGAGTAGGGAGAGTGAAGTTTATAGGGAGGGGTCTTTAATTATAT

AAAGGGTCTCGACTCACCTTTTCGTTTCCGGTTCGTTCCACTCCGGACGGGTGCTTGAAGAAGTCCGACCCAGGTCAGG
GTGGGTGGGAAAGGGGTCAAGTCTCGAGTGTACGGGGGTTCGGAAAGGGACCGGATTTCCCTCATTCCGACCTCACC
ATCGTAAATTGAACGGATCCCTTTAGGGGTTCACGGGTCCGACGTTACGGTGGGAATAGGAACCCCTGAGGTCCGGT
CACTGACGTGGGGTGAGTCCAGGTTCCGCGTCCGTCATTTAAAGAGAGGATATTAATCGCGATAGTTTTTCGTACACA
AAACGTAACCTCTCCTGACCTCTTCGTTACCGTTAAAAAGGGAACATAAGAAAAAGACACAAAAAACTTAAGACAA
AACTTAGGTGATTAATTTTCGAAGAGACTTTCACTTTTAGACGAGAACTGTGGTTCGGGGAAGGCTTACGGACTAAA
CCCTCCGATTCCCCGGTCTCGTCATCGTTCTCTGGACGTTCCGACTACCCCAACCCCCCATCACAAAGAGACTAAAG
AAATCCGAGCAAAATATCTTTGTTTCTTGTTCGAAGAGAGGGGAGGAAGGGTCTGAACACGTATAAATGTCTCGTGT
TGACGTCAAAATTATTTTGTAGACAAGAACGAAGACGACTACTCAAAGGTATTAACAAAAGTCTGTTTAAATTGTCC
TTTTATATTTATATAAATCTTTTATAGATTACGGGTTCGAAAGGGGATATCTGTTCGGTAAGAAAACGGTTTTTTTATT
TTAATTGTACTTTTACCCTTTTCTTTTATATTAGGTTACATCCCTTCTAGCCCGAGTAGACTGGACGAGAGGGGGGA
ACCTCGCTTAGGAAAGTCGTTTAAAGTCGACGTATTAATTTCTAGATTCTCTTTCTTCTCCCGAAGAAGAAAGGAGT
CCTTACTTTCCATTAACCCCCCCCCCTCCACAACCTTTTAATTAATAACGGTTTCCAGAAATTCGTACCTCCTTAAA
TAAACATAAAGACAAAGGGGAAGAGAGGGGGGAAGGTTCGTAAAGACGGAAGAAAGTGGGTTCGTGACCAAGACAAGTA
ATATTTAGCCGCTCTGGAAGTCAGAGACGGGAGACAAATCCCTGCAGTATTTTGTGAATTGAAAGACCCCGACTCC
GAGTGTAAAGACGGAGAGTCCAGCGGTTGGAGGGACTCGGGTGGCTGAAATCCGGGGTTATCTCTTAGAGGGGTTGAA
ACCCCGGGGAAGGAGAGGTCTGTACCAACTCCTCCCCGAGTTCCTTTGGATCCGTAGGAACGACCCCGACATCAG
GTACGACGGGGTCTTCGGGGAAGAAAGAAGGTTCCAGGAAGGGAGGGGGTGGGTCTGTATTTTCTCTAAAAATTTAT
GTATCTTTTAGTTGTGAGTAACCTTTCACTTTGGAGTAATTCTGTAGGATAGAAGGTAGTAAGTTAAACAACATGTAA
CGTTGTAAATTATAGAATCTCCTTTATAGTGAAGTGTACTAAATAGGGGAGATCGTTAGAGAGAGACTTTAACCTCCC
CCCCTGTCTCAATGAGAGAGAAGGCCGGAGAAGATGAACAATGTGGTGTATGGGAATCCCGTCTTTCTCCGACCCCTT
GAGACTCTCTCCGATGAACACAGGTCCCCACGTGACTGAACAAGGGATCGAGTCAGGGGTCTCATGAGGACGGAACA
ACAGAGAGGGGGTGGAGAACGGGAGGTCCCCCGGACTTCGGAGTCGGTCGACAACCTTTGGAGACGAGTCCGTACGG
TACTTAAGCTCCTGTGACCTAAGATCCCGTTTTTCGTTGAGTGTAAACCACAGATACAAACACATGGACGGTTCAGT
GGGGTTTTTACTCTATTGAGTAGATTTTAAAAAAGAATTAAATTTTTTAGGAGGAGATGGGAAAGGGGAAGTAAAG
GAACCCCTCTTCTTTACACTAGATCTTCTCCTTGGAAATATTAATATCCTCAGAAAGGGGAAGGGAAGGTTTATG
GCTGTTTAAAGACAGGAAGACCTCCTAACCCCGACATAAAAGATTTTGGTTGTGTTTAAACAGACCGTCTTCTGAG
AAGTTGGGGATTGCTCCAAGTACATAATTGATTGTACAACAGAGGTGGTCGAGGACGTGTGACAGCGAAACCCCA
AGCTTTGAAATAAAAAGGGGGATCGCTGTGAGTCCCTTTGGAATTGCAATGTCTTCTACTTATTTGCTCAAAAAA
GAGTCGTATCAGGGCAAATACCGAAATAACGATGGGTAATAATGAGGCGAAACAATGTGTCTTCTCTTTCTAGTAT
TTTAGGCCGCTGTAGGCCCAAGAACAATATCGGTTCGAAAGGGGGGGGGTGGGTGGGTTTTTACCACCCCCCCCCC
CCTATACTCGAAAGAGACGAGGGGCGTGTGAGGGGTGGTGGGAAGAAACAGTTGAGTTCCGTAAAAGTCGGATATAG
TGCTATAGTTTAAATTCAAACCTGATTAGTCTCTTTTGGGGGTAGACGTGTGAAAGGCGACAAAGTCAAGCGGATGA
CGCCGAGGGAGTGTTTCGGAGCCTAAGTCGACGGAAGAGCATTGGCAGAGAAGGGAAAAAATCGGGGTCTACCCCC
TTATTTTACCCGCAACTCCAACGACTCTCCGGGTAGGGGTGGTGCCCACTGACCGTGAACCCACTCTCGAGGAATCT
TGGGGAAGACTTCAGAGGTTTCGGAGTATCGAGTCCAGGTCCCCGCGGAGGTGACCGCCACGGGACTCAGACCTCCC
CCGTCCCCCGAAGACTGGTTCCGTCCCCGAACCTCGGACGTCTTCGATTGACAGGTAGTCTCGGGGTGGTTCGGAAG
AGAAAGACCGACCTAGGATACACTTAATGGGGTGTCCGGGAGTGACCGAGGTTTATTATTTAAGAGTGGTTGTATTT
GTCTCAACTACAGAGATCTTTCTAAGAGTCTCAAAAGCAAGAAGGAGAGAGAGAGAAAGAGAGACAGAAAGGGGGGA
GGCCACGGTGGAGGGATGTAAATGAAGAAAAAGAGGAAAAAGATAACGTAAAAAATATGTAAAGGGAAAAATCCGT
TTTTACGACCTTCTCAAAAGGGTTGTCTCCCCAGGACTCTTCTCCGTCTTCCCTCGGTCTCCGTTTCTCTCGTC
TACGCACCGGGTAAACGGGTGAGGTTCGAGACGGTGACATGGGACTTCAGAACGGGGAAATCTGGATCAACAAGGGGT
GTAGGGATGATGTAAGTTCTGAGAGTCCGTGTGAGATGAATGAGTCCCTTGTCCAAAATACTTTAAAGAAAGATGA
AACTAAAAAAAATTAATAACCACTTTTTTTATGATCAAATCACTACCCTGAGTGAGGTACCTCACAGTTTTGTCC
TAAAGGACGTCTTAAAGATCCCTGGGGGGAGGACACGACCCTCCCTCAAGAAGGGAACCTCTTTAGTCTACATTAA
AATCTTCCCCGTCTATATTTTGTACTTTGGTCTAAGTTTATGTCCCAAGAGATCGGGACCGAGACCGGTGTCTG
TGTCGTTTCAATCGACCCCGACTCTCTAAAGGACCATCATGTGAGAGGAAGGGACTCAGGGTTCCGGGACGAGTCC
CTGTACGTGGGGAGGAAGTGGTGGTTTTGGTCAAGGGAATGTATATATATATATATATATAAAAAATTTTCTTTAG
GTTTCAAGATGAAAGTTTCGTGTGTGAATCAGAGTTGATCCCTTAGTTGTCTTCGTCTTCGCTAAAAAAGGGGGAAG
GACTGTAGACCGAACGCTAACCGACCTCCCCAGTCGACTGAAACAGTAAACAGACAGGACCTAACCTCGGCAGG
GATATTGGTAGATCAAGGCTCATGTTTGACCTCTGTCTTTATTTATAATTTCTTTAGTATCTGGCTGGTCCATTTCC
GTTTCCCTACTTAAGGATGAAGTGATTGGGAAGGAATAGGACGGTGGAGCGCCCCCGGTCTGCAGGAGGGGTTCG
AGCGGGAGTTAAGGTGGCGGATACTAGGTCACTCCGTAAAGAGCTGGATACCTCGCCGGCAACGGGTCTTGGCCTAG
ATGAGCTGAGGGAAAATAAGCGGTGTCTCTTACAGCACAAAGTCAAGGTTCGGCCCCCGGCATACTGATACCTAGATT
AAGGAAAATGGTCTCTTTCTGTACGAGAGTTTGACGTCTGTTTTGTGGAATCCTGTATTGTGTGTCTGGAGTTAGC
GAGTCTTAAATCAAGACTCGTCCCGTCTGACGCGGGGTCTGGTCTTTCGGTCATAGGTCTAAATGGGGACCTAC
GTCGCTTACTTAAGCGTGTCAACCACTCAAAATGTGCGAGGCTCTATATTTAATTTATTTGACTTGACCGAAATACTGG

```

CCGAAGGGATCTTCTTGCCCGCCTCTCTCAAAAAATACCGGGGTATTTTTGTTAGTGCAGCAGAGGAGTGGCTCTG
TCCCGGGGGGAGAAGACGGGGGAGTTCGAGCGGAGAGCGAGCGAACGTCCCCGGACCGGGGGAACCGGGGACTGCCTAA
CCTCCCGGGGTTCGCGCCGACTCCCCCGTCTTGTCCCGTCCCCTACTCCCCTCCCTCGTCTTCTATCCCCCTCGGGG
GGTTTTCACCTCAGTCAGACTCCTCTTCCCTCCCCCTCTCCTCTCCGTCTATTACCCCTCTTCAAAGGTTTCTTGAT
CCTACGAACCCCTCTACCCCTCGGGTCAGGAACCTTTTCCCCCGGGAGTAGGGGTAATAGACCCTAAAAAGTTTTT
AGACGAAAATCTGAACGTCGAAAGACAAACGGAACCGGGGCGGGGGGGCGGGGGGGGGTGGTGTGTGGAAAGAA
ACCCTGTGTGTCCAGGGTGAATATCCTCCAGACTTCAACCCAGGAGCGACATGGGGGTACACCCACCCCGTTCC
GTTTTCCCTCTTCCGATCTCGGGATTCTCCGACACTCGACGACCGGTCCAACCCCTCCAGTCTGAAACGATCCGC
TTTGTGACAGCTCTCCTGCGAAACGACTAACGAAAATCACACAAAACACGGGCCTAGAAATCCCCAGCCGATGCCTC
GCCTGGCCTCCGCGCCGGGTCTAGATGAGCGCCATGGTCTGGGACCTTGACCTCTTCTTAAAGTGAAGTTAGCG
ATGGATTGCGCCGCGGCGGTAGCTCTAGCGGTTGCGCGAAACGGACTGGCTCGCTGTCTAGTTTTAGACCAAGGT
CTTGGCGGCCTACTTCACCTTTTTTCTTAGATTAGAGTGTAGGTGAGAGAGCCCCCGCGCCTCCCCGGTGGCGGC
TGTCGGACCCGCTTTTCTCCTTTTCGCCCTTCTGTCTTCTCCTCTTCTCGTCTTTCTCACTGGTCTTGACAGGGAC
GGTGGGGAGAGAGGGAAAGAGGGAGCGAGGGGTGGTTGAGAGGGGATTAGTGTGTGAGACATAAATAGTGACCGTGT
TAATAACACAAAATAAGGGATTTTGTTTAATCCCTCAGTTTGCACCTGGACTTTCAGTCGAGACCTGGGGGAGGG
AGTGGCGTGTGAGAGAAAGTGGTGC GCGGAGGAGGAGGAGCGAGGGAACGATCGAGCAAGAGCCGAACAGATGTCC
GGGAAAAGGGGCAGGTCCGGAACCCCGAGCCTGGGACTTGAGTCTGAGATGTCTAACGGGAGGTTCACTCCTGAAC
CGAGGGGGTGAGGAAGCTGCGGGGGTGGGGGCGGGGGGCACGTCTCTCGGCCGAGGACCCGGACGACCCCGGAGACG
AGGTCCCGGAGTCCCGGGCCGGACCGTCCGCCCTCCCGGCTCCGGGTTCCTCCCGCGCGGAACCGGGGTGTGGTT
GGGGGTCCCGGAGGGGCGTCAGGGACGGATCGGGGAGACGGGGTCGTTTACGGGTCCGGTTCGTTAACATAAATTT
CTTAGGACCCCCAGTAATACCGTAAATGTTTGACACTGGCAAAGACACACTTCTAAAAATCGACATAAACACCAGA
GACATAAATATAAATACAAATCGTGGCAGTCACAAGGATAGGTTAAAGTTTTTTTCTTTTTTTTCTCCCTTTTAATG
TTTTTCTCTTTTTTTTCACTTACTGCAAACAAATCGGTCATCCTCTTTTATTTATTTATTTATTTAGGGAAGCACA
ATGGGAGGACATATTTAGGTTGGAGACCCAGGCAAGAGCTTATAAATTATTTTGACTATAATAAAATTTTGAAAT

```

Se transcribió la secuencia del gen HOXC6 para lo cual primero, se creó seqobjetos para las hebras de ADN de codificación y plantilla.

```

In [21]: trans_seq=seq_record.seq.transcribe()
         print("\n")
         print(trans_seq)

```

```

GCAGAACAGUCCUCCUGUAAGAGCCUAACCAUUGCCAGGGAAACCUGCCCUGGGCGUCCCUUAUUAAGCAGUAUU
UUUUUUAAAUAUAUCUGAUUAAUAUAUUUUUCCCCAUUUAAUUUUUUUCCUCCCAGGUGGAGUUGCCGAAGCU
GGGGGCAGCUGGGGAGGGUGGGGAUGGGAGGGGAGAGACAGAAGUUGAGGGCAUCUCUCUCCUCCCCGACCCUC
UGGCCCCCAAGGGGCAGGAGGAUAGCAGGAGCAGGAGUUGAGCUUGGGAGCUGCAGAUGCCUCCGCCCCUCCUCUCU
CCCAGGCUCUCCUCCUGCCCCCUUCUUGCAACUCUCCUUAUUUUUGUUUGGCUUUUGGAUGAUUAUAUUUUUU
AUUUUUGAAUUUAUAUAAGUAUAUGUGUGUGUGUGUGGAGCUGAGACAGGCUCGGCAGCGGCACAGAAUGAGGGAA
GACGAGAAAGAGAGUGGGAGAGAGAGAGGCAGAGAGGGAGAGAGGGAGAGUGACAGCAGCGCUCGGUAAGUGUUCC
UUAUUGGUUCAAAUAUGUAACUAAUGGUCCGUGGCGGUUUAAGAAUAGGACAGCAGCGGCUUCUCUUGGUUAGGGA
GAAGAGGGGGCUGGCGAGAGGGUGGGCGCGUAGUUAUAUUGGGGAGGAAAAUUGGCGGGGGGAGAGAAACGGGGUU
AAACUGGAAUACCGGUAACUGGUAGUGGGGCGAUUGUGUUGUCCUGAUGUCACUGUGGCAAGUUGGGGGUGGGGA
GGCAAGGGAACAGCACCUGUGGGGCUAAGCCGGCGGUCUAGGGCGCCAGCUCUCUGCUGCCCACUGUGUCAGUACC
UGCUUAUUUAUGAUUCUCCGUCUCUUGGCUAAUGGAACUGGGUCGGAGCGGAACUGGAAGCUCAAGCUCCAGGCAGG
AGAGAGAAAUAGAGGCUGUGCCCAUUGCCCCCUAAGCAAAGUUAGCCCCCAACCCCAAAACAAAUUCCAGCUU
CCCCUCCUGUCUGGCAGCCCCCUCCUCCAGAGAGCGGACUGCUAGAGCUCACACAUGCGCAGUGUGGGCCCAG
GGCCGGGCGCGGAGCAGGAAGCCGGCGCAGCUAGGCGGCGCGGGGCCUGUUAUUGGCAAUUAGGGGGGAGGCU
GGUGGCUGGUGCGGUCAGCCGAGAGGAGAGCGUCUGCCACCCCCUGCUCCCGCCCCCACUCGGGCGGAUGGAAGG

```

GUGGGAGGUGCCCUGCGUUGGGUGGAGGGUGGAGGUUGUAGGGUGGGGGUGGGGGAUGCUGUACUAAAAGCCAUCU
 UGUGCUCAGAGAAAAGAGGCCUACCGGCUUUCUUCCGGGGUCCGGCGCCCCUACCCCCAGCCGCGGCCAUCCCA
 GCCGGGAUGCCCACUGGACCGGAUGCCCGCUCGCCACGCAUGGCUGCUCUGGGCUAGGACCUGCCUCGCCUCGGGG
 GGAGGUGGAAAUGGGGAAGGCCCUAGGCUGGGCGGGGACAGGCCCAGUCCUCCUGGUAGAGUCUCCUAGGUGCUGUG
 GAGGCAGUGAGGAUUCGGGGGCGUCUGGCUUAUGGGGUAGAGGCGUCAGCCUGGGGCGGCGGGAGGGUCCUGGGCA
 CUCGGGAUCUCUCUAAUAAUUGGUGCUAAUGGCCUGUUGUAUAGUCUUAACCCGACGCAUAUGGUUUAUAACACAG
 UGGCUUAAUUCUUCCAAUGUACGUGGCCUGAAUGUGUUGCAGUUUGAUUAUAGACCCCCGCCUGCUCUCCCGGCC
 UGGGUCCGAGGCUCGAGGCGGCCCAUAAGUGUGAAUUCUGUUUAUAAACGGCGACGCACACGCAAACGCCUACA
 CACCCAGGCAGACACACCCGGAGGCGUGGGGGAACAUAUUCAGACGAGGUCCGAGAGAACCUGGGGGCCAGACUUA
 CCUUAUUCUAAGCCAAGACACCCCUUACCCCAACCCUCUCCAGAGGCGCGACUGGGAUUCUGCGCCAAGUCCCA
 GUGGGGGAGGGGAUCGCUCCACGAUUUCCUGUUUAAACCCUGACCCGUCGCCUGUGGGGGGGCGGGGAUGGGGGAAG
 AAACCUGUUUUUAUCAAUUUUUAAAAUACAGUUUGCAUUCAGAUCAAACCGUUUGGGCCAGAGCUGACGGUGAAU
 UUUUUCUCCCCUCCCCCACAACCCCUUCCCUUUCUUGCCCCCGUAAAUCCCCCUUCGCAGCCAUAUAA
 AGGCCUCUGAGGUUUCUCCCGCGGGAAGAAAUGGCAUUUUCUCCCCUCCCCCACCACCCCAAGUAGGA
 CUUGUGUGUCGGCAGAGGCGUCUGACGAGGGGCUAAUUCGCUUCCUUGUUUACGAUCGAGAAAAGCGCUUGGCUC
 UCCCCAAUUGGGCCCAGCCCGCCGCCUGCUCAGGCUGCCGCGCGCUCUCCCGCCGCUAGGCACGGCCCAGUGGCUG
 AUGCCCCGCGGAUCCAGCCUGCGGCAGGCCGGAACCCGGAGUCUGUGGGACCUGAAAAGGGCAAAGGGAGUGCGGUG
 GGACACAAGAGGGGGCCGAGAGGCCGAAAAAGGGAGCCUUUCUCCUGGAGUAAUGUGGGAAUCGCCGACCGGUGAG
 GCCUCUAUUUCUCUCUUUGCCUGCAUUUUGUCCGCCCCCGCCCCUGUGGAUGGCCGUGAUGGGGGGAACACAAG
 UGUCCUUUGUCCGUCCCCGAAAGAAAACCUUUUCCAGAGAAGGCCAUUCAGGCCAGUGGAGAUUACCUAGAAUAAU
 AGAAGGGCCCCUCCACCCGAGGGGCGGGCCUGGGCGAGGCUGAGGUUUCUUAUCCUAGAACGACCUCACU
 GAGUGUACCCACAGCGAGAUUUGGAGGUGCUAGAGUUCGAGGUUGGCCGCUAACAGCACCAUCCUGAGGUGUCUUU
 AUAUAUAUUAUUAUUGAUUUUCUGUCCCUUCCUGGCUCACGCCUUCGAGCUCUCCCCACCCCCAUCCUAUGCCCUUC
 ACCCUUAGACCCAUCCUUAACAAGACAGAGGCAAGCAGAGAAGGCAUAGCAGCAGCGACCGGCGCUCUGUUUUAUUU
 UCCACUCUGGCCAGGGGAUAAACUGGACCCAGUGGACUCCAACACCAACUCUCUGCUUCCAGUUACUUUCGAUUU
 UCUUCCCUUCUCCUUCUUCUCUCCCUUCUCUGGCAUUCAAACAGAAGCACAAGAGGACCCACCAGUGGCAACUCUA
 AGCCUCCUUGCAGGCCACAUUCCUGCUGACGUGGUCCUUCUGCAGGAGAGCCCUUCUGUGAUUAUCUAAUAAUAG
 CCAGGGUGGCUGAGAAUAGAAUCUGGCUCAAGGACUCUGCACACCCGGGGCAGAGACACUGAAUCUGUGGGUCAGGG
 CCUGGCAACAUUCCUUGGAGCCUCCCCCUCUGAGCCCACUCUUGGGCACCCUAGGCUUGCCUGCCCCCUUAUAA
 GGGGCUUUGUCAGCCCUGCCAUCCGUGUAUUUCAUGUGCUGGUUCCUAGCUAAGGAGAAUUGGGCAGAGCAAACCC
 UUUUCACCCUGGUCUGACAGUUGUGAUCCCGACAGAAUGCGACCUUGUUUGGAGAAGGCGACCUUGUUGGGUUGGC
 UGUGUUCUGAAACCAGGCUUGGGAGUGUUAUAUAUUUAUUAUAUACAGGCUUAAAUAUACAGACUCAGGCUUAUGCA
 UGAAUAUAUACAGACUCAUAUUUCCUUGGUGGUAAUUGUAUAUUAUACCAUGCCUAUAUUCGCUAAACACCAUUUUA
 UGUUCAUGCAAAAUUUUGUAGUCCUACUAGAUGAGAGUGGGCUAUGGAGGCACCGGGUGCACUGAGCAGGCCCCU
 CUGUGCAUGCUCUGCUGGAAAUAUGUGCUCACUGGAAUCGGAGUAUGCGGGUCCAGUAACCAAUGGAUGCCAUA
 GGCUUUCUCAGACCUUAUUGCUUUUGUUGCCCAAGCAGCUGGACUGGGGAAUACUAAAUCCCUUCUGCUUUCUCUC
 CACACCUCCUGGCAAACUCUCCCAACUUAAGAAUUAUUUAACAAGGGAGACCAGUUGUCUUGAAAUUGAGUUCUGCG
 GGAUUCUUUACGGGCAUGGUGGGGCGCAAGCAGUUCGGAGUAGUCAGGGCCCUGGCGUGGAAGCCUCACUGGA
 GCGAUGAAGGGAUCAGAGUGGUUAAUAAGGUGGCCUCUGGACAGUGAUCUGGGGACGUCUGGCCAGGGUGCUUGG
 AGAAGCUUGUGGAAAAGGCAAGCCAAAAGGGUAGGGAUGCAGUGAAUAUCUGUUUCUCCCCUGCUUGCCCCAGCC
 GGCCGCUACCCCCGCCCCAGCCCGAUUUGCCUGCAUCCGUGGCGGCAGAUUGGGAGUUCUCUAAAGGUCCCCCAAUA
 CCCAGAUCCUGGUUACUGGAUGAACUGCGUUUUAUCCUCGUCGCUAUCUGAGCCAGGACAAGACCUGGUCAGGGAGA
 CGGCGUGGUCGGGCGACGCGGAGAGGACGUGGACCCGUGCCUGUGGCCUGGAGCACUGAGUAUGGGCUACCAAG
 GCUGAGAAUUGAUGGGCCUCCUUUCCAGCCUUUUUCCCUUCCAGAGACACAAAAAGACAGAGCAGAGGGGAAUAU
 CAAUGUCAGUGGAAAACCCCAUAAAAGAAAGCAGGGACCAGAGCGGGGCUUGAGGGGUUUUCCGCUACUCUCUCUAG

GUGAUCGGGUUCCUCCCAAGACUCUUGUCCAGACAGGCUGCACCAAGAGGUGUGCAGCUUUCUUCCACCUUCCUCCC
 GCCCUUCAAGGAGGGCAGCCUUAAGGCUUUCUGGGCCUUGGUCACCCCAGGCACUGGGGUGACAUUACUCCAGGAUUG
 UGCCCCUCGCCCAGCCCACGGCCUCUGCCACGGCGCCGCAGCCGCUCUCGUUUUUAUGUGUCUGUGAGUUGUUUACUU
 GUUUUGUCUGCUGCCACCCUGCGUCAGAAACCCGCCAAACCAGCGAACUCGCCCAGGAGAACAAGUGUCUGUGCUAU
 CCCCAGACUCCCUCCCUUUCAGCAACCCCCUCCUGCCUCACAAGCUCUCCUGCAUUCUCCCUUUCGGAGGGGAGGA
 AGCAGAAAUAGUCUAAAAACAGGGCAGCUGUUUAUAGGGUCAGGGGAAUUAUGGGGACACCAUAAAACCUCUGGGUGU
 CCCUGUGGUGGACAGACUUCUGAGCAGCCCUUCCCUUGCCUAGGCUUUUCUCCUCCCCUGGUGACCACUUGGUUA
 AGGGGGCCACACUCAUUGAGAAAGUUGAGCCCACCCCCAUCAACCCAGUCCCCACUACCCACUCCUUUCUGCAGG
 AGUUUCCUGAGACACAGGUUGUUGGGUUUAAAUCAGGGUAAUGUUGGUGGAUUAUAAAGCUAGAUAUAGCAUGU
 CAUUCUUCUCCACACUCCCUCCUCAGCCCCCAAAUUGCAGUGAAUUAAGCAAUCCUACAAAUAUAAUUG
 CCCCUIUGCUAAAGAUCUUAUUGUCUCCAAUCCUGCCAGGUGUAGAGUCAGAUAGUUGGCUUUUGGCCUUUAUUU
 UGCUGCUAGCUAUUAAAUUUAAAAGAGUAUCAUUGAAGUAAGUAAUUCUCAUUCUGAAAUAAUUUCCCCUCC
 CAGUCUCUCUCUCUUCUUUUGCCAGCACAGAUUAAAUGGUUCUUUCAGAUUCUGAAACAAUAAAAGGAUAAAAUUA
 UUUUAUAGGGAAUAGGAUGCUGUUUAUUGUUGGCAUUAUUGUCCAGAACAUACAACAUUUUAUUUUAAGCAGG
 AGAGGUUAGACCAGGGGGAGAAAGAGGGGGAAAGGAAAAACAUGAAAGGAGGGCAGGAACCCACACACAAAACUCC
 UGCCAUGUAUAACUGGUAGGACAUUUACAAUUCUAUAUGACAACUCACUGUGAUGUCAUAAUACCGAGUUCUUGAA
 UUUUAUAAUUCUCAUGAACCUCUCAAAUGGGUCCAAGCAGUGCUUCAUGUGGAGGGGUUAUAGAAGAAGCCAGGAGGC
 UGGGGAUCCUUUGACCCAGUGCAGAGCAGCACACCCCCUAUUGCGGCUCUUCUCCUCAUGUGACAUCUCCUUGAGGCC
 CUAUUCUGGGGGCUGAAAAUAUUCUGGAUAUUCUGCAUUAUUAUUUUCUCCUGGGCCAAGCAGGGCCACUUGCUIU
 GCCCCCUCAAGUGUCCCUAGGCAUCUAAGCUCUCUCCCCAUUCCAGAUACAGAGAGCUGUGUCCCCAGGCUCUC
 CCCAAGCCCCCUCCUCAGCUAAUACUCAGCUGCUUGGAAUAGAGGUGGGAGGCUUGGGCACCAGCCCCCAAAAGCC
 AGAAAUGAUUUUUUCCAGAUUGGAGCUGGAAGGGGGCUCAGGGCUGAGCGUGAUGGGGUAGAGAUUUUGGGACCUAU
 UUUACGUUUUAACCUUAAGUCAGAGAAGUGAAUUGAGAAGCCAAACAAUUGCCAAUGUGAGAGUUUGUAUCUGCUGA
 UGGUAGCAGCAUAAGUGAAGUCUGGUAAGGAAAGGUCAGGCAGAAUUCUGGAAAGAGAGAGGGAGGAGAGGAAU
 GUGACAUGAGUUAGCAGAGCCCUUCCAUAUACCUUUCUUCUCCGUUUAAUUGUAAUAAAGCAAAAGCAGCUCGGGUA
 GCCUCUCAACCUCUUAAGCUCAAAAGAGGAAUACCCAGCGCAGGCCUUGGAGGUGUUUGGGGGCAGUUCUCAGCAGCU
 UAACAUGCCCAAUCUGCCCCUCUCCUCCUUUCCCCUCCUGCUGAUCCCCAGCUCUCCUUAACUGGUACAGCCGGG
 CCCUGUCCCCCAAAUAUUAAGCACCCCCAAUAGUAUCCAUAUUCUGUGUAGGUGUGUCAAAUUAUACUUCUCUCCAU
 GCCUAAACUUGAAAGGAGGAGUGUACCUUUUUGGAAUGCGGUGUGUGUAGCGGGGGCAGGUUAUGUCUCAGACAGC
 UUUCCCCAUCAUAUCUAUGGGAAACUGGAAGCCCCUAGGUGAGCCAUGCUUUGGUGUUCUCCAAACAGCAUUCUCCUG
 CAUUCAGGUUCUCUGGCAGACCAGGAGGCUCUUGCGAUUCUGAUCCUUGACUAGGAGUCCAUGGAGAUCAAAGGA
 GCUUAUUUAUACACAAAAACACCAGGCAUUGGAGUAGAAUUUAUUUUUAAUUAUUAUAAAAAGUCUAAUUAUA
 GACCAUUCGGGCCAACUGAGCAGUCGCUCAGAACAGCAAAGCCUUAACCAAUAGUGUGAUUAUAGAGGGGAGCUGA
 GCAAUUAUCUGAAACUCGGCUUUCUCCGUAAAGAGGGAUUUUAAGCCUUCGCCCAGUGGCACAAAAGGGGGCAGCGGGC
 AACCGUUCUUCUCUUGCUCAGAAAAGUGAGACUUGGGGGUGGUGGGGGCAGGGUCGGGCAGCUGCGGUCGCGUCCAG
 GCAGCUGAGCACCCAGGCCAGGCGCUGAGGCUGAGCACUGAGUCCUGAGGCUCAGUGCUUGGGUCCCCGAGCAGCC
 UCAGGAGGCACCGAGCUGUAGUUGCAGCGCGGCAGUCAGGAAUGGAGUAUGUGUAUCUGUGUGCGAGGGUUGGUGAG
 UUACAAACAUCCGCGCAGAGUUAAGGGGAUCUGGGGAGAGAGAGGAGACAGGGAGAACAACUAAGGAGGACCUCUG
 GAAAAUAUUGUCCCCAAUAAAAAUUCCAGAUUCUAUCCUUCUACUUAUAAUUAUCCUCCCCAGGAAAUAAUUAUA
 UUUCCAGAGCUGAGUGGAAAGCAAAGGCCAGCAAGGUGAGGCCUGCCCACGAACUUCUUCAGGCUGGGUCCAGUCC
 CACCCACCCUUUCCCCAGUUCAGAGCUCACAUGCCCCAGCCUUUCCUGGCCUAAAGGGGAGUAAGGCUGGAGUGG
 UAGCAUUUAACUUGCCUAGGGAAUCCCCAAGUGCCCAGGCUGCAAUGCCACCCUUAUCCUUGGGGACUCCAGGCCA
 GUGACUGCACCCACUCAGGUCCAAGGCGCAGGCAGUAAAUUUCUCCUAUAAUUAAGCGCUAUCAAAAGCAUGUGU
 UUUGCAUUGGAGAGGACUGGAGAAGCAAGUGGCAAUUUUUCUUGUAUUCUUUUUCUGUGUUUUUGAAUUCUGUU
 UUUGAAUCCACUAAUUAAGCUUCUCUGAAAGUGAAAAUCUGCUCUUGACACCAAGCCCCUCCGAAUGCCUGAUUU

GGGAGGCUAAGGGGCCAGAGCAGUAGCAAGGAGACCUGCAAGCCUGAUGGGGUUGGGGGUAGUGUUCUCUGAUUUC
 UUUAGGCUCGUUUUAUAGAAACAAAGAACAAGUUCUCUCCCCUCCUCCCAAGACUUGUGCAUAUUUACAGAGCACA
 ACUGCAGUUUAAUAAAACAUUCUGUUCUUGCUUCUGCUGAUGAGUUCCAUAUUUGUUUUCAGACAAAUUUAAACAGG
 AAAAUUAUUAAUUAUUUAGAAAAAUCUAAGUCCCGAGCUUCCCCCUAUAGACAGCCAUCUUCUUUGCCAAAAAAUAA
 AAUUAACAUGAAAUGGGAAAAAGAAAAUUAUAAUCCAUGUAGGGAAGAUCGGGCUCAUCUGACCUGCUCUCCCCCU
 UGGAGCGAAUCCUUUCAGCAAAUUUCAGCUGCAUAAUUAAGAUCUAAGAGAAAAGAAGAGGGCUUCUUCUUUCCUCA
 GGAAUGAAAGGUAAUUGGGGGGGGGGAGGUGUUGAAAAUUAUUUUGCCAAAGGUCUUUAAGCAGUGGAGGAAUUU
 AUUUGUAUUUCUGUUUCCCCUUCUCUCCCCCUUCCAGCAUUUCUGCCUUUUUACCCAGCAGCUGGUUUCUGUUCAU
 UAUAAAUCGGCGAGACC UUUCAGUCUCUGCCUCUGUUUAGGGACGCUAUAAAACACUUAACUUUCUGGGGCUGAGG
 CUCACAUUCUGCCUCUCAGGUCGCCAACCUCCUGAGCCACCAGACUUUAGGCCCAAUAGAGAAUCUCCCCAACUU
 UGGGGCCCCUUUCCUCUCCAGACAUGGUUGAGGAGGGGCUCAAGGAAACCUAGGCAUCCUUGCUGGGGGCUGUAGUC
 CAUGCUGCCCCAGAAGCCCCUUCUUUCUCCAAGGUCCUCCCCUCCCCACCCAGACAUAAGAGAUUUUUAAAUA
 CAUAGAAAAUCAACACUCAUUGAAAGUGAAACCUCAUUAAGACAUCUUAUCCAUCAUUCAAUUUGUUGUACAUU
 GCAACAAUUUAUAUCUUGAAGGAAUAUCACUGACAUGAUUUAUCCCUUAGCAAUCUCUCUCUGAAAUUGGAGGG
 GGGGACAGAGUUACUCUCUUCGCGCCUCUUCUACUUGUUAACACCACUACCCUAGGGCAGAAAGGAGGCUGGGAA
 CUCUGAGAGAGGCUACUUGUGUCCAGGGGUGCACUGACUUGUUCUCCUAGCUCAGUCCCCAGAGUACUCCUGCCUUGU
 UGUCUCUCCCCACCUCUUGCCCUCCAGGGGGCCUGAAGCCUCAGCCAGCUGUUGGAAACCUCUGCUCAGGCAUGCC
 AUGAAUUCGAGGACAGCUGGAUUCUAGGGCAAAAAGCAACUCACAUUGGUGUCUAUGUUUGUGUACCUGCCAAGUCA
 CCCCCAAAUGAGAUACUCAUCUAAAAUUUUUUUUCUAAUUUAAAAAUCCUCCUCUACCCUUUCCCCUUCAUUUC
 CUUGGGGAGAAAGAGUGUGAUCUAGAAGGAGGAACCUUUAUAAUUAUAGGAGUCUUUCCCCUCCCCUCCAAUAC
 CGACAAAUUUCUGUCCUUUCUGGAGGAUUGGGGCUGUAUUUUCUAAAACCAACACAAAUUGUCUGGCAGGAAGACUC
 UUCAACCCCCUAACGAGGUUCAUGUAUUAACUAACAUGUUGUCUCCACCAGCUCUCCUGCACACUGUCUGCUUUUGGU
 UCGAAACUUUAUUUUUCCCCCUAGCGACACUCCAGGGAAACCUUAACGUUACAGAAGAUGAAUAAACGAGUUUUU
 CUCAGCAUAGUCCCGUUUAUGGCUUUAUUGCUACCCAUUGAUUACUCCGCUUUGUUAACAGAAAGGAAAAGAUCAUA
 AAAUCCGGCGACAUCGCGGUUCUUGUUAUAGCCAGCUUCCCCCCCCCAACCCACCCAAAAAUGGUGGGGGGGGGG
 GGAUAUGAGCUUUCUCUGCUCUCCCGCACACUCCCCACCACCCUUCUUUGUCAACUCAAGGCAUUUUCAGCCUAUAUC
 ACGAUUAUCAAUUAAGUUUGGACUAAUCAGGAGAAAAACCCCAUCUGCACACUUUCCGCUUUCAGUUCGCCUACU
 GCGGCUCUCCUCAAGCCUCGGAUUCAGCUGCCUUUUCGUAACCGUCUCUCCCCUUUUUUUAGCCCCAAGAUGGGGG
 AAUAAAAUGGGCGUUGAGGUUGCUGAGAGGCCCAUCCCCACCACGGGUGACUGGCACUUGGGUGAGAGCUCUUAAGA
 ACCCUUUCUGAAGUCUCCAAGCCUCAUAGCUCAGGUCCAGGGGCGCCUCCACUGGCGGUGCCCUGAGUCUGGAGGG
 GGCAGGGGGGCUUCUGACCAAGGCAGGGGCUUGAGCCUGCAGGAAGCUAACUGUCCAUCAGAGCCCCACCAGCCUUC
 UCUUUCUGGCUGGAUCCUAUGUGAAUUAACCCACAGGCCCUCACUGGCUCCAAUAAUAAAUUUCACCAACAUAUA
 CAGGAGUUGAUGUCUCUAGAAAGAUUCUAGAGUUUUCGUUCUUCUUCUCUCUCUUCUUCUCUGUCUUUCCCCCU
 CCGGUGCCACCUCUCCUACAUUUUACUUCUUUUUCUCCUUUUUCUUAUGCAUUUUUUAUACAUUUUCCCUUUUAGGCA
 AAAAUGCUGGAAGAGUUUCCCAACAGAGGGGUCCUGAGAAGGAGGCAGAAGGGAGCCAGGAGGCAAAGGAGAGCAG
 AUGCGUGGCCCAUUGGCCACUCCAGCUCUGCCACUGUACCCUGAAGUCUUGCCCCUUUAGACCUAGUUGUCCCCA
 CAUCCCUACUACAUUCAAGACUCUCAGCCACACUCUACUUAACUCAGGGGAACAGGUUUUAUGAAAUUUCUUUCUACU
 UUUGAUUUUUUUUAAUUAUUGGUGAAAAAAUACUAGUUUAGUGAUGGGACUCACUCCAUGGAGUGUCAAACAGG
 AUUUCUGCAGGAUUUCUAGGGACCCCCUCCUGUGCUGGGAGGGAGUUCUCCCCUUGGAGAAAUCAGGAUGUAAUU
 UUAGAAGGGGGCAGAUUAUAAACAUGAAACCAGAUUGAAAAUACAGGGGUUUCUCUAGCCCUGGCUCUGGCCACAGC
 ACAGCAAAGUUAGCUGGGGCGAGAGAUUUCUGGUAGUACAGCUCUCCUCCCCUGAGUCCCAAGGCCUGCUCAGG
 GACAUGCACCCUCCUUCACCACCAACCAGUUCCCUUACAUAUAUAUAUAUAUAUAUAUUUUUUAAAAGAAAUC
 CAAGUCUUAUUCAAAGCACACACUUAAGUCUAAUAGGGAAUCAAACAGAAGCAGAAGCGAUUUUUUCCCCCUUC
 CUGACAUCUGGCUUGCGAUUGGCUGGGAGGGGUCAGCUGACUUUGUCAUUUUGUCUGUCCUGGAUUGGAGCCGUCC
 CUUAACCAUCUAGUUCGAGUACAAACUGGAGACAGAAAUAUUAUAAAGAAAUCAUAGACCGACCAGGUAAAGG

```

CAAAGGGAUGAAUUCUACUUCACUAACCCUUCUUAUCCUGCCACCUCGCCGGGGGCCAGGACGUCCUCCCCAACG
UCGCCCUCAAUUCCACCGCCUAUGAUCCAGUGAGGCAUUCUCGACCUAUGGAGCGGCCGUUGCCCAGAACCGGAUC
UACUCGACUCCCUUUUAUUCGCCACAGGAGAAUGUCGUGUUCAGUUCAGCCGGGGGCCGUUGACUAUGGAUCUAA
UUCUUUUUACCAGGAGAAAGACAUGCUCUCAAAACUGCAGACAAAACACCUUAGGACAUAACACACAGACCUCAAUCG
CUCAGGAUUUUAGUUCUGAGCAGGGCAGGACUGCGCCCCAGGACCAGAAAGCCAGUAUCCAGAUUUUACCCUGGAUG
CAGCGAAUGAAUUCGCACAGUGGUGAGUUUUACAGCUCCGAGAUUAAAAUAAAUAACUGAACUGGCUUUUAGACC
GGCUUCCCUAGAAGAACGGGCGGAGAGAGUUUUUAUGGCCCCAUAAAAACAAUCACGCUCGUCUCCUCACCGAGAC
AGGGCCCCCUCUUCUGCCCCCUCAGCUCGCCUCUCGCUCGCUUGCAGGGGCCUGGGCCCCUUGGGCCCCUGACGGAUU
GGAGGGCCCCAAGGCGGGCUGAGGGGGCAGGAACAGGGCAGGGGAUGAGGGGAGGGAGCAGAAGAUAGGGGAGCCCC
CCAAAGUGGAGUCAGUCUGAGGAGAAGGGAGGGGGAGAGGAGAGGCAGAUAAUGGGGGAGAAGUUUCCAAGGAACUA
GGAUGCUUGGGGAGAUGGGGGAGCCAGUCCUUGAGAAAGGGGGGCCCUAUCCCCAUUAUCUGGGAUUUUUAAAAA
UCUGCUUUUAGACUUGCAGCUUUCUGUUUGCCUUUUGCCCCGCCCCCCCGCCCCCCCCCCCCACCACACACCUUUCU
UGGGACACACAGGUCCCACCUUAUAGGAGGUCUGAAGUUGGGGUCCUCGCUGUACCCCCAGUGGGGGUGGGGCAAGG
CAAAAGGGAGAAGGCUAGAGCCCCUAAGGAGGCUGUGAGCUCUGGCCAGGUUGGGAGGGUCAGGACUUUGCUAGGCG
AAACAGUCUGCAGAGGACGCUUUGCUGAUUGCUUUAGUGUGUUUUGUGCCCGGAUCUUUAGGGGUCGGCUACGGAG
CGGACCGGAGGCGCGGCCCGCCAGAUUCUCGCGGUACCAGACCCUGGAACUGGAGAAGGAAUUUCACUUCAAUCGC
UACCUAACGCGGCGCCGCGCAUCGAGAUCCGAACGCGCUUUGCCUGACCGAGCGACAGAUCAAAAUCUGGUUCCA
GAACCGCCGGAUGAAGUGGAAAAAAGAAUCUAAUCUCACAUCACUCUCGCGGGGGCGGCGGAGGGGCCACCGCCG
ACAGCCUGGGCGGAAAAGAGGAAAAGCGGGAAGAGACAGAAGAGGAGAAGCAGAAAGAGUGACCAGGACUGUCCUG
CCACCCUCUCUCCCUUUCUCCUCGCUCUCCACCAACUCUCCCUAAUCACACACUCUGUAUUUAUCACUGGCACA
AUUGAUGUGUUUGAUUCCCUAAAAACAAAUAAGGGAGUCAAACGUGGACCUGAAAGUCAGCUCUGGACCCCCUCCC
UCACCGCACAAUCUCUUAUCCACGCGCCUCCUCCUCCUCGCUCUCCUUGCUAGCUCGUUCUGGCUUGUCUACAGG
CCCUUUUCCCCGUCAGGCCUUGGGGGCUCGGACCCUGAACUCAGACUCUACAGAUUGCCCUCCAAGUGAGGACUUG
GCUCCCCACUCCUUCGACGCCCCACCCCCGCCCGUGCAGAGAGCCGGCUCUGGGCCUGCUGGGGCCUCUGC
UCCAGGGCCUCAGGGCCCGGCCUGGCAGCCGGGGAGGGCCGGAGGCCCAAGGAGGGCGCGCCUUGGGCCCCACACCAA
CCCCCAGGGCCUCCCCGAGUCCUUGCCUAGCCCCUCUGCCCCAGCAAAUGCCCAGCCCAGGCAAAUUGUAUUUAAA
GAAUCCUGGGGGUCAUUAUGGCAUUUACAAACUGUGACCGUUUCUGUGUGAAGAUUUUAGCUGUAUUUGUGGUCU
CUGUAUUUAUUAUUAUGUUUAGCACCGUCAGUGUCCUAUCCAAUUUCAAAGGAAAAAAGAGGGAAAAUUAUAC
AAAAAGAGAGAAAAAAGUGAAUGACGUUUGUUAGCCAGUAGGAGAAAAUAAAUAUUAAAUAUUUCCCUUCGUGU
UACCCUCCUGUAUAAAUCCAACCUCUGGGUCCGUUCUGGAAUUAUUAAUAAAACUGAUUAUUUUUAAAACUUUA

```

Traducción del ARNm.

Se obtuvo la proteína de nuestra secuencia y la cantidad de codones lo guardamos en una variable.

```

In [24]: seq_tradu=seq_1.translate(stop_symbol='')
          print(seq_tradu)
          lenseq_tradu=len(seq_tradu)
          print(lenseq_tradu)

```

```

AEQSSLEPNHCQGNLPWALPSLAVFFLNSDLFFPHLIFFPPRWSCRSWGQLGRVGMGGERQKLRASLSSFPTLWPPR
GRNAGAGVELGSCRCLRPSSLPGSSSCPLLATLLNFVWLLDD
YNYFYFYIYIKYMCVCVELRQARQRHRMREDEKESGREGRGREGERESDSSARVFPYWFKYVTNGPWRFQEDSSGFSWL
GRRGGWREGGRVVNYGEENWRGERNGVKLEYRLVVGRLLCCPDV
TVCQVGGGEAREQHLLWGSSRRSRAPALCCPLCQYLLIIDLRSLGWNWVGAELEAQAPGRRENLEAVPIAPAKLAPQP

```

PNKLSSFPSSLGSGSPLPPRERDCSSHMRSVGPGRAGSRRSA
 AGGACLAIRGEAGGWCASAERRASAHPLLPPLGRMEGWVPCVGVWRVEVVGWGWGMLYSKAILCSEKRGLPAFPSPG
 VRRPSPAAAIPAGMPTGPGCPLATHGCSGLGPASPRGEVEMG
 KALGWAGTGPVLLVESPRCCGGSEDSGGCWLMGRRQPGAAGGSWALGISLIIGANGLLYSLNPTHMVSHSGLISSKC
 TWPCCSLIYDPALLPGLGPRLRGGPISVNFCLTATHTQTPTH
 PGRHTRRRGGTYSRRGPREPGGQTYLNLSDTPTSTPLSQRRDWDSAPSPSGGGDRSTISCLNPDPSVGGRGWGKK
 PVFIKFLKIQFAFRSNRLGQSRILSPLPPQPPSLPFLPPLNSP
 FAAIKASEVFSRGKKWHFLSPSPTPPPVGLVCRQRLTRGFRFPCLRSRKALGSPQMGPARLLRLPRALPPLRH
 GPVADAPRIQPAAGRNPESVGPEKGKGSAGVGHKRGPRGPKKGA
 FLLECGNRRPVRPLFLSFACILSAPAPVDGRWGGTQVSFVRPRKKTFSREGHSQWRFTNIRRAPSHPRGGPGRGGI
 SLNPRTTSLSVPTARFGGARVRGWPLNSTILRCLYIIFLIIFCP
 FLATALQLPHPHMPFTPRPILTRQQAQEAQQRALCFHFPLWPGDKLDPSGLQHQLSASQLLSIFFPFSFFLSLS
 GIQTEAQEDPPVATLSLLAGHIPCRGPSCRRALLLSNNSQGGG
 NLAQGLCTPGAETLNLWVRAWQHSLEPPPLSPLFGHPACLPPLGALSALPSAVFHVLPSPGELGRANPFSPWSDSCD
 PDRMRPCLEKATLLGWLCSERLGSVIYLLYTGLNIQTQAYAI
 YTDSYFLGGLYITMPIFAQTFFYVHAKFCSSLDESGLMEAPGALSRRPSVHALLEICAHWNRSRMRVPVTNGCHKGFLR
 PYCFCCPSSWTGEILNPFCLSTPPGKLSQLRNYFTRETSCLE
 IEFCEGILFTGMVGLQAVRSSQPGAGSLTGAMKGSEWLIRWPLDSDLGQLLARVLGEACGKGKPKGVGMQISVSPPA
 CPSRPLTPPQPDLPASVAADWEFSKVPQIPRSGYWMNCVLILV
 AIARTRPGQGDGWSGQHGRGRWTRCLWPGALSMGYQGELMGLLSQPFSLPETQKDRAGKYQCQWKTPKKAGTRAGL
 EGFSATLSRSGSSQDSCPDRLHQEVCSFLPPSSRPSRRAALGF
 LGLGHPRHWDITPGLCPSPSPRPLPRRRSRSCLLCLVYVLFCLLPCCVRNPPNQRTRPGEQSAVLSPDLSLFPFATP
 SCPHKLHSPFRRGGSRLKGTQQLFIGSGELWGHKPLGVPV
 VDRLLSSPSLAAFLPSPGDHLVKGATLIEKVEPTPINVPHYPTPFCSRFPETQVVGFIRVMLVDSSIHVILLPHSL
 SSAPQNCSEFKQILQNIAPLLKILYVSNPARCRVRLAFLGCC
 LLNLKEYHSHKYSHFNFSPPSLSLSLFASTDMVLSDLQTIKDKIYREMRMLFYCWHIVPEHNNILFSSRRGTRGRKR
 GKGNMKGQEPHTKTPAMYNWDIYNFYMTTHCDVLIIPSSILF
 STSQMGPGSSASCQGVKKPGGWSGSDPVQSSTPIAASSCDILLRPPYSGGKYSYSAFFFSWAKQGHLLCQPQVSLG
 IALSPFPDHRELCPQALPKPPPQLILSCLERWEAWAPAPKKPE
 MIFSRWSWKAQGAWGRDWDLFYVLTLSQRSELRSQTNANVRVCICWQQVKCWGKVRQKSGKREGGEEMHELAEPF
 HIPFFSVLSKSSSGSLSTSLKRGITSAGLGGVWGQFSAHAQS
 ALFLLSPSLLIPSSLQLVSPGPVPKILAPPNDIHFVRCVKFHFSPCLNFERRSVPFWNAVCVAGAGMCSDSFPHHI
 YGKLEAPVSHALVFPNSIPCIQVLWQTRRLFAISFDEFHGDQR
 SLFIHKTPIGWSRIYFLISKKSLLIDQSGQLSSRSRTAKPPNSVIMRGAEQLLKLGLFLRKRDFSPPRVAQKGPAGQP
 FFSCSEKDLGVVAGSGSCGRVQAAEHQALRLSTESGSSAW
 VPEQPQEAPSCSCSAAVRNGVCVSVCEGWVTNIRAEGLDGRERRQGEHNGGPLENIVPKKFQISSLSLQISLPRKL
 IYFPELSGKQRPARGLPNTFFRLGPVPPTLSPVQSSHAPSLSL
 ARGVRLEWHLTCLGKSPSAQAAMPPLSLGTPGQLHPTQVQAGSKFLSYNRYQKHVFCIGEDWRSKWQFFPCILFLC
 FLNSVFESTNSFSESENLLLTSPSPFRMPDLGGGARAVARRPAS
 LMGLGGSVLFLARFIETKNKFSPLLPKTCAYLQSTTAVLIKHLFLLLLMSFHNCFTNLTKYKYIKNLSPSFPPID
 SHSFAKKIKLTNGKKKISNVGKIGLIPALPPWSESFQQISAAL
 KIEKEEGFFFPQEKVIGGGGGVENFCQRLSSGGIYLYFCFPFSPPSSISAFFTQQLVSVHYKSARPFSLCPLFRDV
 IKHLTFWGGSHSASQVANLPEPTDFRPQRIPTLGLPLSSPDMV
 EEGLKETASLLGAVVHAAPEAPSFQGPSLPHPDIKEIFKYIENQHSCLKVPHDILSSIIQFVVHCNNLISRKYHHD
 LSLQSLSEIGGGDRVTLSSGLFYLLHHYPGRKEAGNSERGYLC
 PGVHLVPLSPQSTPALLSLPHLLPSRGPEASASCWKPLLRHAMNSRTAGFGKKQLTLVSMFVYLPSPHPKNEITHLKF
 FFLIKSSSTLSPSFPWGEESVIKEEPLLESFPFPSKYRQISVL
 SGGGLGLYFLKPTQIVWQEDSSTPRGSCINHVSTSSCTLSAFGFETLFFPPSDTPGKPRYRRINEFFLSIVPFMALL
 LPIDYSALLHRRKRSNPATSGFLLPAFPPPNPPKNGGGGGIAF
 SAPRTLPTTLLCQLKAFSAYITISNVWTNQEKTPICTLSAVSVRLRLRPHKPRIQLPFRNRLFPFFSPKMGENGRC
 EAHPHHGLALGELLRTPFSLQASLRSGASTGGALSLEGAGGL
 LTKAGAACRKLTVHQSPSTLLFLAGSYVNYPTGPHWLQIINSHQHKQELMSLERFSEFSFFLSLSFSLSFPLRCHLP
 TFYFFFFSFFYCFYTFSLGKNAGRVPTEGSEGGRRPVGKG
 EQMRGPFSAHSSSATVPSLAPLDLVVPHIPTTFKTLSTLLTQGNRFYEISFYFFFFNYWKYFSDGTHSMECQNRIS
 CRISRDPPPVLGGSSSLGEIRMFKGADIKHETRLKIQGFPLWL
 WPQHSHKSVSWGEISWYSSPSLSPKALLRDMHPSFTTKVPPLHIYIYIYIFKKSYSYFQSTHLVSTRESTEAEAIFFPL
 PDIWLAIGWEGVSLCHFVCPGLEPSLPSSEYKLETEINIKEI

```
IDRPGKKGKMNSYFTNPSSLCHLAGGQDVLPNVALNSTAYDPVRHFSTYGAAVAQNRIYSTPFYSPQENVVFSSSRG
PYDYGSNSFYQEKDMLSNCRQNTLGHNTQTSIAQDFSSEQGRT
APQDQKASIQIYPWMQRMNSHSGEFYSSEIIKTELALPASLEERAERVFYGPIKTTITLVSSPRQGPLFCPLSSPLAR
LQGGPGLGPRIGGPQGGRLRGQEQGRGGEGAEDRGAPQSGVSLR
RREGERRGRVGRSFQGTRMLGEMGEPSPERGPSSPLSGIFQKSAFRLAAFCCLPFAPPPRPPPHHTPFFGTHRSHLIG
GLKLGSSLYPQWGWGKAKGRRLEPGGCELLARLGGSGLCAKQS
AEDALLIAFSVFCARIFRGLRSGPEARPPDLLAVPDPGTGEGISLQSLPNAAPAHRDRQRALPDRATDQNLVPEPP
DEVEKRISHIHSLGGRRRGHRRQPGRKRGKAGRDRRGEAERT
RTVPATPLSLSPSLPTNSPLITHSVFITGTIDVFFPKTKLGSQWTWTKSALDPLPHRTTLFHHAPPPPRSLASSFSAC
LQALFPVQALGARTLNSDSTDCPPSEDLAPLLRRPHRPPCR
EPAPGPAGASAPGPQGPAPWQPGRAGGPERRARLGPTPTPRASPSQLSPSPSAPANAQPRQIVFKESWGSLLWHFTNCDRF
CVKIFSCICGLCIYIYVHRQCSYPISKKEKKEGKLQKERKKVN
DVCLASRRKINKIPSCYPPVIQPLGPFNSINYYFNF
```

Número de codones: 4476

```
In [35]: standard_table = CodonTable.unambiguous_dna_by_name["Standard"]
print(standard_table)
```

Table Standard, SGC0

	T	C	A	G	
T	TTT F	TCT S	TAT Y	TGT C	T
T	TTC F	TCC S	TAC Y	TGC C	C
T	TTA L	TCA S	TAA Stop	TGA Stop	A
T	TTG L(s)	TCG S	TAG Stop	TGG W	G
C	CTT L	CCT P	CAT H	CGT R	T
C	CTC L	CCC P	CAC H	CGC R	C
C	CTA L	CCA P	CAA Q	CGA R	A
C	CTG L(s)	CCG P	CAG Q	CGG R	G
A	ATT I	ACT T	AAT N	AGT S	T
A	ATC I	ACC T	AAC N	AGC S	C
A	ATA I	ACA T	AAA K	AGA R	A
A	ATG M(s)	ACG T	AAG K	AGG R	G
G	GTT V	GCT A	GAT D	GGT G	T
G	GTC V	GCC A	GAC D	GGC G	C
G	GTA V	GCA A	GAA E	GGA G	A
G	GTG V	GCG A	GAG E	GGG G	G

Se importó la secuencia del gen abd-A


```
In [17]: t_seq2=seq_2.count('T')#TIMINAS
print(t_seq2)
```

6481

Se

obtuvo la proteína de nuestra secuencia y la cantidad de codones lo guardamos en una variable

```
In [18]: seq_tradu_2=seq_2.translate(stop_symbol='')
print(seq_tradu_2)
lenseq_tradu_2=len(seq_tradu_2)
print(lenseq_tradu_2)#numero de aminoacidos,codones
```

YCAINVSLLVTPQRSTFRSGFAFASLLARKAVSPKKKKTKNIKENPLRRQTKLKVSNDKRLSRCRAAVSDAAARAGSKNYQLHEIFILFKFRFIIIFIVARSSVCGTQLSTLWNF
 GPPPIPIRIIRLYDYTTVVRQYDSVSDSPCLLFRFGRRLIVTMLIANIKLCVKFKTEKTKTLWFITHCVRFVVKIKKTSRKKVLASPFQTSTIVLPACFMTFTKLKQNVGRKHFSGH
 THALEKSLKAAGNSLCHKNTRVPLKKIARLITRTKRRHDVHSNFRVIDDRILIAFTTNKRYTHPKHTHATHTLVPQLQIQNRPRPYANLTNFVNSPASSEKSNRNRKAANTSCSPKVR
 RKLEKGNKMHFTSRNREKLKVGQAKFPNSVSLTSLKKYLLAISYSTVKVTHILTRWHWQGNLAEKYIKHLDYIFFSLEFCIMLPERFMFNPINLVNLFMVSEFLTCTYFCFAIFNS
 CLSHWTPSLFLFLVPFNSINFVSDLYTKSNVVCIRLSRKVKYRGRCSLKTSKFTLIQIKLLCKSYEKVKEKIDFKMLFTVLNVKVIQVDLCAYRSSRNFCFIMARKRSSPRIQTKFV
 KVTQRGNFKLNEILNLAINQTQIKNPNHSNSESETKIEQQVDTKFRSDPRPESKICAMSKFVFDMLPKYPQFPFISSHLTTTPNSSSAVAAAAAASASVSASSSSNN
 NSSNTIAGSNTSNTNNSSSPSSSSNNNSNLNLGGSLSPSHLSQHLGQSPHSPVSSSPFQQHHPQVQQHNLHQQQHLLHQQQHLLHQYSSLSAALQLQQQHHISKLAAAASHG
 HAHQQLLLTPPSAGNSQAGDSSCSPSPASGSSSLHRSNDNSPGSASASASAAASSVAAAAAASSFAIPTSKMYPYVSNHPSSHGGLSGMAGFTGLEDKSCRDLPLCLSMGC
 RFLIGMGFMSNRKMLNDPNRCKYKVNDSILSNMIFKYLECIDRTSQTVSYFKIYANLVPLALPGFKTFKKMPGLKIQKKRNTQLFQELISRALNLASGNASDQLNLQTLQNRKS
 VSNALLSGLGWSNDVHFLGHRPFLNGHKVGKLEQPVVAAGNLSPATRFARTLTFRCGLHISTTVMALHVVQTLCCVGVFGISVQLELTAVWKGHANGHLISILYGRWSWERTMRAG
 ITNNGYFLSPTPTAGTRTLSTATSRAYLPRHLHSSLSINMPQPPHPRYPVPARSAWTRWEMPAHSPPPACQGGERAEPVSPICPGIPGRQLQVSQNSILDSAEIYYTVKKLPDSDS

7281 codones

Generamos un gráfico de barras para la comparación de nucleótidos:

```
In [19]: import matplotlib.pyplot as plt
fig, ax = plt.subplots()
plt.title("NUMERO DE NUCLEOTIDOS")
plt.xlabel('SECUENCIAS')
plt.ylabel('NUCLEOTIDOS')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [nuc_homo, nuc_mosca])
plt.show()
```

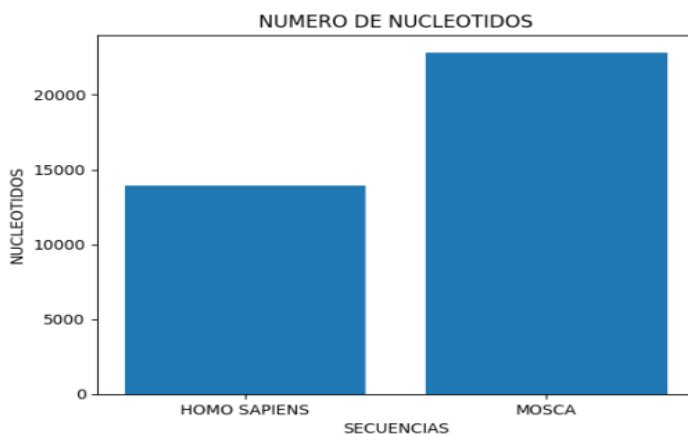


Figura 15- Comparación del número de nucleótidos del gen HOXC6 del Homo sapiens y abd-A de Drosophila melanogaster. (Elaborado por el autor)

```
In [20]: fig, ax = plt.subplots()
plt.title("NUMERO DE ADENINAS")
plt.xlabel('SECUENCIAS')
plt.ylabel('ADENINAS')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [a_seq1, a_seq2])
plt.show()
```

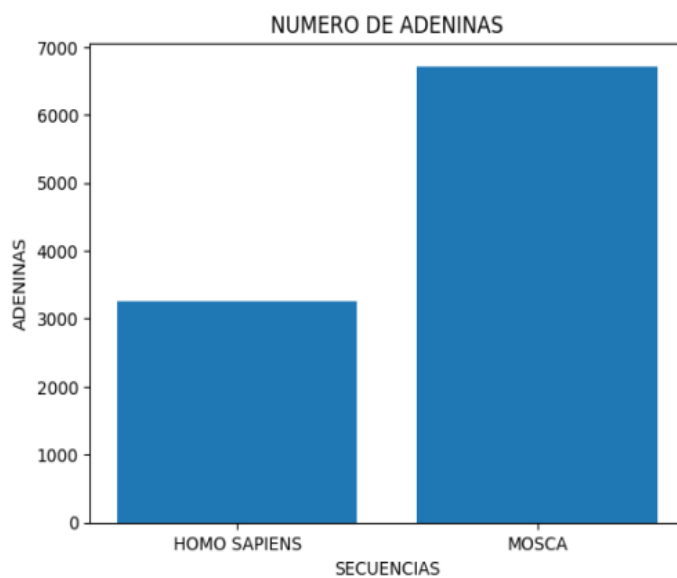


Figura 16. Comparación del número de adeninas del gen HOXC6 del Homo sapiens y abd-A de Drosophila melanogaster. (Fuente: el autor)

```
In [21]: fig, ax = plt.subplots()
plt.title("NUMERO DE TIMINAS")
plt.xlabel('SECUENCIAS')
plt.ylabel('TIMINAS')
ax.bar(['HOMO SAPIENS', 'MOSCA'], [t_seq1, t_seq2])
plt.show()
```

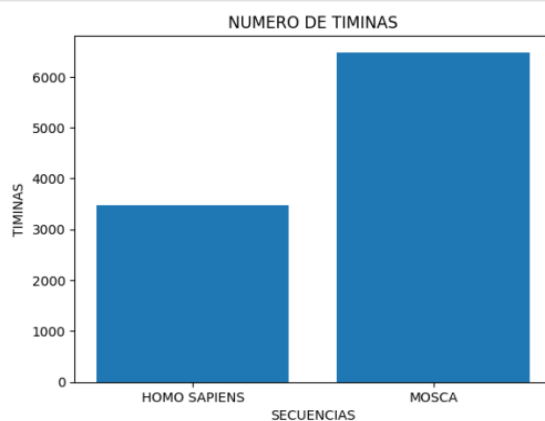


Figura 17 Comparación del número del número de timinas del gen HOXC6 del Homo sapiens y abd-A de Drosophila melanogaster. (Fuente: el autor)