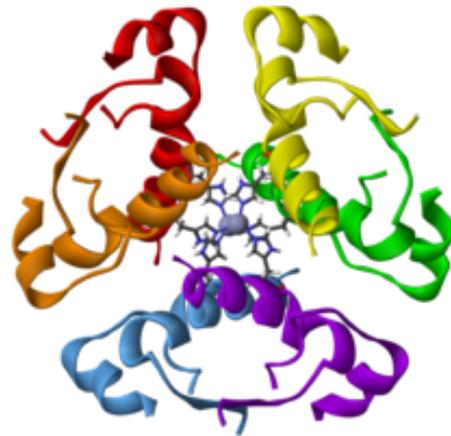




INSTITUTO IES RAYUELA DE MÓSTOLES



SECUENCIAS DE PROTEÍNAS COMO INDICADOR DE LAS RELACIONES EVOLUTIVAS DE LOS CORDADOS

**Realizado por: Lydia Flores Gigante 2ºA Bachillerato
Tutor: Arturo González Laguna**

RESUMEN

Las proteínas cumplen funciones clave en todas las especies de seres vivos. Además, pueden servir para reconstruir la filogenia de estas. El objetivo de este estudio es, a partir de las relaciones taxonómicas actualmente conocidas de los cordados, comprobar si las secuencias de aminoácidos de diferentes especies en tres proteínas (insulina, citocromo c e histona H1) siguen estas mismas relaciones y ver de entre esas tres, cuál ofrece mejores resultados al reconstruir los árboles evolutivos. Escogí 15 especies del grupo de los cordados y después de alinear sus secuencias, hacer los árboles y analizarlos obtuve que la insulina es la que mejor resultados muestra para reconstruir la filogenia, después el citocromo c, porque las secuencias entre las especies se parecían demasiado y un pequeño cambio hacía que una especie no estuviera emparentada con la que tendría que estar, y por último la histona H1, porque era bastante variable. Con esto, la conclusión que se obtiene, es que estas tres proteínas pueden reconstruir las relaciones taxonómicas actuales de los cordados, pero la insulina da mejores resultados.

ABSTRACT

Proteins fulfill key functions in all species of living beings. In addition, they can be used to reconstruct their phylogeny. The objective of this study is, based on the current taxonomic relationships of chordates, to demonstrate if the amino acid sequences from different species of three proteins (insulin, cytochrome c and histone H1) follow these same relationships and to see among those three, which one offers better results by reconstructing evolutionary trees. I chose 15 species from the chordate group and after aligning their sequences, making the trees and analyzing them, I obtained that insulin is the best to reconstruct the phylogeny, then cytochrome c, because the sequences between the species were too similar and a small change made a species not show its relationship with which it should be, and finally the histone H1, because it was quite variable. With this, the conclusion that is obtained is that these three proteins can reconstruct the current taxonomic relationships of the chordates, but insulin gives better results.

PALABRAS CLAVE

- Proteína
- Evolución
- Árbol filogenético
- Alineamiento
- Secuencia
- Especie
- Cladograma
- Cordado

KEYWORDS

- Protein
- Evolution
- Phylogenetic tree
- Alignment
- Sequence
- Species
- Cladogram
- Chordate

AGRADECIMIENTOS

Quiero en primer lugar, agradecer a mi tutor, Arturo González, por guiarme y aconsejarme en todas las fases del proyecto, y por la dedicación y el incondicional apoyo para que este trabajo saliera adelante.

Además, dar las gracias a Óscar Gómez, coordinador de los Trabajos de Investigación y al centro IES Rayuela por brindarme esta oportunidad de investigar y aprender en el proceso.

ÍNDICE

1.- INTRODUCCIÓN	5
2.- CONTEXTUALIZACIÓN DEL TRABAJO	7
2.1.- SECUENCIA, ESTRUCTURA Y FUNCIONES DE LAS PROTEÍNAS	7
2.2.- TIPOS DE PROTEÍNAS	13
3.- ALINEAMIENTO DE SECUENCIAS Y SU FUNCIÓN	15
4.- MATERIAL Y MÉTODOS	18
5.- TRABAJO DE CAMPO	21
5.1.- PROTEÍNAS Y ESPECIES ESCOGIDAS	21
5.2.- RESULTADOS	28
5.2.1.- Insulina	28
5.2.2.- Citocromo c	32
5.2.3.- Histona H1	36
5.2.4.- Estructura tridimensional	40
6.- CONCLUSIONES	42
7.- FUENTES CONSULTADAS	44

1.- INTRODUCCIÓN

En 1962, Emile Zuckerkandl y Linus Pauling publicaron un artículo sobre el reloj molecular, que en genética, es una técnica para datar la divergencia de dos especies. Ellos comenzaron sus investigaciones con las cadenas α y β de la hemoglobina (hemoproteína de la sangre) y demostraron que las diferencias de aminoácidos entre especies encajaban con el ritmo evolutivo, que estaba basada en la evidencia fósil.

El tema de mi investigación está relacionado estrechamente con este descubrimiento. Este proyecto nos transmite el paso del tiempo y la fundamental importancia de estas moléculas formadas por aminoácidos, llamadas proteínas, presentes en todos los organismos, que han ido evolucionando mediante mutaciones producidas en sus genes.

Después de reflexionar sobre el tema que iba a investigar, me decanté por las proteínas. Desde que lo vimos en clase, me pareció muy interesante que estas moléculas tuviesen tantas funciones en el organismo y que estuvieran organizadas con esa estructura. También me atrajo, el hecho de que muchas proteínas sean similares en diferentes especies muy alejadas en el árbol evolutivo y que estas tengan la misma función.

La hipótesis a demostrar es, a partir de las relaciones taxonómicas conocidas actualmente de los cordados, ver si las secuencias de aminoácidos (proteínas) de 15 especies del filo de los cordados siguen estas mismas relaciones. Para realizar este trabajo, he escogido tres proteínas diferentes, la histona H1, el citocromo c y la insulina, ya que están en todos los organismos eucariotas y hay más secuencias de aminoácidos en bases de datos de estas proteínas. El objetivo es ver cómo cambian su secuencia de aminoácidos a lo largo de la evolución y qué partes de la proteína son las más importantes para su función, es decir, están más conservadas, y de entre estas tres proteínas, deducir cuál sigue mejor las relaciones taxonómicas de los cordados para poder reconstruir árboles filogenéticos. Para ello, a partir de la base de datos de proteínas secuenciadas como las que hay en el NCBI (National Center for Biotechnology Information), voy a comparar las secuencias de aminoácidos de las especies (alineación) y saber así en qué se diferencian una especie de otra, para después hacer cladogramas o árboles evolutivos y ver cuál de las tres proteínas da mejores resultados. En

teoría, la proteína que debería seguir mejor las relaciones taxonómicas sería la histona H1 porque es una proteína estructural, que cambia poco a lo largo de la evolución, después el citocromo c y la insulina.

Las conclusiones y resultados que quiero lograr son los que validen mi hipótesis, esperando obtener que las secuencias de aminoácidos de las especies se parezcan a las relaciones taxonómicas actuales.

Espero que este trabajo aporte a los lectores una base sobre el funcionamiento de las proteínas y su evolución. Así pues, profundizar también en estos contenidos que se dan muy por encima en bachillerato.

Quería hacer este trabajo porque tengo un espíritu muy emprendedor, y me gusta ponerme retos difíciles. De este proceso he aprendido de mí misma, de mejorar mi rendimiento del tiempo y el esfuerzo empleado. También he aprendido a manejar fuentes, sintetizar la información, mejorar la redacción y solventar los problemas cuando se han presentado. Además de usar bases de datos complejas como el NCBI y programas como Jalview o Clustal Omega.

2.- CONTEXTUALIZACIÓN DEL TRABAJO

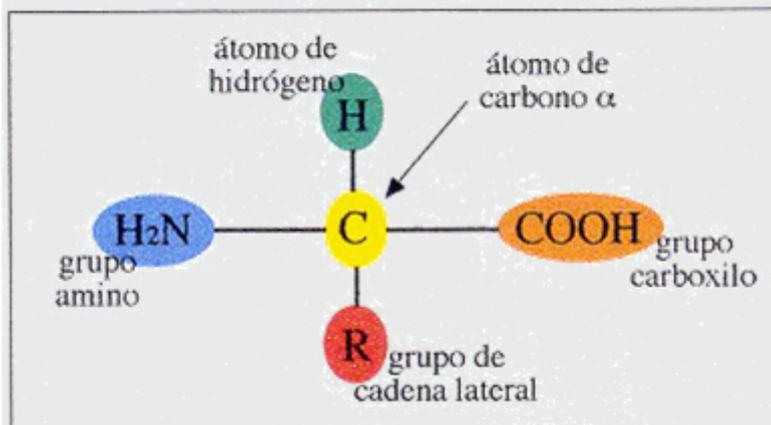
2.1.- SECUENCIA, ESTRUCTURA Y FUNCIONES DE LAS PROTEÍNAS

Las proteínas son macromoléculas formadas por cadenas lineales de aminoácidos. Son los materiales que desempeñan un mayor número de funciones en las células de todos los seres vivos, como formar parte de la estructura básica de los tejidos y desempeñar funciones metabólicas y reguladoras. También son los responsables de la expresión del código genético y como por ejemplo, de los sistemas de reconocimiento de organismos externos en el sistema inmunitario.

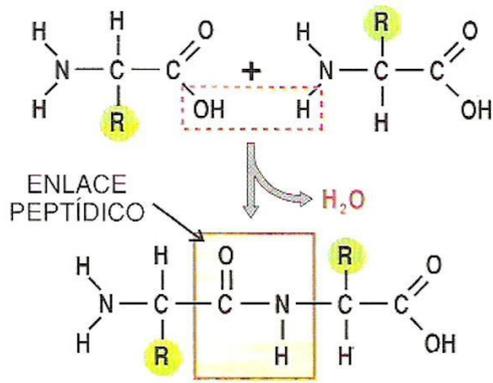
Desde el punto de vista de su composición elemental, todas las proteínas contienen carbono (C), hidrógeno (H), oxígeno (O) y nitrógeno (N) y algunas contienen azufre (S).

Los aminoácidos son las unidades estructurales o monómeros de las proteínas. Son moléculas orgánicas que cuentan con un grupo amino ($-NH_2$), un grupo carboxilo o ácido ($-COOH$), un átomo de hidrógeno y una cadena lateral o radical variable ($-R$).

La fórmula general de un aminoácido es:



La unión de varios aminoácidos se hace mediante una reacción de condensación entre el grupo amino de un aminoácido y el carboxilo de otro, liberando una molécula de agua y formando así un enlace peptídico entre los aminoácidos. La unión de varios de estos aminoácidos da lugar a las cadenas llamadas péptidos (entre 2 y 100 aminoácidos), que al superar este número, son llamados proteínas.



De todos los aminoácidos presentes en la naturaleza, tan sólo 20 son los constituyentes básicos de las proteínas y son los que llamamos aminoácidos proteicos (α -aminoácidos).

Las proteínas son biomoléculas de elevado peso molecular, por eso se las llama macromoléculas y presentan una estructura química compleja. Estas cadenas de aminoácidos no son polímeros al azar, sino que cada una de ellas posee una determinada composición química, un peso molecular y una secuencia ordenada de aminoácidos.

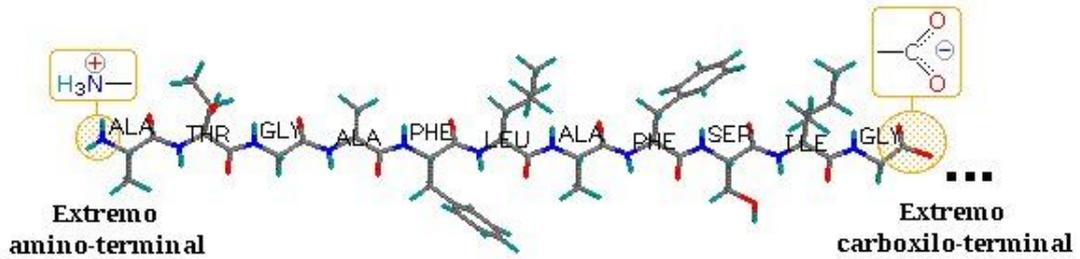
La actividad biológica de una proteína depende en gran medida de la disposición espacial de su cadena polipeptídica. Dicha cadena sufre una serie de plegamientos que se realiza mediante rotaciones de los enlaces simples del esqueleto, pero hay una serie de restricciones a la libertad de giro de estos enlaces, la mayoría de ellas debido a la interacción de la cadena polipeptídica con las moléculas de agua que la rodean, las cuales determinan que solo haya una posible configuración tridimensional que sea funcional.

La organización de una proteína viene definida por cuatro niveles estructurales denominados: estructura primaria, estructura secundaria, estructura terciaria y estructura cuaternaria.

Niveles estructurales de las proteínas

Estructura primaria

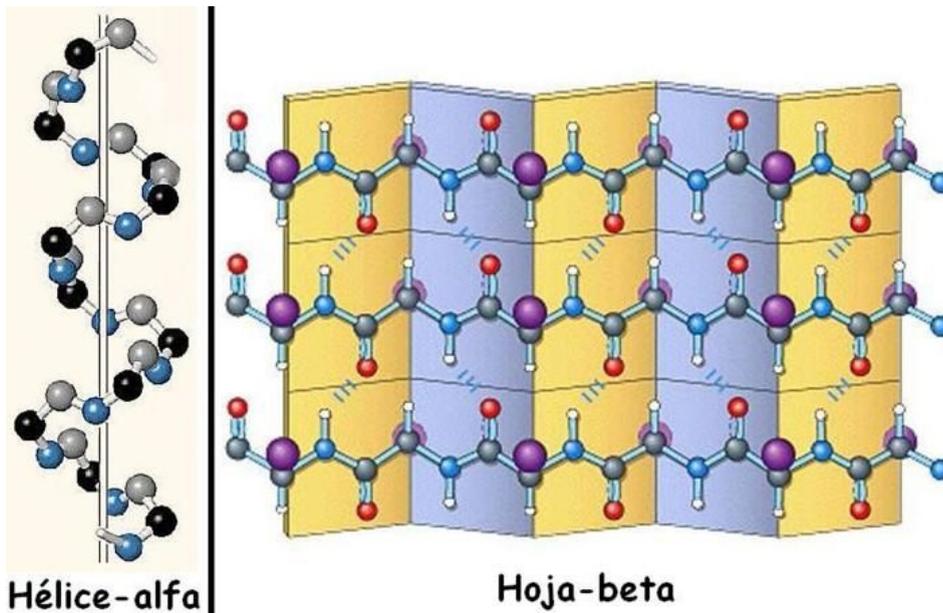
Esta estructura es la secuencia lineal y ordenada de aminoácidos unidos mediante enlaces peptídicos.



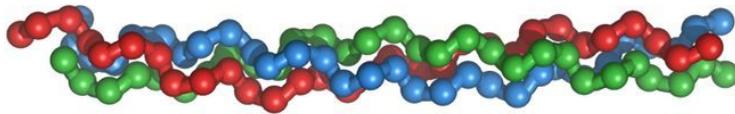
Estructura secundaria

La estructura primaria se pliega en el espacio, estableciéndose enlaces o puentes de hidrógeno en el esqueleto de la propia cadena polipeptídica. Los tipos básicos de la estructura secundaria son:

- α -hélice: plegamiento en hélice de la cadena polipeptídica sobre sí misma.
- Lámina plegada o lámina β : el plegamiento origina una estructura laminar y la cadena polipeptídica se dispone en paralelo en forma de zig zag.

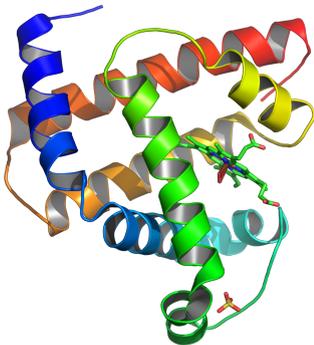


- Hélice de colágeno: consiste en tres cadenas peptídicas enrolladas en espiral y tienen una secuencia repetida de aminoácidos.



Estructura terciaria

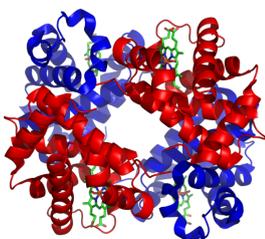
Se refiere a la disposición de la estructura secundaria de un polipéptido en el espacio producidos por enlaces entre las cadenas laterales de los aminoácidos. De la estructura terciaria depende por lo tanto la función de la proteína, por lo que cualquier cambio que se produzca en la disposición de esta estructura puede provocar la pérdida de su actividad biológica, proceso que conocemos con el nombre de desnaturalización. Dicho proceso puede estar causado por cambios de temperatura, variaciones del pH, entre otros.



Estructura cuaternaria

Se produce por la unión de varias cadenas polipeptídicas con estructura terciaria para formar un complejo proteico (grupo de dos o más proteínas). Cada una de estas cadenas polipeptídicas recibe el nombre de una subunidad proteica, que pueden ser iguales o muy parecidas o completamente diferentes y dependiendo del número de subunidades que se asocian, pueden denominarse dímeros, tetrámeros, pentámeros...

Ejemplo: hemoglobina



Factores que determinan la estructura tridimensional de una proteína

Los factores que influyen son el tipo de enlace, las fuerzas eléctricas y los puentes de hidrógeno entre otros.

En las proteínas existen dominios que contribuyen a esta estabilidad que son comunes a muchas proteínas. La estructura es estable bajo determinadas condiciones de salinidad, pH y temperatura. Si cambian estas condiciones pueden modificar la estructura y por tanto las propiedades y su función. Aunque, dependiendo del tipo de proteína y circunstancias en las que se ha desnaturalizado, la proteína puede quedar modificada permanentemente o puede volver a su configuración original y ser funcional de nuevo (renaturalización).

Propiedades de las proteínas

La secuencia de aminoácidos de cada proteína es específica de cada especie. Las propiedades de un péptido o proteína dependen de su secuencia, especialmente de los aminoácidos orientados al exterior, y de la estructura espacial.

Solubilidad

Son solubles las que presentan grupos polares o iónicos (átomos de diferente electronegatividad) al exterior (generalmente globulares) y al contrario son insolubles las que presentan muchos grupos apolares (que no llevan carga ni positiva ni negativa).

Especificidad de función

La estructura y propiedades tan variables de las proteínas hacen que puedan cumplir numerosas y diversas funciones biológicas. Esto se debe, a que su secuencia puede estar formada por 20 posibles aminoácidos, y pueden crear moléculas con muchísimas posibles propiedades químicas y de ahí su variedad funcional. Dentro de una misma proteína pueden existir distintos lugares que determinan la estructura, función o regulación de la proteína.

Proteínas homólogas

Al igual que otras características de los seres vivos existen proteínas semejantes en especies próximas y van variando con la escala filogenética. Esto es un reflejo de la evolución desde

un antepasado común, en el que pueden realizarse árboles filogenéticos para estudiar las variaciones de las proteínas, que es lo que se va a hacer en este trabajo. El ritmo de variación de una proteína depende de la función que tenga. Hay algunas muy variables, como las de reserva y otras muy estables, como las estructurales o algunas enzimas.

Regulación proteínica

Algunas proteínas se encuentran siempre en el mismo estado (su función es siempre la misma en las condiciones fisiológicas adecuadas). Otras proteínas presentan varios posibles estados con actividad diversa como la regulación a corto plazo, que depende de las condiciones cambiantes del medio externo o por cambios internos.

Hay otro tipo de regulación, la de largo plazo, en el que una modificación en la estructura de la proteína la modifica indefinidamente. Este cambio puede ser reversible o irreversible (como las enzimas digestivas). Estas regulaciones de las proteínas tienen que ver siempre con su estructura tridimensional.

Funciones de las proteínas

Las proteínas son las moléculas más versátiles de la célula, como antes he mencionado, ya que cumplen multitud de funciones en las células.

- Son estructurales: dan forma y soporte a orgánulos, tejidos, células y organismos.
- Son transportadores en agua: solubilizan sustancias hidrófobas.
- Enzimas: aumentan las velocidades de reacción química.
- Estructuradoras de proteínas: ayudan al plegamiento de las proteínas o las eliminan si no son funcionales.
- Transportadoras de membrana: transportan sustancias a través de las membranas biológicas.
- Regulación génica: expresan o inhiben la expresión de determinados genes.
- Receptores de estímulos: captan un estímulo y mandan un mensajero intracelular.
- Movimientos: producen movimiento por desplazamiento de unas moléculas sobre otras.
- Reserva de aminoácidos: algunas proteínas almacenan aminoácidos que son utilizados como elementos nutritivos y unidades estructurales por el embrión en desarrollo.
- Toxinas: dañan células u organismos.
- Hormonas y neurotransmisores: moléculas mensajeras intercelulares.

- Protección frente a parásitos: lucha inmunitaria frente a microorganismos, endoparásitos y virus.
- Reconocimiento: informan de la identidad celular.
- Coagulación: evitan la salida de líquidos a presión.
- Regulación del pH: atemperan los cambios de pH.

2.2.- TIPOS DE PROTEÍNAS

Se pueden aplicar distintos criterios para clasificar a las proteínas y no existe un sistema universal de clasificación.

Según su complejidad pueden ser:

- Proteínas simples u holoproteínas. Están compuestas por aminoácidos.

Dentro de estas, se pueden diferenciar las proteínas fibrosas, que suelen estar ordenadas a lo largo de una sola dimensión con una estructura secundaria bien definida y proteínas globulares, que se caracterizan por doblar su estructura secundaria en una forma esférica.

- Proteínas conjugadas o heteroproteínas: formadas por la unión de aminoácidos y otras moléculas no proteicas (grupo prostético). Se pueden diferenciar varios tipos en función del grupo prostético que lo acompañe como nucleoproteínas (ácidos nucleicos), glucoproteínas (glúcidos), fosfoproteínas (ácido fosfórico) o lipoproteínas (lípidos).

Enzimas

Son generalmente proteínas con función catalítica, es decir, aumentan la velocidad de las reacciones bioquímicas y son responsables del metabolismo celular. En estas reacciones, las enzimas actúan sobre unas moléculas denominadas sustratos, las cuales se convierten en moléculas diferentes denominadas productos. Casi todos los procesos en las células necesitan enzimas para que ocurran a unas tasas significativas. Constan de un centro activo que es la zona de la enzima al cual se une el sustrato, para que la reacción se produzca. La estructura y composición del centro activo está configurado para que únicamente un determinado sustrato tenga la semejanza suficiente como para unirse a esta zona de la enzima. Y también tienen interacción alostérica, que consiste en el cambio estructural de una proteína, es decir, sus

receptores cambian. Las enzimas alostéricas poseen dos sitios activos, en el primero se unirá un sustrato determinado, mientras en el segundo se unirá un efector el cual producirá un cambio en la proteína.

3.- ALINEAMIENTO DE SECUENCIAS Y SU FUNCIÓN

Los ácidos nucleicos y las proteínas son secuencias lineales de monómeros, nucleótidos y aminoácidos, respectivamente.

Para entender la utilidad de la comparación de secuencias y de la búsqueda de homólogos, es decir, proteínas que tienen un origen evolutivo común, se necesitan saber algunos conceptos acerca de cómo evolucionan las proteínas.

A lo largo de la evolución unas especies van dando lugar a otras nuevas. Detrás de estas especiaciones está la variación genética de los organismos, con la evolución de los genomas, sus genes y las proteínas. Generalmente, cuanto más similares sean dos secuencias, más parecido tenderán a ser las funciones de las proteínas, y cuando las secuencias sean más distintas, quiere decir que las especies están más alejadas filogenéticamente. Sin embargo, puede haber cambios muy pequeños en las secuencias que lleven a que tengan funciones muy diferentes.

La base de la evolución en los genes y las proteínas son las mutaciones en el ADN. Cuando el ADN se replica se pueden producir errores al realizar la copia o también pueden introducirse mutaciones por efecto de agentes externos. Dentro de las mutaciones se pueden distinguir las duplicaciones, que permiten a una copia evolucionar más rápidamente llegando a adquirir nuevas funciones biológicas. Además, muchas proteínas están constituidas por entidades estructurales independientes, denominadas dominios. Mediante recombinaciones se puede producir un barajado de dominios, obteniendo así nuevas proteínas.

Cuando observamos diferencias entre secuencias de aminoácidos que sabemos que tienen un origen evolutivo común, estas diferencias están hablándonos de las propiedades funcionales de las proteínas. Algunas diferencias estarán allí porque la mutación no ha alterado la función, otras diferencias quizás tengan que ver con el hecho de que esas proteínas, aunque tengan un origen común, realicen funciones distintas y otras tendrán que ver con el modo de vida del organismo.

Por otra parte, el hecho de que determinadas posiciones de las secuencias permanezcan invariables, nos indica que estas tienen una especial importancia para el mantenimiento de la estructura o la función de la proteína.

Los organismos responden a una mutación dependiendo de si estas producen alteraciones en la estructura o en la función de la proteína. Sin embargo, en el caso de que se trate de un gen duplicado, la pérdida de función del nuevo gen no es determinante ya que existe otra copia. Según lo beneficiosos o perjudiciales que sean los cambios en las proteínas, los organismos que los porten serán seleccionados positiva o negativamente, lo que determinará que los cambios pasen a generaciones posteriores. Las secuencias que observamos ahora reflejan toda una historia evolutiva en la que las proteínas han ido divergiendo, adquiriendo nuevas funciones, adaptándose a nuevos entornos...

La comparación de secuencias, es descubrir qué partes de las secuencias son más importantes (están más conservadas). Encontrar qué proteínas tienen un origen común puede servir para predecir la estructura y función de las proteínas.

La elección de comparar secuencias de nucleótidos o de aminoácidos depende de la información que se quiera buscar. El parecido entre secuencias de nucleótidos con un origen común se pierde más rápidamente que el parecido en las secuencias de aminoácidos.

En este trabajo en concreto, se va a trabajar con el alineamiento múltiple de secuencias de aminoácidos.

Un alineamiento múltiple de secuencias es un alineamiento de tres o más secuencias biológicas, generalmente proteínas, ADN o ARN. En general, se asume que el conjunto de secuencias tienen una relación evolutiva por la cual comparten un linaje y descienden de un ancestro común. Del alineamiento puede llevarse a cabo el análisis filogenético para evaluar los orígenes evolutivos compartidos por las secuencias. El alineamiento múltiple de secuencias a menudo se utiliza para evaluar la conservación de los dominios proteicos, las estructuras terciarias y secundarias, e incluso aminoácidos o nucleótidos individuales.

Como puede ser difícil alinear a mano tres o más secuencias de longitud biológicamente relevante, y casi siempre consume mucho tiempo, se utilizan algoritmos computacionales para producir y analizar los alineamientos. Los alineamientos múltiples requieren metodologías más sofisticadas que los alineamiento de pares porque son computacionalmente más complejos de producir.

Este es un ejemplo de alineamiento. En él aparece la secuencia de aminoácidos de la especie con guiones, que se emplean para que se pueda realizar el alineamiento múltiple de todas las especies correctamente. Cada aminoácido está representado con una letra.

```
>NP_001022529.2 INSulin related [Caenorhabditis elegans]
MR-----NSFQLILLVLTISYSNSMLF-ND-----VEMDKPSYLSSEAWKMLNELLKEPK-----HHHHHRHKG YCG
VKAVKCLKQICPDLC SNVD-----D-----NLLMEMCSKNLTDDDILQRCCPE-----
-
```

4.- MATERIAL Y MÉTODOS

He utilizado diversas fuentes de información y programas que a continuación voy a explicar para realizar el trabajo de campo.

Obtención de secuencias de proteínas

Las secuencias de proteínas que he escogido, las he obtenido del National Center for Biotechnology Information (NCBI) o el Centro Nacional para la Información Biotecnológica. Fue fundado el 4 de noviembre de 1988 y forma parte de la Biblioteca Nacional de Medicina de Estados Unidos y una rama de los Institutos Nacionales de Salud. Tiene como finalidad ser un recurso nacional con acceso libre y gratuito para todo el público de información sobre biología molecular y actualiza información constantemente de secuencias genómicas, artículos científicos, biotecnología, bioquímica, genética y genómica.

Este es un ejemplo de cómo están representadas las secuencias de cada especie. Cada secuencia comienza con un mayor que (>) seguido de un código de letras y números que corresponde a esa secuencia de ese organismo en concreto y sirve para buscarla en la base de datos rápidamente. Seguido de esto, comienza la secuencia de aminoácidos en cuestión. Cada aminoácido está representado con una letra y empieza por el grupo amino de la cadena y termina por el carboxilo.

```
>NP_001335969.1 insulin-like 3 protein isoform 2 [Ciona intestinalis]  
MSETSRVVSATRPSSGGTYRRLCLVTCIVVTCMAIFPGSSIAHGRKRFAALSCGSRLVNSMRYMCLQEVWTPKLCVSVFRTKFCGTEINFHPI  
CRFTCITESRITSCFPIPAQQH
```

El NCBI está formado por un amplio y diverso banco de bases de datos, herramientas y otros medios que hace posible la búsqueda de información sobre las ciencias de la vida y la salud. Con esto, es posible acceder a las principales bases de datos de secuencias de proteínas y ADN disponibles, como Nucleotide, dbSNP, Proteins, PubChem Substance o Gene.

Este programa agrupa sus bases de datos esenciales en tres grandes sectores: Literature Databases, Molecular Databases y Genomes. Estas dos últimas clases comprenden un grupo amplio de bases de datos biológicas cuya información procede de los resultados de experimentos científicos, suministrados por los laboratorios o instituciones que los realizan. La información contenida en estas bases de datos comprende: funciones, estructura y

localización de los genes, los efectos clínicos de las mutaciones, así como las similitudes entre secuencias y estructuras biológicas. Además, el NCBI ofrece algunas herramientas bioinformáticas para el análisis de secuencias de ADN, ARN y proteínas, como BLAST.

Similitud de secuencias

A continuación utilicé un programa para alinear todas las secuencias de aminoácidos de las especies. Empleé Clustal Omega, que las alinea utilizando un algoritmo basado en un método progresivo, alineando primero las secuencias con mayor parentesco, incorporando las menos relacionadas y generando nuevos alineamientos por orden de similitud.

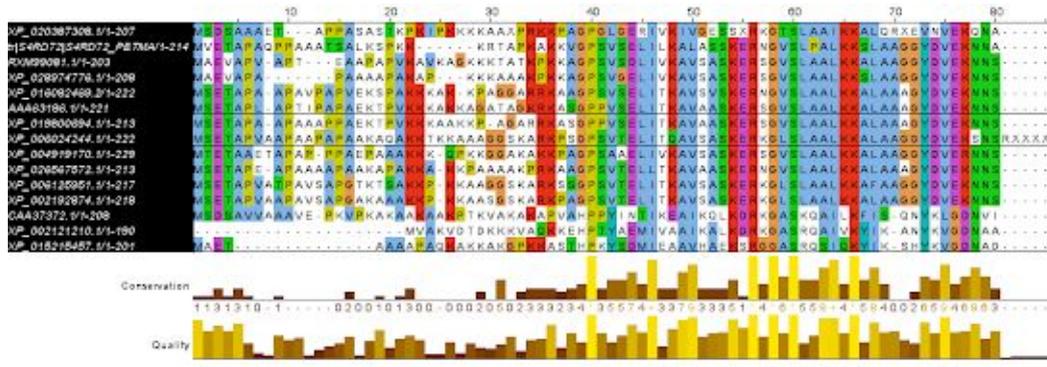
Generación de árboles de semejanzas

Para finalizar, utilicé Jalview, un programa de software de bioinformática que se utiliza para observar y editar múltiples alineaciones de secuencias. Jalview tiene una amplia gama de funciones además de la generación, visualización y edición de alineación de múltiples secuencias, busca semejanzas y distancias que pueden coincidir con árboles filogenéticos reales y la visualización de estructuras moleculares.

Para analizar estos datos Jalview utiliza métodos de distancia empleando un algoritmo para construir un único árbol a partir de los datos. Este algoritmo convierte las secuencias alineadas en matrices por diferencia por pares, dando lugar a que la distancia que hay entre cada taxón, sean los caracteres que tienen diferentes. Es decir, cuanta más diferencia hay entre dos taxones, es más probable que haya pasado más tiempo desde que divergieron del ancestro común. Por tanto, dos especies están más emparentadas cuando menor es la distancia molecular entre ellas.

Además, Jalview hace gráficos de semejanza, describiendo qué partes de las secuencias son las más conservadas. En la siguiente imagen, se puede ver un ejemplo de estos gráficos. Los aminoácidos son representados con letras a los que se les asigna un color, para hacerlo más visual y para que a simple vista sepas dónde las secuencias se parecen más. La primera fila se refiere a la conservación (conservation), que mide el número de propiedades físicoquímicas conservadas para cada columna de la alineación, cuanto más amarillo sean las barras, significa que ese aminoácido está más conservado. La segunda fila hace referencia a la

calidad (quality). Si tiene un puntaje de alta calidad de alineación para una columna sugiere que no hay ninguna mutación. En conclusión, las columnas más amarillas, permanecen más invariables, por lo que esta parte indica que tiene una especial importancia para el mantenimiento de la estructura o la función de la proteína.



Los datos de secuencia, tanto de ADN como de proteínas, tienen ventajas que potencian el uso de técnicas estadísticas de deducción filogenética. La cantidad de información de la que se dispone es mayor y son comparables mediante los programas antes mencionados.

Aún así, existen problemas comunes con las filogenias moleculares. Las especies que se comparan pueden haber divergido mucho o muy poco y el alineamiento de las secuencias puede resultar complejo, ya que se pueden hacer diferentes alineaciones de la misma región.

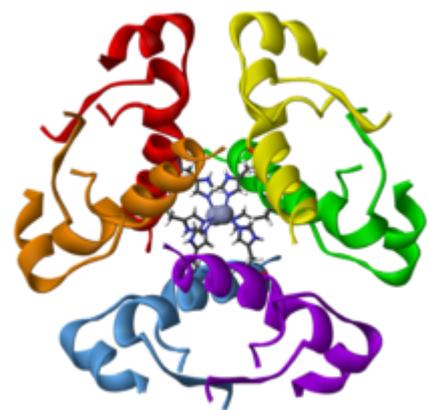
5.- TRABAJO DE CAMPO

5.1.- PROTEÍNAS Y ESPECIES ESCOGIDAS

Para realizar el trabajo de campo, he escogido tres proteínas, la insulina, el citocromo c y la histona H1 y 15 especies del grupo de los cordados. Además, he escogido una especie adicional, un nematodo, para tener un grupo externo. El grupo externo está menos relacionado con los demás taxones, por ello, proporciona información sobre características que aparecen en el grupo interno y que no aparecen en el grupo externo.

INSULINA

La insulina es una hormona polipeptídica que interviene en el aprovechamiento metabólico de los nutrientes, sobre todo en el anabolismo de los glúcidos, cuya función es regular los azúcares. La insulina permite disponer a las células del aporte necesario de glucosa para los procesos con gasto de energía.



En su forma activa, tiene 2 cadenas, una de 21 aminoácidos, la A y otra de 30 aminoácidos, la B, unidas por 2 enlaces disulfuro y existe un tercer enlace disulfuro dentro de la cadena A. La estructura secundaria es muy compleja para el tamaño de la molécula, presentando giros β en ambas cadenas y láminas β en la cadena B.

Las células de nuestro organismo, tienen dificultades para hacer que proteínas tan pequeñas se plieguen correctamente en estructuras estables. Esto se soluciona al traducir el ARNm de la insulina como un precursor de una única cadena polipeptídica llamado preproinsulina de 110 aminoácidos.

La conformación estructural de la insulina es esencial para su actividad como hormona. La insulina es sintetizada y almacenada en el cuerpo en forma de un hexámero, es decir, una unidad compuesta por seis insulinas, mientras que su forma activa es la de una hormona monomérica, es decir, la molécula de insulina sola. Seis moléculas de insulina permanecen

inactivas por largo tiempo en su forma hexamérica, como forma de almacenamiento de disponibilidad rápida.

La insulina es un ejemplo típico de proteína globular en la que encontramos los aminoácidos no polares en el centro formando un núcleo hidrofóbico y los aminoácidos polares y cargados en torno a este.

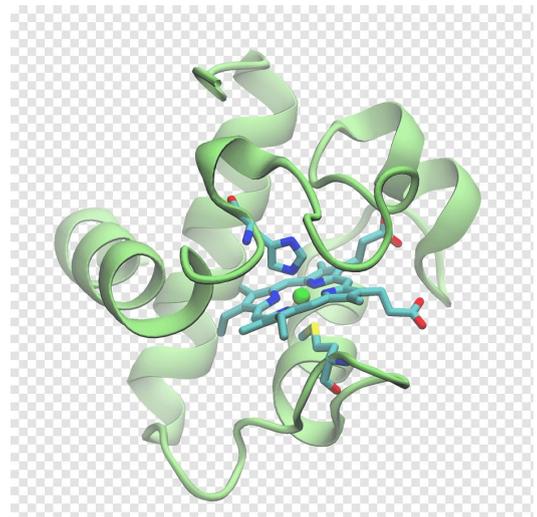
Entre los vertebrados, la insulina conserva una gran semejanza estructural. Por ejemplo, la insulina bovina difiere de la humana en solo tres aminoácidos o en la insulina del invertebrado *Caenorhabditis elegans*, un nematodo, es muy similar en estructura, tiene efectos celulares muy parecidos y se produce de manera análoga a la de los humanos. De modo que es una proteína que se ha preservado a lo largo de la evolución del tiempo, sugiriendo su papel fundamental en el control metabólico.

CITOCROMO C

El citocromo c es una proteína pequeña, que es componente de la cadena de transporte de electrones en la mitocondria.

Se encuentra situado en las mitocondrias de la mayoría de células eucariotas, en la membrana mitocondrial interna y está codificado por el ADN nuclear.

Se trata de una proteína monomérica, es decir con un solo polipéptido y tiene un grupo prostético, hemo C unido a esta estructura.



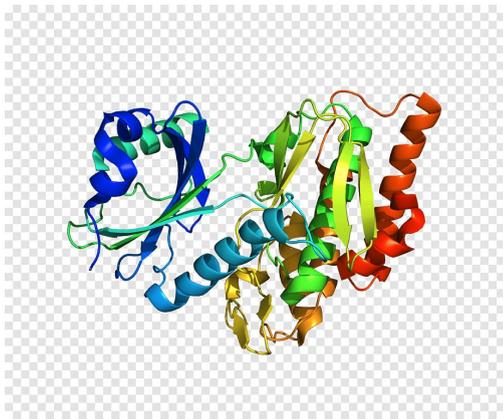
El citocromo c de humanos y chimpancés está formado por 104 aminoácidos, exactamente los mismos y en el mismo orden. El citocromo del caballo difiere del humano en 12 aminoácidos, lo cual, el grado de similitud refleja la proximidad del ancestro común y permite reconstruir la filogenia de estos organismos.

HISTONA H1

Las histonas son proteínas básicas, de baja masa molecular, y están muy conservadas entre los eucariontes. Forman la cromatina, junto con el ADN, que resuelve el problema del tamaño del ADN dentro del núcleo, compactándolo.

La histona H1 es un protector principal de la estabilidad del genoma y una proteína vital para el organismo. La supresión de la histona H1 produce daño celular y lesiones en el ADN. El descontrol en una zona habitualmente reprimida de la cromatina, llamada heterocromatina, genera que se transcriba información genética impropia, que da lugar a que se acumulen en la cromatina híbridos de ADN y ARN, los llamados R-loops, que son letales. La heterocromatina contiene información genética que no transcribe para proteínas, sino que son secuencias repetitivas que la célula quiere tener silenciadas y que están rígidamente controladas.

En los mamíferos hay 7 subtipos de histona H1, H1.1-H1.5, H1.0 y H1x. A pesar de que todos los subtipos de H1 son capaces de unirse y compactar la cromatina, existen evidencias que sugieren su diferenciación funcional de cada uno de los subtipos.



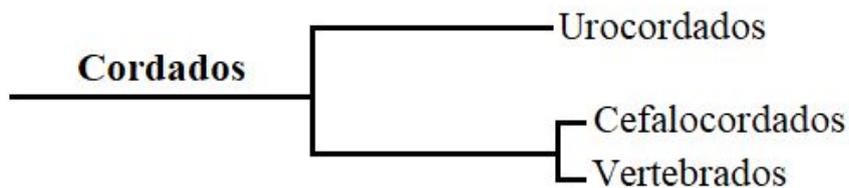
ESPECIES ESCOGIDAS

Las 16 especies que he escogido son del filo de los cordados perteneciente al reino animal (excepto *Caenorhabditis elegans* que es un nematodo y funciona como grupo externo). Se caracterizan por la presencia de una notocorda, tubo neural hueco en posición dorsal, hendiduras branquiales y cola, por lo menos en alguna fase del desarrollo embrionario. Se

conocen casi 65.000 especies actuales, la mayoría pertenecientes al subfilo de los vertebrados, de los que casi la mitad son peces.

Los cordados han demostrado a lo largo de su historia evolutiva notables adaptaciones al medio terrestre, al acuático o anfibio. En los cordados se destaca la capacidad de autorregulación y organización interna. Algunos, como las aves y los mamíferos, pueden elevar y mantener constante la temperatura del cuerpo. Estos y otros factores han añadido complejidad a este grupo de animales, permitiendo un mayor control sobre las reacciones metabólicas y el desarrollo de un complejo sistema nervioso.

Los cordados se subdividen en tres subfilos: urocordados, cefalocordados y vertebrados, tal y como se puede observar en el siguiente cladograma (es un diagrama ramificado usado en la cladística, que es una rama de la biología que define las relaciones evolutivas entre los organismos).



En el siguiente cladograma, se encuentran las supuestas relaciones filogenéticas (algunas no son definitivas o se encuentran en duda) de las especies que he escogido. Este cladograma va a servirme de ayuda para saber si los árboles que haga con las proteínas escogidas, tienen las correctas relaciones entre especies. He escogido especies del filo de los cordados porque es uno de los grupos más diversos y cuentan con los suficientes datos para el estudio.

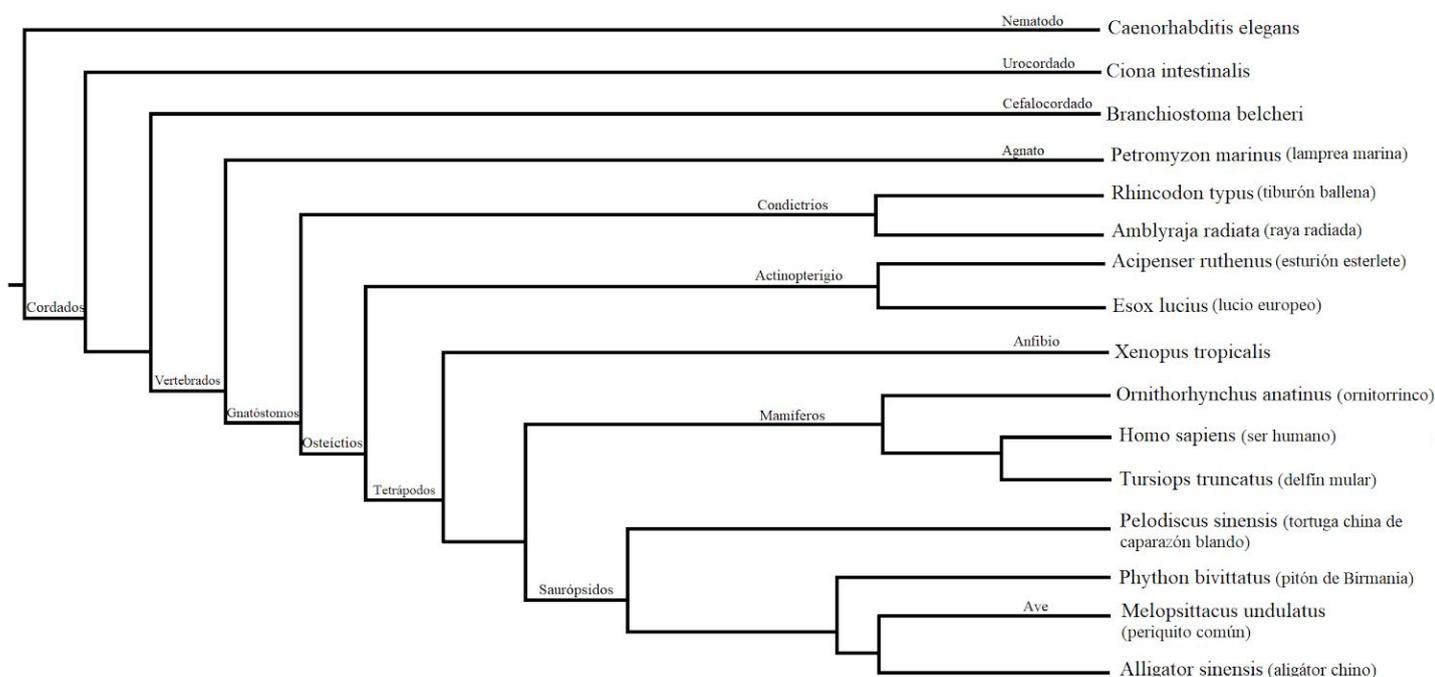


Imagen	Nombre científico	Nombre común	Algunas características
	<i>Ciona intestinalis</i>	Ascidia	Es una especie de urocordado, pariente lejano de los vertebrados. Es un animal marino filtrador del que se ha secuenciado su genoma, y se ha convertido en un modelo de referencia para el estudio de la biología del desarrollo.
	<i>Branchiostoma belcheri</i>	Pez lanceta	<i>Branchiostoma belcheri</i> es uno de los pocos géneros vivos de cefalocordados. Estos pequeños animales son parientes cercanos de los vertebrados. Poseen una boca adyacente a las branquias y son filtradores.
	<i>Petromyzon marinus</i>	Lamprea marina	La lamprea marina es una especie de pez agnato o pez sin mandíbula, que se caracteriza por su cuerpo anguiliforme, la presencia de siete pares de orificios branquiales, un orificio nasal impar y por la ausencia de aletas pares, mandíbulas y escamas.

	Rhincodon typus	Tiburón ballena	El tiburón ballena se encuentra dentro del grupo de los gnatóstomos o peces con mandíbula y es el pez existente más grande del mundo, con aproximadamente 12 metros de longitud. Tiene una alimentación basada en la filtración y habita en aguas cálidas tropicales y subtropicales.
	Amblyraja radiata	Raya radiada	La raya radiada es una especie de pez rajiforme y se distingue, por las grandes espinas que recorren la línea media del dorso desde la nuca hasta la primera aleta dorsal.
	Acipenser ruthenus	Esturión esterlete	El esturión esterlete es una especie de osteíctio o pez óseo. Se reconoce entre los peces por la presencia en su cuerpo de placas dorsales y laterales y su cabeza prolongada en un hocico en punta.
	Esox lucius	Lucio europeo	El lucio europeo es una especie de actinopterio. Tiene un cuerpo largo, de forma casi cilíndrica y una boca con forma de pico de pato que se abre hacia arriba.
	Xenopus tropicalis		Xenopus tropicalis es una especie del clado de los tetrápodos. Es de la clase de los anfibios anuros con la forma típica de una rana pípida, con un cuerpo aplanado.
	Ornithorhynchus anatinus	Ornitorrinco	El ornitorrinco es una especie de mamífero semiacuático. Es una de las cinco especies que perviven en la actualidad del orden de los monotremas (grupo de mamíferos que ponen huevos). Tiene una inusual apariencia, es venenoso, con hocico en forma de pico de pato, cola de castor y patas de nutria, lo que ha hecho que se convierta en un importante sujeto de estudio en el campo de la biología evolutiva.
	Homo sapiens	Ser humano	Homo sapiens es una especie del orden de los primates perteneciente a la familia de los homínidos. Poseen capacidades mentales que les permiten inventar, aprender. Además son animales sociales, conscientes de sí mismos...
	Tursiops truncatus	Delfín mular o nariz de botella	El delfín mular o nariz de botella es una especie de cetáceo. Viven en grupos de hasta 10 o 12 individuos y se sirven de un sistema de ecolocalización para localizar su alimento y, a menudo, cooperan entre ellos para acorralar a sus presas.

	Pelodiscus sinensis	Tortuga china de caparazón blando	La tortuga china de caparazón blando corresponde al clado de los saurópsidos, al que pertenecen los reptiles y aves. El cuello y la cabeza son largos, y el hocico largo y fino le permite respirar mientras está en el fondo de lagos o ríos.
	Python bivittatus	Pitón de Birmania	La pitón de Birmania es la especie de serpiente constrictora más grande de la India y una de las seis serpientes más grandes del mundo. Es una serpiente dócil y un icono en la India. Las pitones birmanas adultas son grandes, pesadas, y su piel tiene un patrón de tipo red.
	Alligator sinensis	Aligátor chino o aligátor del Yangtze	El aligátor chino o aligátor del Yangtze es una especie de saurópsido cocodriliano completamente acorazado y con un color oscuro con bandas más claras a lo largo del cuerpo.
	Melopsittacus undulatus	Periquito común	El periquito común es una especie de ave introducido en muchas zonas del mundo por ser muy común su uso como mascota. Es el ave de jaula más popular del mundo por sus vivos colores, su fácil adaptación y su alegre canto.
	Caenorhabditis elegans		Caenorhabditis elegans es una especie de nematodo que mide aproximadamente 1 mm de longitud. Esta especie se utiliza como modelo para diversos estudios genéticos, muy especialmente en genética del desarrollo, ya que es uno de los organismos animales más simples que cuentan con sistema nervioso y digestivo bien definidos, entre otras características.

Una vez escogidas las 16 secuencias de las tres proteínas, las comparé para ver sus variaciones.

5.2.- RESULTADOS

5.2.1.- Insulina

Primero, busqué las 16 secuencias de aminoácidos en el NCBI. La única secuencia que no encontré fue *Branchiostoma belcheri* (cefalocordado).

Obtuve las siguientes secuencias:

```
>NP_001335969.1 insulin-like 3 protein isoform 2 [Ciona intestinalis]
MSETSRVVSATRPSGGTYRRLCLVTCIVVTCMAIFPGSSIAHGRKRFALSCGSRVNSMRYMCLQEVWTPKLCVFRFKFCGTEINFHPI
CRFTCITESRITSCFPIPAQQH

>XP_032836479.1 insulin [Petromyzon marinus]
MHLCVAGCEPPAWATSLTPQYSPTPVLLQPLHHHPPTMSPRSPRAPSAMLTLLLLLPALLLLLLSLMSPVPAHASALTGAGGTHLCGSHL
VEALYVVCGRDGRFFYTPSKTRRELSDVIRTEADAGEQQGGVVAEGEGELWGGAAAPGAREESPRTRRGIVEQCCHRKCSIYDMENYCN

>XP_020376878.1 insulin [Rhincodon typus]
MVLWNQIVLVAVLFI FSCP KLETQPSQHLCGSQLVDALYFVCGSKGFYYPKTKRGIEQLLGLRGEESAQENEAQQKFLFNKLLGRMMKR
GIVDHCCHNTCSLYDLEGYCNQ

>XP_032894479.1 insulin [Amblyraja radiata]
MMFWSQIVLVCALIIFACPKGLETLP SQHLCGSHLVDALYFVCGPKGFYYLPKAKRGQEEFLELRGGEESTQENEAQQFPFHKLLGRMIK
RGIVEHCCHSTCSLFDLEGYCNQ

>RXM98941.1 Insulin [Acipenser ruthenus]
MALWFRVLSLLALLVLS SPHVSKAAANQHLCGSHLVEALYLVCGERGFYTPNKVKRDMTSPFSFLSGKTASENDVDEFPLKQQGEAKVKR
GIVEQCCHSPCSLYDLENYCN

>XP_010895347.1 insulin [Esox lucius]
MASWLQATSLLVLLSLSPGANSAAAQHLCGSHLVDALYLVCGEKGFYYPNPKRDVDSLIGFLSPKSAQENEAENPFKDQMEMVVRGIVEQC
CHKPCNIFDLQSYCN

>AAI35266.1 ins protein [Xenopus tropicalis]
MALWMQCLPLVLVLLFSTPNTALANQHLCGSHLVEALYLVCGDRGFYYPKIKRDIEQAMVNGPQDNELDGMQLQPQEQYQKMKRGIVEQC
CHSTCSLFLQLESYCN

>AAA59172.1 insulin [Homo sapiens]
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFYTPKTRREAEDLQVGVQLVGGPGAGSLQPLALEGSLQKRGV
VEQCCTSIQSLYQLENYCN

>XP_028915363.1 insulin [Ornithorhynchus anatinus]
MALWVTVLPLLALLALSMPAPGGAFPQNQHLCGSHLVEALYLVCGEKGFYI PRGRRDAEEPVVRGLLQNLNDLEFPQSDLVGKRGIVEEC
CKGVCSMYQLENYCN

>XP_004317908.1 insulin [Tursiops truncatus]
MVLWLRPLPLLALLALWAPAPARAFVNQHLCGSHLVEALYLVCGERGFYTPKARREVEGPQVGAVELGGPGAGGLQPPALEGPPQKRGV
VEQCCTSIQSLYQLENYCN

>XP_006134976.1 insulin [Pelodiscus sinensis]
MALWIRSLPLLALLALS GPPI SHAAANQHLCGSHLVEALYLVCGERGFYSPKARRDLEQPLVNGHLQNEVELPFQQQEFQQAARGIVEQC
CHNTCSLYQLENYCN

>XP_007424294.1 insulin [Python bivittatus]
MTLWIRSLPLLLLLAVSAPTISYAAPNQHLCGSHLVEALYLVCGDRGFYSPRSRRNMEQPLVNGPFQNEVEVPLQEFQKSKRGIVEQCCE
NTCSLYELENYCN
```

>XP_006033708.1 insulin [Alligator sinensis]
MALWIHSLPLLAFLLALSSPSVSYAAANQRLCGSHLVDALYLVCGERGFFYSFKRRDLEQPLVNGPLRNEVEELAFQQQYEEKVVRGIVEQ
CCHNTCSLYQLENYCN

>XP_005149257.1 insulin [Melopsittacus undulatus]
MALWILSLPLLALLLALSGPGTSHAAATQHLGSHLVEALYLVCGERGFFYSFKARRDVEQPLVSGPLHGEVGEVLPFRPEEFQKVRGIVEQ
CCHNTCSLYQLENYCN

>NP_001022529.2 INSulin related [Caenorhabditis elegans]
MRNSFQLILLVLTISYSNSMLFNDVEMDKPSYLSKEAWKMLNELLKEPKHHHHHHHRKGYCGVKAVKKLKQICPDLCSNVDDNLLMEMCS
KNLTDDDILQRCCPE

Después, estas secuencias las copié en el programa de alineamientos múltiples, Clustal Omega. Los guiones que aparecen en el alineamiento se emplean para que se puedan alinear todos los aminoácidos correctamente.

ALINEAMIENTO

>NP_001022529.2 INSulin related [Caenorhabditis elegans]
MR-----NSFQLILLVLTISYSNSMLF-ND-----VEMDKPSYLSKEAWKMLNELLKEPK-----HHHHHHHRKGYCG
VKAVKKLKQICPDLCSNVD-----D-----NLLMEMCSKNLTDDDILQRCCPE-----
-

>NP_001335969.1 insulin-like 3 protein isoform 2 [Ciona intestinalis]
-----MSETSrvvsa---TRPSSGTYRRLCLVTCIVVTCMAIFPGSSIAH---GRKRFALSCG
SRLVNSMRYMCLQEVW--TPKLCsvfRtkfcgTEINFHPI-CRFTcITESR--ITSCFPIPAQH-----
-

>XP_020376878.1 insulin [Rhincodon typus]
-----MVLWNQIVLVA---V--LFIFSCPkg-----LETQPSQHLCG
SQLVDALYFVCGSKGFYYEPKT-KRGIEQLLGLRG-EESA-----QEN--EAQQKFLFNKLLGRMMKRGIVDHCCHNTCSLYDLEGYCN
Q

>XP_032894479.1 insulin [Amblyraja radiata]
-----MMFWSQIVLVC---A--LIIFACPKG-----LETLPSQHLCG
SHLVDALYFVCGPKGFYYLPKA-KRGQEEFLELRGGEEST-----QEN--EAEQQFPFHKLLGRMIKRGIVEHCCHSTCSLFDLEGYCN
Q

>XP_032836479.1 insulin [Petromyzon marinus]
MHLCVAGCEPPAWATSLTPQYSPTPVLLQPLHHHPPTMSPRSPRAPSAMLTLLLLL PAL---LLLLLSLMSVPVPAHASALTGAGGTHLCG
SHLVEALYVCGDRGFFYTPSKTRRELSDVIRTEADAGEQQGGVVAEGEGELWGGAAAPGAREESPRTRRGIVEQCCHRKCSIYDMENYCN
-

>XP_010895347.1 insulin [Esox lucius]
-----MASWLQATSLl---V--LLSLSPGA--N-----SAAQHLCG
SHLVDALYLVCGEKGFYFNPKR---DVDSLIGFLSPKSA--Q-----EN---EVAENPFKD-QMEMVKGIVEQCCHKPCNIFDLQSYCN
-

>RXM98941.1 Insulin [Acipenser ruthenus]
-----MALWFRVLSLl---A--LLVLSSPHVSK-----AAANQHLCG
SHLVEALYLVCGERGFFYTPNKVKRDMTSPFSFLSGKTA--S-----EN---DVDEFPLKQQGEAKVKGIVEQCCHSPCSLYDLENYCN
-

>AAA59172.1 insulin [Homo sapiens]
-----MALWMRLLPll---A--LLALWGPDPAA-----AFVNQHLCG
SHLVEALYLVCGERGFFYTPKT-RREAEDLQVGQV-ELG--G-----GP---GAGSLQPLALEGSLQKRGIVEQCCTSIcslyQLENYCN
-

```

>XP_004317908.1 insulin [Tursiops truncatus]
-----MVLWLRRLGPLL---A--LLALWAPAPAR-----AFVNQHLCG
SHLVEALYLVCGERGFFYTPKA-RRVEEGPQVGAV-ELG--G-----GP---GAGGLQPPALEGPPQKRGIVEQCCTSICSLYQLENYCN
-

>AAI35266.1 ins protein [Xenopus tropicalis]
-----MALWMQCLPLV---L--VLLFSTPNT-E-----ALANQHLCG
SHLVEALYLVCGRGDRGFFYYPKI-KRDIEQAMVNGP-QDN-----ELDGMQLQPQEQYQKMKRGIVEQCCHSTCSLQLESYCN
-

>XP_028915363.1 insulin [Ornithorhynchus anatinus]
-----MALWVTVLPLL---A--LLALSMPAPGG-----AFPNQHLCCG
SHLVEALYLVCGERGKGFYIPRG-RRDAEEPVVRGL-LQN-----GLNDLEFPS-DLQVKGKRGIVEECKGVCSMYQLENYCN
-

>XP_007424294.1 insulin [Python bivittatus]
-----MTLWIRSLPLL---L--LLAVSAPTISY-----AAPNQHLCCG
SHLVEALYLVCGRGDRGFFYSPRS-RRNMEQPLVNGP-FQN-----EVE---VPLQEFQKSKRGIVEQCCHNTCSLYELENYCN
-

>XP_006033708.1 insulin [Alligator sinensis]
-----MALWIHSLPLL---A--FLALSSPSVSY-----AAANQRLCCG
SHLVDALYLVCGERGFFYSPKG-RRDLEQPLVNGP-LRN-----EVEELAFQQEQYQKVKRGIVEQCCHNTCSLYQLENYCN
-

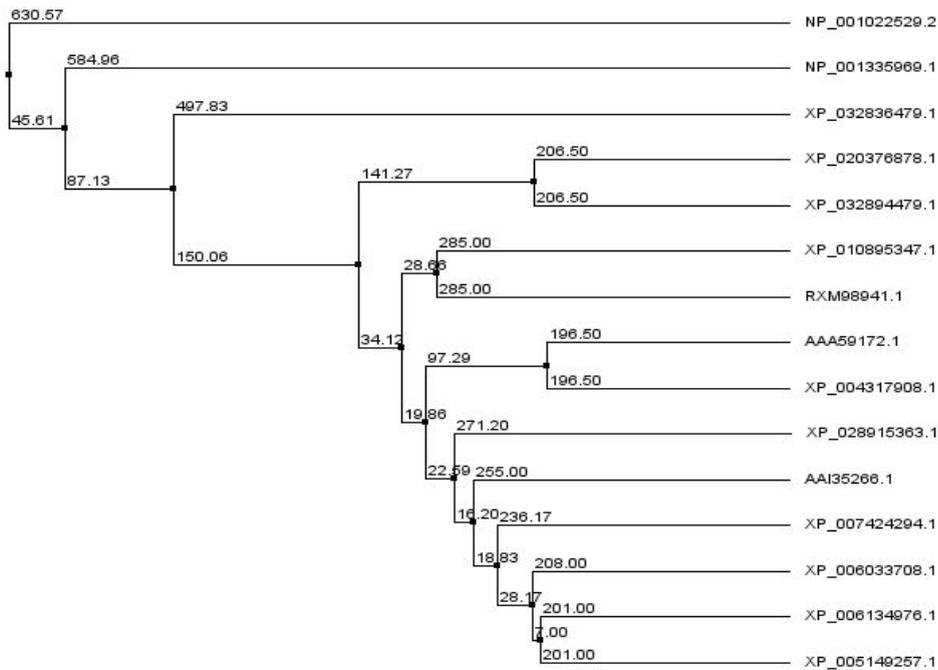
>XP_006134976.1 insulin [Pelodiscus sinensis]
-----MALWIRSLPLL---A--LLALSGPPISH-----AAANQHLCG
SHLVEALYLVCGERGFFYSPKA-RRDLEQPLVNGH-LQN-----EV-ELPFQQEQFQAKRGIVEQCCHNTCSLYQLENYCN
-

>XP_005149257.1 insulin [Melopsittacus undulatus]
-----MALWILSLPLL---A--LLALSGPGTSH-----AAATQHLCG
SHLVEALYLVCGERGFFYSPKA-RRDVEQPLVSGP-LHG-----EVGELPFRPEEFQKVKRGIVEQCCHNTCSLYQLENYCN
-

```

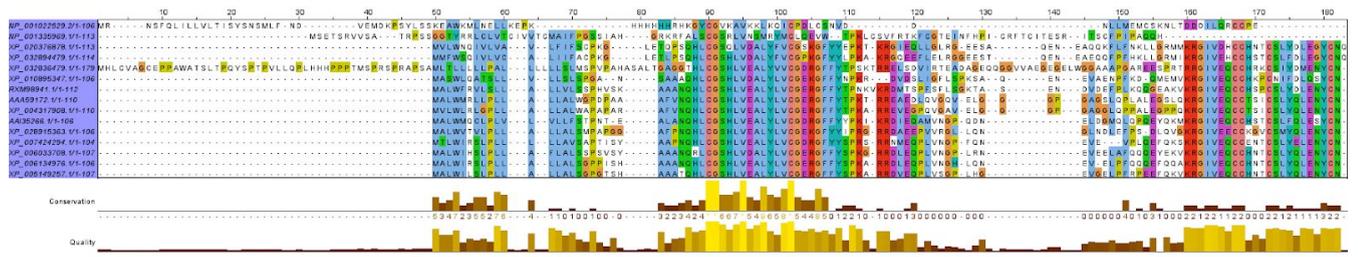
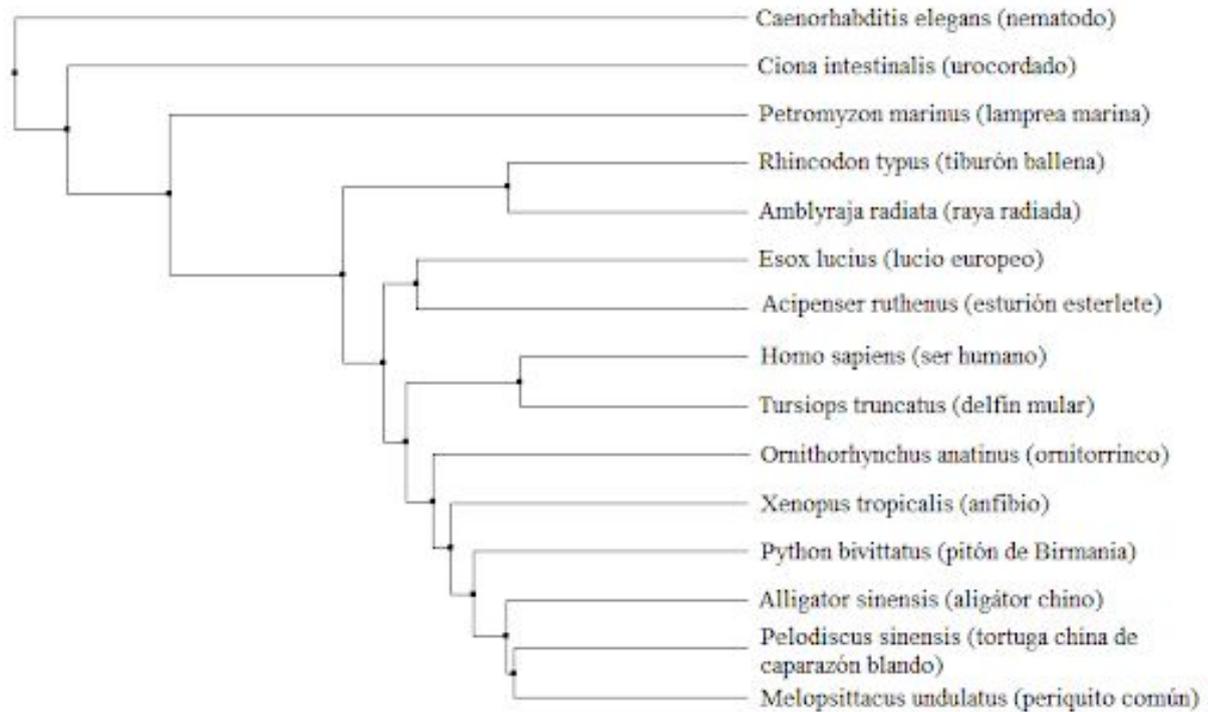
Finalmente, para ver las diferencias entre unas secuencias y otras y construir el cladograma, utilicé el programa Jalview.

El resultado que obtuve fue el siguiente cladograma. Los finales de cada rama muestran el código de letras, seguido de unos números, que corresponden a una especie en concreto. Para una mejor comprensión, he cambiado cada código por la especie a la que corresponde.



Como se puede observar, el resultado obtenido es muy similar al cladograma inicial que mostraba cómo tenían que ser las relaciones filogenéticas entre las especies. Las tres únicas diferencias entre este cladograma y el cladograma “base” son que *Xenopus tropicalis* (anfibio) tiene un ancestro común con los saurópsidos (reptiles y aves) y tendría que estar situado en una rama anterior a los mamíferos, ya que mamíferos y saurópsidos son amniotas, pero los anfibios no.

Otra pequeña divergencia es que *Ornithorhynchus anatinus* (ornitorrinco), tendría que estar más relacionado con los mamíferos *Homo sapiens* (ser humano) y *Tursiops truncatus* (delfín mular). Por último, señalar que *Alligator sinensis* (aligátor chino) y *Melopsittacus undulatus* (periquito común) tendrían que estar más próximos en las ramas evolutivas. Sin embargo, se puede observar que la insulina de *Pelodiscus sinensis* (tortuga china de caparazón blando) tiene más relación con el ave.



En el caso de la insulina, se puede observar, que la parte más conservada es la central porque puede ser una estructura fundamental para su función.

5.2.2.- Citocromo c

Al igual que en la insulina, primero seleccioné las secuencias de las especies. En el citocromo c, no encontré dos especies en la base de datos, ni *Petromyzon marinus* (lamprea marina), ni *Amblyraja radiata* (raya radiada).

```
>XP_002130585.1 cytochrome c, somatic [Ciona intestinalis]
MGKKAAGDVEKGGKIFIQCKQCHTYSDCTNKTGPTLHGLFGRQSQVAGFDYTDQNRDKGVIWGEDTLDVYLTNPAKIYPGTMVFAG
IKKKGERADLIAFLKDATK

>BAF47392.1 cytochrome c [Branchiostoma belcheri]
MADAVAKGNPDGKGVFVQKCAQCHTVEKGGKHTGPNLNLGFRKRTGQASGFSYTDANKNKGITWAEDTLWVYLENPKKIYPGTMVFAG
LKKKGERADLISYLKESTSS
```

```

>XP_020365536.1 cytochrome c [Rhincodon typus]
MGDVEKGGKVFVQKCAQCHTVECGGKHKHTGPNLCGLFGRKGTQAEGFSYTDANKNKGITWNEETLMVYLENPKKYIPGTKMIFAGLKKKSE
REDLIAYLKSATAK

>XP_033856686.1 cytochrome c [Acipenser ruthenus]
MGDVAKGKKVVFVQKCAQCHTVEGGKHKHTGPNLHGLFGRKGTQAEGFSYTEGNKNKGITWSEDTLMEYLENPKKYIPGTKMLFSGIKKKSE
RQDVIAYLKESTSA

>NP_001291051.1 cytochrome c [Esox lucius]
MGDIAKGGKAFVQKCAQCHTVEEGGKHKVGPNLWGLFGRKGTQAPGFSYTDANKSKGITWSEDTLMEYLENPKKYIPGTKMIFAGIKKKGE
RQDLIAYLKSATS

>NP_001008176.1 cytochrome c, somatic [Xenopus tropicalis]
MGDVEKGGKIFVQKCAQCHTVEKTKGKHKHTGPNLWGLFGRKGTQAPGFSYTDANKSKGIVWGEDTLFEYLENPKKYIPGTKMIFAGIKKKNE
RADLIAYLKKSTSE

>AAA35732.1 cytochrome c [Homo sapiens]
MGDVEKGGKIFIMKCSQCHTVEKGGKHKHTGPNLHGLFGRKGTQAPGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEE
RADLIAYLKKATNE

>XP_028926987.1 cytochrome c [Ornithorhynchus anatinus]
MGDFEKGKKIFVQKCAQCHTVEKGGKHKHTGPNLNGLFGRKGTQAPGFSYTDANKSKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKGE
RADLVAYLKKATKE

>XP_033718988.1 cytochrome c [Tursiops truncatus]
MGDIEKGGKIFVQKCAQCHTVEKGGKHKHTGPNLHGLFGRKGTQAVGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKGE
RADLIAYLKKATNE

>XP_006138438.1 cytochrome c [Pelodiscus sinensis]
MGDIEKGGKIFVQKCAQCHTVEKGGKHKHTGPNLNGLIGRKTQQAAGFSYTDANKNKGITWNEDTLMEYLENPKKYIPGTKMIFVGIKKKAE
RADLIAYLKDATSK

>XP_007420882.1 cytochrome c [Python bivittatus]
MGDVEKGGKIFTMKCGTCHTVEEGGKHKHTGPNLHGLFGRKGTQAVGFSYTEANKNKGITWGEDTLMEYLENPKKYIPGTKMVFTGLKSKKE
RTDLIAYLKEATSK

>XP_006027096.1 cytochrome c [Alligator sinensis]
MGDVEKGGKIFVQKCAQCHTVEKGGKHKHTGPNLHGLIGRKTQQAAGFSYTEANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKPE
RADLIAYLKEATAN

>tr|S4RXV7|S4RXV7_PETMA Cytochrome c domain-containing protein OS=Petromyzon marinus
OX=7757 PE=3 SV=1
MGDVEKGGKVFVQKCSQCHSVEKGGKHKHTGPNLSGIFGRQTQQAAGFSYTDANRDKGIVWNEDTLFLYLENPKKYIPGTKMIFAGIKKEGE
RKDLIAYLKQSTAE

>XP_005154206.1 cytochrome c [Melopsittacus undulatus]
MGDIEKGGKIFVQKCSQCHTVEKGGKHKHTGPNLNGIFGRKGTQAEGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKSE
RADLIAYLKSATK

>CCD68708.1 Cytochrome c 2.1 [Caenorhabditis elegans]
MSDIPAGDYEGKGVYKQRCLQCHVVDSTATKTGPTLHGVI GRTSGTVSGFDYSAANKNKGIVVWRETFLFEYLLNPKKYIPGTKMVFAGLK
KADERADLIKYIEVESAKSL

```

A continuación, realicé el alineamiento múltiple con Clustal Omega y el árbol con Jalview.

En este alineamiento del citocromo se puede observar que hay muy pocos guiones . Esto puede ser debido a que el citocromo ha cambiado poco entre estas especies.

ALINEAMIENTO:

>CCD68708.1 Cytochrome c 2.1 [Caenorhabditis elegans]
 -MSDIPAGDYEKGGKVKYQRCQLQCHVVDST-ATKTGPTLHGVIGRTSGTVSGFDYSAANKNKGVVWTRETLFEYLLNPKKYIPGTKMVFAG
 LKKADERADLIKYIEVESAKSL

>XP_002130585.1 cytochrome c, somatic [Ciona intestinalis]
 MGKAAAAGDVEKGGKIFIQKCKQCHTYTSDCTNKTGPTLHGLFGRQSGQVAGFDYTDQNRDKGVIWGEDTLDVYLTNPAKYIPGTKMVFAG
 IKKKGERADLIAFLKDATK---

>BAF47392.1 cytochrome c [Branchiostoma belcheri]
 MADAVAKGNPDGKKVVFVQKCAQCHTVEKGGKHKGTGPNLNLGFRKTGQAGFSYTDANKNKGITWAEDTLWVYLENPKKYIPGTKMVFAG
 LKKKGERADLISYLKESTSS--

>XP_033856686.1 cytochrome c [Acipenser ruthenus]
 -----MGDVAKGKKVVFVQKCSQCHTVENGGKHKGTGPNLHGLFGRKTGQAEFGFSYTEGKNKGITWSEDTLMEYLENPKKYIPGTKMLFSG
 IKKKSERQDVIAYLKESTSA--

>tr|S4RXV7|S4RXV7_PETMA Cytochrome c domain-containing protein OS=Petromyzon marinus
 OX=7757 PE=3 SV=1
 -----MGDVEKGGKVFVQKCSQCHSVEKGGKHKGTGPNLSGIFGRQTGQAPGFSYTDANRDKGIVWNEDTLFLYLENPKKYIPGTKMIFAG
 IKKEGERKDLIAYLKQSTAE--

>XP_007420882.1 cytochrome c [Python bivittatus]
 -----MGDVEKGGKIFTMKCGTCHTVEEGGKHKGTGPNLHGLFGRKTGQAVGFSYTEANKNKGITWGEDTLMEYLENPKKYIPGTKMVFTG
 LKSKKERTDLIAYLKEATSK--

>XP_020365536.1 cytochrome c [Rhincodon typus]
 -----MGDVEKGGKVFVQKCAQCHTVECGGKHKGTGPNLNLGFRKTGQAEFGFSYTDANKNKGITWNEETLMVYLENPKKYIPGTKMIFAG
 LKKKSEREDLIAYLKSATAK--

>NP_001291051.1 cytochrome c [Esox lucius]
 -----MGDIAKGGKAFVQKCAQCHTVEEGGKHKVGPNLWGLFGRKTGQAPGFSYTDANKSKGITWSEDTLMEYLENPKKYIPGTKMIFAG
 IKKKGERQDLIAYLKSATS---

>AAA35732.1 cytochrome c [Homo sapiens]
 -----MGDVEKGGKIFIMKCSQCHTVEKGGKHKGTGPNLHGLFGRKTGQAPGFSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVG
 IKKKEERADLIAYLKKATNE--

>NP_001008176.1 cytochrome c, somatic [Xenopus tropicalis]
 -----MGDVEKGGKIFVQKCAQCHTVEKTGKHKGTGPNLWGLFGRKTGQAPGFSYTDANKSKGIVWGEDTLFEYLENPKKYIPGTKMIFAG
 IKKKNERADLIAYLKKSTSE--

>XP_006138438.1 cytochrome c [Pelodiscus sinensis]
 -----MGDIEKGGKIFVQKCAQCHTVEKGGKHKGTGPNLNLGFRKTGQAAFGFSYTDANKNKGITWNEDTLMEYLENPKKYIPGTKMIFVG
 IKKKAERADLIAYLKDATSK--

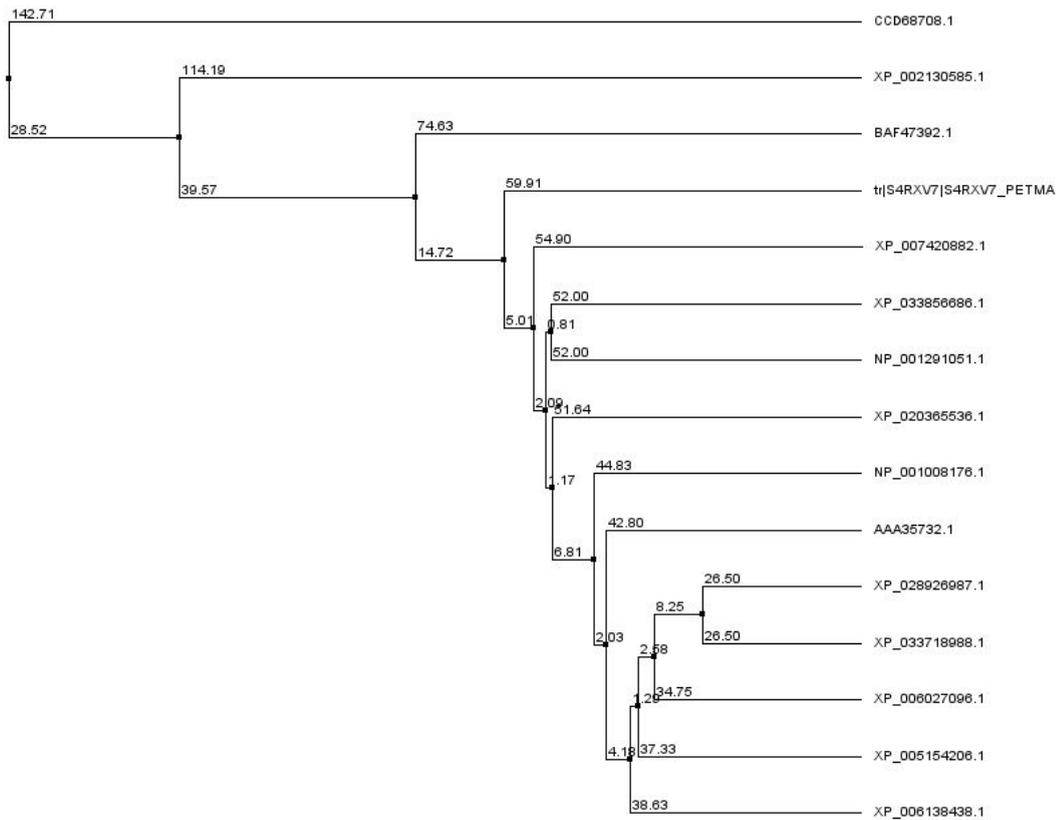
>XP_005154206.1 cytochrome c [Melopsittacus undulatus]
 -----MGDIEKGGKIFVQKCSQCHTVEKGGKHKGTGPNLNLGFRKTGQAEFGFSYTDANKNKGITWGEDTLMEYLENPKKYIPGTKMIFAG
 IKKKSERADLIAYLKSATK---

>XP_006027096.1 cytochrome c [Alligator sinensis]
 -----MGDVEKGGKIFVQKCAQCHTVEKGGKHKGTGPNLHGLGFRKTGQAPGFSYTEANKNKGITWGEETLMEYLENPKKYIPGTKMIFAG
 IKKKPERADLIAYLKEATAN--

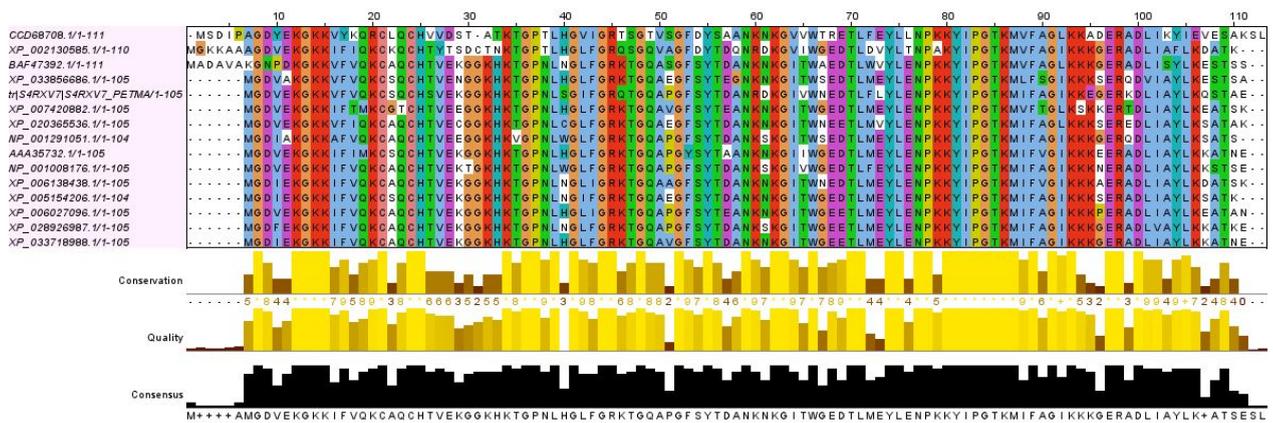
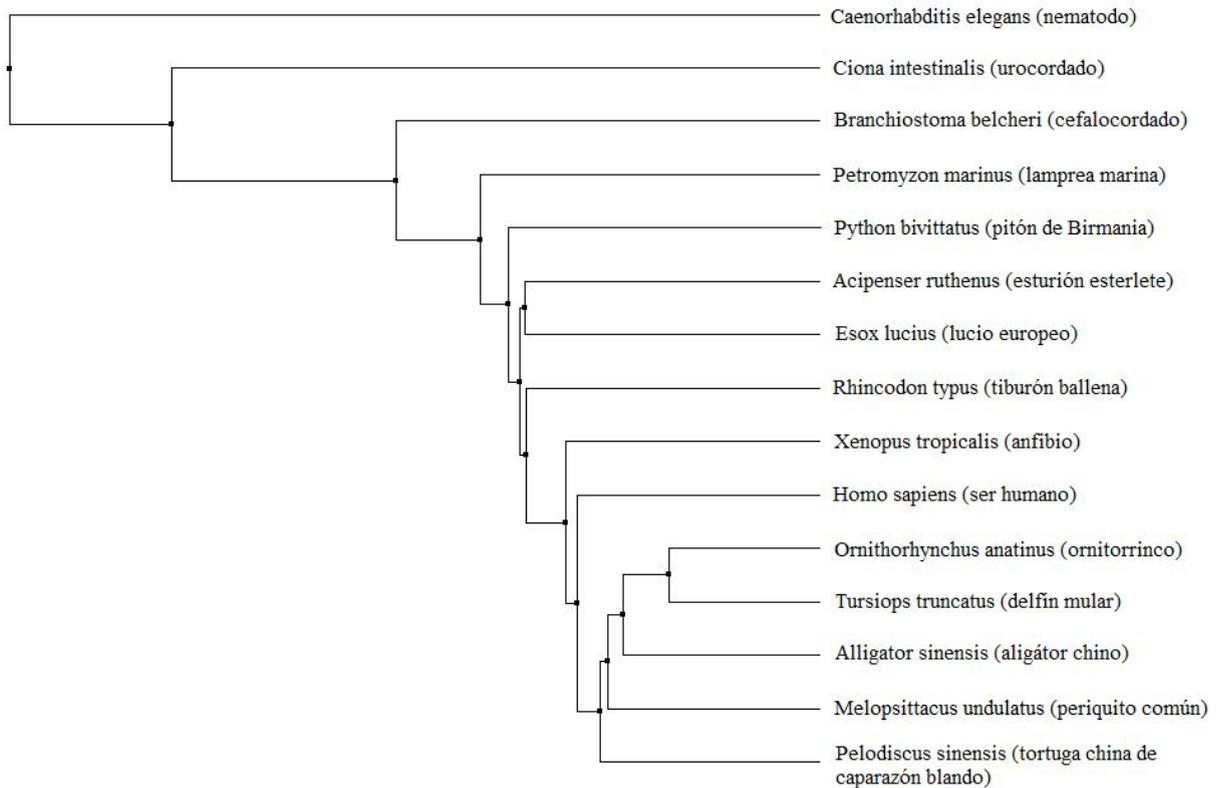
>XP_028926987.1 cytochrome c [Ornithorhynchus anatinus]
 -----MGDFEKGKKIFVQKCAQCHTVEKGGKHKGTGPNLNLGFRKTGQAPGFSYTDANKSKGITWGEETLMEYLENPKKYIPGTKMIFAG
 IKKKGERADLVAYLKKATKE--

>XP_033718988.1 cytochrome c [Tursiops truncatus]
 -----MGDIEKGGKIFVQKCAQCHTVEKGGKHKGTGPNLHGLFGRKTGQAVGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFAG
 IKKKGERADLIAYLKKATNE--

Al igual que en la anterior proteína, he cambiado cada código de cada secuencia de las especies para mejorar la comprensión del cladograma.



Las grandes diferencias de este cladograma, comparado con el cladograma “base” son que *Python bivittatus* (pitón de Birmania) tendría que estar con los demás reptiles, pero en este caso, se puede observar que tiene un antepasado común con los peces. Este cladograma difiere del anterior en que *Rhincodon typus* (tiburón ballena) tendría que estar más relacionado con el grupo de los peces. Sin embargo, según este cladograma, tiene un antecesor de los tetrápodos. Una gran diferencia es que dos de las tres especies de mamíferos están incluidos dentro del grupo de los reptiles.



En este gráfico se puede observar que el grado de similitud es muy alto entre todas las secuencias, lo que hace que un cambio en un aminoácido, puede llevar a que las especies que están más emparentadas aparezcan distantes en el cladograma.

5.2.3.- Histona H1

Al igual que en la insulina y el citocromo c, primero seleccioné las secuencias de las especies.

En la histona H1, no encontré en la base de datos la especie *Amblyraja radiata* (raya radiada). Además, no todas las histonas H1 son del mismo tipo, ya que, existen hasta siete subtipos. En el NCBI, he tenido problemas para encontrar el mismo subtipo en todas las especies, por lo que he tenido que escoger diferentes subtipos, aunque las diferencias entre ellos son mínimas.

```
>XP_002121210.1 histone H1.0 [Ciona intestinalis]
MVAKVDTDKKKVAQKKEHPTYAEMIVAAIKALKDRKGSRQAIIVKYIKANYKVGDNAAATQCKHSLKRLVTSEKIVNTVGVGASGRFKVPQA
AVPTTTKKVAKKTEKEAKPAQKKTAAKKKTAAKKSPKKQAKKSPAKKPAQKKAAPKVKVTKSPKKPTKKPAAKKSAPAKKAAKPAAKK
SAKKTTKK
```

```
>tr|S4RD72|S4RD72_PETMA H15 domain-containing protein OS=Petromyzon marinus OX=7757 PE=3
SV=1
MVEVAPAPQPPAAATSALKSPKKKRTAPKAKKVGPSVSEILIKALASSKERNVSLPALKKSLAAGGYDVEKNARLKLAIKKLVAKDAVAQ
VKGSGASGSFKLNKKEKAAAKAPKKSSTAKPAAKPPAAAKVKKPAAAKKTAPKKSPPKAKKSAAAKSPKKVKSAPKPKAAKAVAKPKKS
ALKKRAAAVRVMDPDPFASKSAGPAKVPAPKK
```

```
>XP_020387308.1 histone H1-like [Rhincodon typus]
MSDSAAAETAPPASASTKPKIPKPKKKAAXPRKKPAGPGLGERIVKIVGESSXRKGTSLAAIKKALQRXEVNVEKQNAQIRXAIKRCCLAKGS
LVXVKGQGVSGSFKLPKNPVKAKVGGKAKPAASAKKPAVKKTTAKKVTAKKSPAKKAAVKKAAATPKKAVKKAAPKKTTPVKVKKTKSPKA
AKPVFKSGKPKAKPKAKVTTKAAKK
```

```
>RXM99081.1 Histone H1 [Acipenser ruthenus]
MAEVAPVAPTEAAPVPKAVKAGKKKTATKPKKAGPSVSDLIVKAVSASKERSGVSLAALKKALAAGGYDVEKNNSRVNTAVKSLVSKETL
LQTKGVGASGSFRLNKKVAEAKDKVAKKAVPKPKKPAVKKAAAPKAPAKKPAKPKATKPKATKSPKKAKKPAKKAATKSPKKAKKPAKPKPTK
SPKKVAKPQAAKTKRAAPKRR
```

```
>XP_028974776.1 histone H1 [Esox lucius]
MAEVAPAPAAAAPAKAPKKAAPKPKKAGPSVSELIVKAVSASKERSGVSLAALKKSLAAGGYDVEKNNSRVKLVKSLVTKGTLVQTKGT
GASGSFKLNKATETTKPKAKKAAAPKAKKVAAKKPAKAAKPKKAAAKKPAKAAKSPKKAKKPAKPKAAKSPKKVKKPAAAAPKAAKSPK
KATKAAKPKAAKPKAAKPKKAAAPKPK
```

```
>XP_004919170.1 histone H1B [Xenopus tropicalis]
MTEVAPAPAPPPAEPAKAAKQPKKGGAKAKKPAKPSAAELIVKAVSASKERSGVSLAALKKALAAGGYDVERNNSRLKLALALVTKG
TLTQVKGSGASGSFKLNKPLESKEKAAKPKPAAPKVKKPAKAAKPKKSPKKPKKVSAAAKSPKKVKKPAKAAKSPKKPKAAKPKKVAKS
PAKKAAPKAAKSPAKKPTKPKATKSPAKAKAAKPKAAKAKKAAAPK
```

```
>AAA63186.1 histone H1 [Homo sapiens]
MSETAPLAPTIPAPAEKTPVKKKAKKAGATAGRRKASGPPVSELITKAVAASKERSGVSLAALKKALAAGGYDVEKNNSRIKGLKSLVSK
GTLVQTKGTGASGSFKLNKKAASGEGKPKAKKAGAARKPRKPAKAAKPKKVAAGATPKKSIKKTTPKVKKPPATAAGTKKVAKSAAKVKTPQ
PKKAAKSPAKAKAPKAAKPKSGKPKVTKAKKAAAPKPK
```

```
>XP_016082469.2 histone H1.4-like [Ornithorhynchus anatinus]
MSETAPAPAVPAPVEKSPAKKAKKPAKGGARRKAAGPSVSELITKAVSVSKERNVSLAALKKALAAGGYDVEKNNSRIKGLKSLVAKG
TLVQTKGTGASGSFKLNKKAAGSEKAKPKAKKPAKAAKPKKPAKAAKPKKPAVSGARKSVKKTPKKAKKPAVAAKAAKSPKKTAVK
PKKAAKSPAKPKVVKPAAKPKVAKPKTKSKKAAASKKK
```

```
>XP_019800694.1 histone H1.2 [Tursiops truncatus]
MSETAPAPAAAAPAEKTPVKKKAAKPAKARRKASGPPVSELITKAVAASKERSGVSLAALKKALAAGGYDVEKNNSRIKGLRSLVSKG
TLVQTKGTGASGSFKLNKRGATGEAKPKAKKAGAARKPKAAKPKKATGAATPKKAAKTPKAKKPAATVTKKVAKSPRKAKAAK
KKAASAAKAAKTKAAKPKVAKPKKAAAPKPK
```

```
>XP_006024244.1 histone H1.01-like [Alligator sinensis]
MSETAPVAAPAPAPAAKQAKKTKKAAAGGSKARKPSGVSVELITQAVSASKERKGLSLAALKKALAAGGYDVEKSNRXXXXXXXXXX
TTSRIKVLKSLVSKGILVQTKGTGASGSFKLGGKPKGEVKEKAPKPKKAVAAKPKKAAKPKAAAVKSPKAKKPAKAAKKSAAKSPR
KAKAAKPKRAAKSPAKAVKPKSPKAAKAKKAAAPKPK
```

```
>XP_005150318.2 histone H1 [Melopsittacus undulatus]
```

MSETAPVAAPAVSAPGAKAATKKPKKAAGGSKARKPAGPSVTEELITKAVSASKERKGLSLAALKKALAAGGYDVEKNNSRIKLGKSLVSK
GTLVQTKGTGASGSFKLNKRPGETKERATKKKPAAKTKKPAAKKPASAACKPKKAAVKKSPKKAKKPVVAAAKKATKSPKKAAGRPKK
AAKSPAKAKAVKPKAAKPKAAKPKAAKAKKAAPKKK

>XP_007445091.1 histone H1-like [Python bivittatus]
MSETAPEAPAAAAAPAAKAPAKKAKKPAGAAKPRKASGPSVTELLTKAVAASKERNVSLAALKKALAAGGYDVEKNNSRIKLGKSLVSKG
TLVQTKGTGASGSFKLNKKQQEGKDKVVKKPAVKSKKPAAKKPASAACKPKKPAATKSPKKAVKKPAAGAKKAAKSPKKAKPAKPTKAAK
SPAKAKTKAAKPKTPKAKVAKPKKTAAKK

>XP_006125951.1 histone H1 [Pelodiscus sinensis]
MSETAPVATPAVSAPGKTTSAKKPKKAAGGSKARKSSGPSVTEELITKAVSASKERKGLSLAALKKAFAGAAGGYDVEKNNSRIKLGKSLVSK
GVLVQTKGTGASGSFKLNKPKGETKEKAPKKKPAAKPKKPAAKKPASAACKPKKAAVKKSPKKAKKPAASAACKVAKSPKKAKPAKPKKA
AKSPAKAKAVKPKAAKPKPTKPKAGAKKAAPKKK

>CAA37372.1 histone H1 [Caenorhabditis elegans]
MSDSAVVAAVEPKVPKAKAAKAAKPTKVAKAKAPVAHPYINTIKEAIKQLKDRKGASKQAILKFISQNYKLGDNVIQINAHHRQALKRG
VTSKALVQARGSGANGFRVPEKAAAACKPAAKPKPAAKPKPAAKPKATGEEKAKKPAAPKPKAATGDKKVKKAKSPKKVAKPAKPKVAK
SPAKAAPKKIAKPAAKKAAKPAAKA

A continuación, realicé el alineamiento múltiple con Clustal Omega y el árbol con Jalview.

ALINEAMIENTO

>XP_020387308.1 histone H1-like [Rhincodon typus]
MSDSAAAET-AP---PASASTPKPKIPKKKKAAXPRKKPAGPGLGERIVKIVGESSXRKGTSLAAIKKALQRXEVNVEKQNA-----
----QIRXAIKRCLAKGSLVXVKGQVSGSFKLPKNPVKAKVKG--KAKPAASAKKPAVKKTTAKKVTAKKSPAKKAAVKAATPKKA
VKKAAPK--KKTVPVK--VKKTKSPKAA-----KPV-----PKSGKPKAK-----PKAKVTKKAACK--
--

>tr|S4RD72|S4RD72_PETMA H15 domain-containing protein OS=Petromyzon marinus OX=7757 PE=3
SV=1
MVETAPAQPAAA-TSALKSPKK-----RTAPKAKKVGPSVSELILKALASSKERNVSLPALKKS LAAGGYDVEKNNA-----
----RLKLAIKKLVAKDAVAQVKGSGASGSFKLNKKEKAAKAK-----PKSSTAKPAAKPPAAK-----KVKKPAA--AKKTAPK
KSPKK--AKKSA-----AAKSPKVKSAVKPKKAAKAVAKPKKSALKKRAAAVRVDMPPDPP-ASKSAGPAKV-----PAPKK--

>RXM99081.1 Histone H1 [Acipenser ruthenus]
MAEVAPVA-P----TEAAPAPVKAVKAGKKTATKPKKAGPSVSDLIVKAVSASKERSGVSLAALKKALAAGGYDVEKNNS-----
----RVNTAVKSLVSKETLLQTKGVGASGSFRLNKKVAEAKDKVAKKAVP--KPKKPAVKAAPK--APA--KKP-----ATKPKPAT
KSPKK--AKKPAK---KAATKSPKAKKPAK--KKPTKS-----PKKVAKPQAATKRAAPKRR-

>XP_006024244.1 histone H1.01-like [Alligator sinensis]
MSETAPVAAPAAP-APAAKAQAKKTKKAAAGGSKARKPSGPSVTEELITQAVSASKERKGLSLAALKKALAAGGYDVEKSNRXXXXXXXXX
XTTSRIKVLKSLVSKGILVQTKGTGASGSFKLGKKPGEVKEKAPKKKAVAAAKPKKAAAKPKA-----AAVK
KSPKK--AKKPAAAAAKKSAKSPRKA-KAAKPKRAAKSPAKAKAV-----KPKPSKPKAAKAKKAAPKKK--

>XP_004919170.1 histone H1B [Xenopus tropicalis]
MTETAETAPAPP--PAEPAAAKKQPKKGGAKAKKPAGPSAAELIVKAVSASKERSGVSLAALKKALAAGGYDVERNNS-----
----RLKLALVTKGTLTQVKGSGASGSFKLNKPLESKEKAAK--KPAA----PKVKKPAAAKKAPKSPKPKKVA-----AA
KSPKK--VKKPAK-----AKSPKPK--KAAKPKKVAKSPAKKAAKPKAAKS---PAKKPTKPKATKSPAKAKAAKPKAAKAKKAAPKK--

>XP_028974776.1 histone H1 [Esox lucius]
MAEVAPA-----PAAAAPAKAPK---KAAAKPKKAGPSVGEELIVKAVSASKERSGVSLAALKKSLAAGGYDVEKNNS-----
----RVKLAVKSLVTKGTLVQTKGTGASGSFKLNKATETKPKAK--AAPKAKKVAACKPA-AA-----KKPKKAAA-KKPAAAA
KSPKK--AKKPATP---KKAASPKKVKKPA---AAAKKAAKSPK-----ATKAAKPKAAKPKAAKPKKAAAPKKK--

>XP_016082469.2 histone H1.4-like [Ornithorhynchus anatinus]

```

MSETAPA-APAVP-APVEKSPAKKKAK-KPAGGARRKAAGPSVSEELITKAVSVSKERNGVSLAALKKALAAAGYDVEKNNS-----
----RIKLGKLSLVAKGTLVQTKGTGASGSFKLNKKAAGSEGKAKPKAKKPAAA---AKPKKPASAA-----KKPKKPAV-SGARKSVK
KTPKK--AKKPAAVA-AKKAASKPKKT-KAVKPKKAAKSPAKPKVV-----KPKAAKPKVAKPKTTKSKKAAASKKK

>AAA63186.1 histone H1 [Homo sapiens]
MSETAPL-APTIP-APAECTPVKKKAKKAGATAGKRKASGPPVSEELITKAVAASKERSGVSLAALKKALAAAGYDVEKNNS-----
----RIKLGKLSLVSKGTLVQTKGTGASGSFKLNKKAASGEGKPK--AKK-AGA---AKPRKPAGAA-----KKPKKVA GAATPKKSIK
KTPKK--VKKPATAAGTKKVAKSARKV-KTPQPKKAAKSPAKAKAP-----KPKAAKPKSGKPKVTAKKAAPKKK-

>XP_019800694.1 histone H1.2 [Tursiops truncatus]
MSETAPA-APAAA-PPAEKTPVKKKAAKPP-AGARRKASGPPVSEELITKAVAASKERSGVSLAALKKALAAAGYDVEKNNS-----
----RIKLGKLSLVSKGTLVQTKGTGASGSFKLNKRGATGEAKPK--AKK-AGA---AKPKKAAGAA-----KKPKKATGAATPKKAAK
KTPKK--AKKPAATVTKKVAKSPRKA-KAAKPKKAAKSAAKA-----AKTKAAKPKVAKPKKAAPKKK-

>XP_007445091.1 histone H1-like [Python bivittatus]
MSETAPE-APAAA-APAAKAPAKKA-KKPAGAAKPRKASGSPVTELLTKAVAASKERNGVSLAALKKALAAAGYDVEKNNS-----
----RIKLGKLSLVGKTLVQTKGTGASGSFKLNKKQEGKDKVV---KKKPAVKSKKPAAKKPASAA-----KKPKKPAA-----TK
SPKKA--VKKP-AA-GAKKAASKPKKA-KPAKPTKAAKSPAKAKTK-----AA---KPKTPKAKVAKPKKTAAKKK-

>XP_005150318.2 histone H1 [Melopsittacus undulatus]
MSETAPVAAPAVS-APGAKAATKPP-KKAAGGSKARKPAGSPVTEELITKAVSASKERKGLSLAALKKALAAAGGYDVEKNNS-----
----RIKLGKLSLVNKGTLVQTKGTGASGSFKLNKRPGETKERAT---KKKPAAKTKKPAAKKPASAA-----KKPKKAAA-----VK
KSPKK--AKKPVVA-AAKATKSPKKAAKAGRPKKAAKSPAKAKAV-----KPKAAKPKAAKPKAAKAKKAAPKKK-

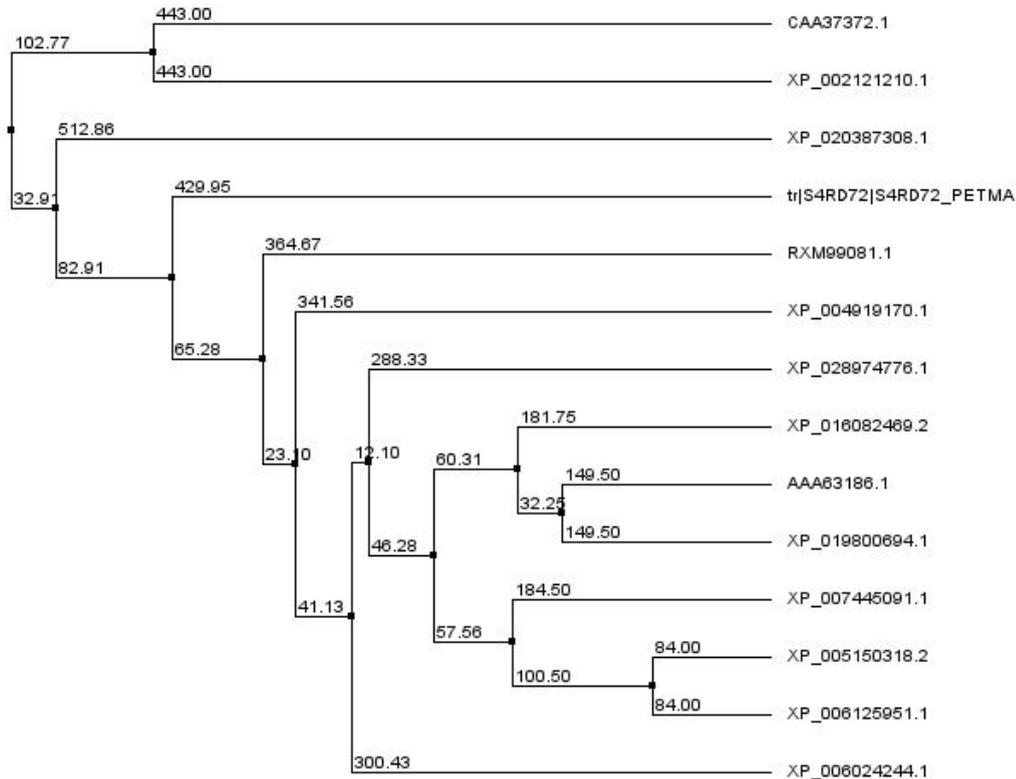
>XP_006125951.1 histone H1 [Pelodiscus sinensis]
MSETAPVATPAVS-APGTKTSAKPP-KKAAGGSKARKSSGSPVTEELITKAVSASKERKGLSLAALKKAFAGGYDVEKNNS-----
----RIKLGKLSLVSKGVLVQTKGTGASGSFKLNKPKGETKEKAP---KKKPAAKPKKPAAKKPASAA-----KKPKKAAA-----VK
KSPKK--AKKPAAS-AAKVAKSPKKA-KPAKPKKAAKSPAKAKAV-----KPKAAKPKPTKPKAGKAKKAAPKKK-

>XP_002121210.1 histone H1.0 [Ciona intestinalis]
-----M--VAKVDTDKKKVAQKKEHPTYAEMIVAAIKALKDRKGASRQAIIVKYIK-ANYKVGDNAA-----
TQC---KHSKRLVTSEKIVNTVGVGASGRFKVPQAAVPTTTKKVAKK--T-EKEAKPAQKKTAAAKKTA-AKSP-----KKQAK
KSPAKKPAQKKAAPKVKTTKSPKPTTKPAA---KSPAKKAAK-----KPAAKKSAKKTTKK--

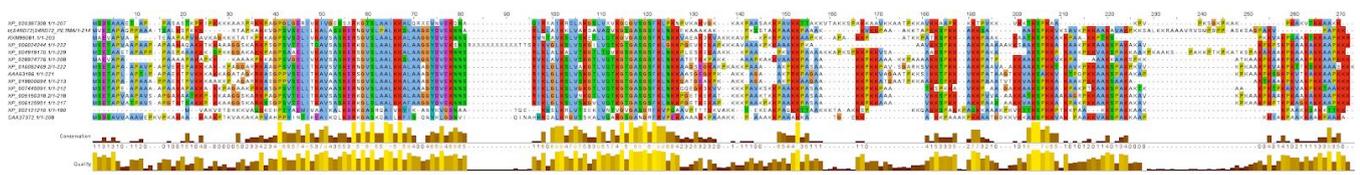
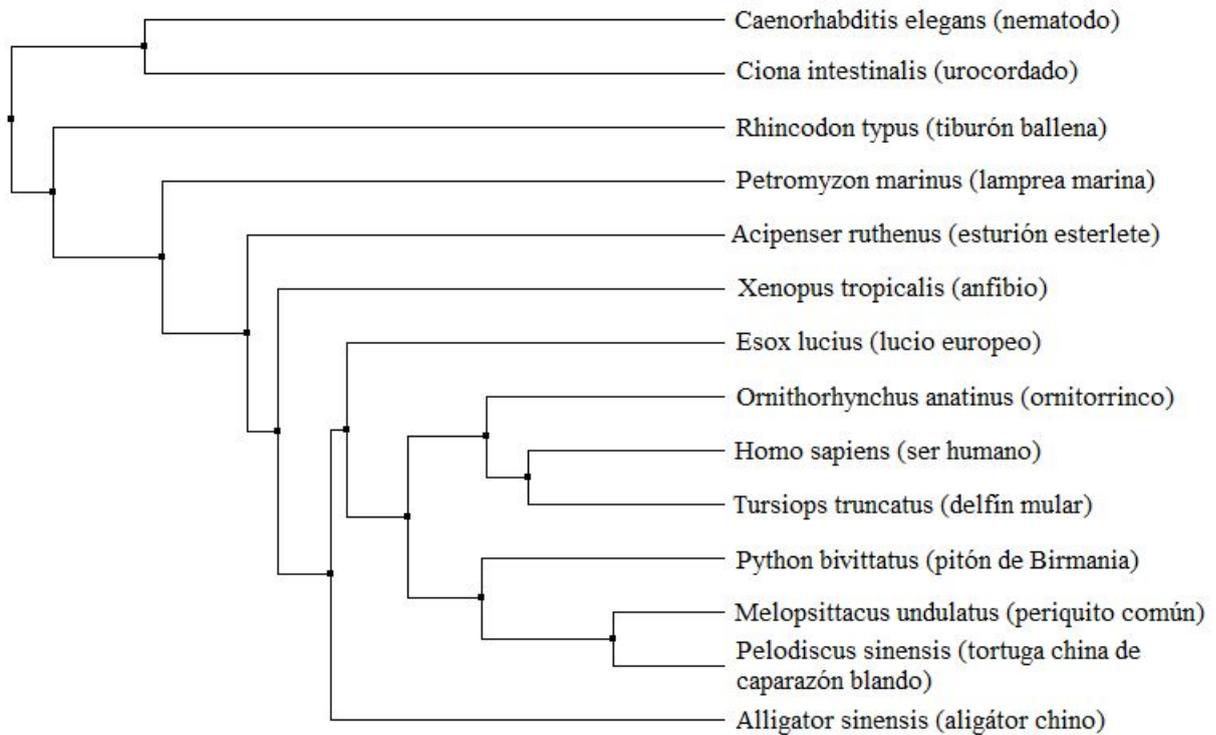
>CAA37372.1 histone H1 [Caenorhabditis elegans]
MSDSAVVAAAVEPKVPKAKAA--KAAKPTKVAKAKAPVAHPPYINTIKEAIKQLKDRKGASKQAILKFIS-QNYKLGDNVI-----
QINAHHRQALKRGVTSKALVQARGSGANGRFRVPEKAAAANKPAAAK--P-AAAKKPAAKKA-----TG--EKK-----A-
KPPAAAKPKKAATGDKKVKKAKSPKPKVAK-PAAKKVAKSPAKKAAP-----KKIAKPAAKKAAPAKA--

```

Al igual que en las dos anteriores proteínas, he cambiado cada código de las especies para mejorar la comprensión del cladograma.



En este cladograma hay más diferencias que en los dos anteriores. La primera gran divergencia es que *Ciona intestinalis* (urocordado) está más relacionado con *Caenorhabditis elegans* (la especie del grupo externo) que con el filo de los cordados. Otra diferencia es que *Esox lucius* (lucio europeo), tiene un antepasado común con el clado de los tetrápodos, cuando la especie que tendría que estar en ese lugar es *Xenopus tropicalis* (anfíbio). Por último, la especie *Alligator sinensis* (aligátor chino), no está en el clado de los saurópsidos, sino que tiene otra rama que tiene un antepasado común, al que deriva la especie *Esox lucius* (lucio europeo).



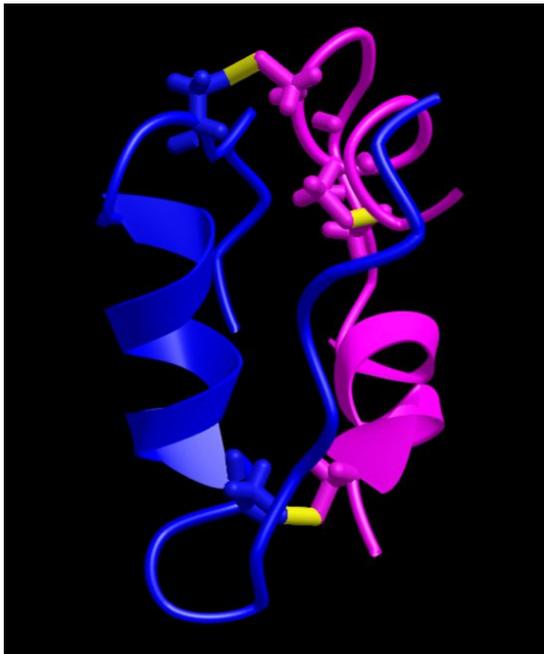
En este gráfico se pueden ver tres partes bien diferenciadas, que tienen que ver con la estructura de la proteína H1. Estas partes pueden que hayan cambiado menos porque son importantes para la función de la proteína y se han mantenido igual a lo largo de la evolución.

5.2.4.- Estructura tridimensional

Para demostrar la estructura más conservada en un esquema tridimensional, he escogido la secuencia de insulina de *Homo sapiens* (ser humano), porque es el mejor gráfico y árbol que ha salido.

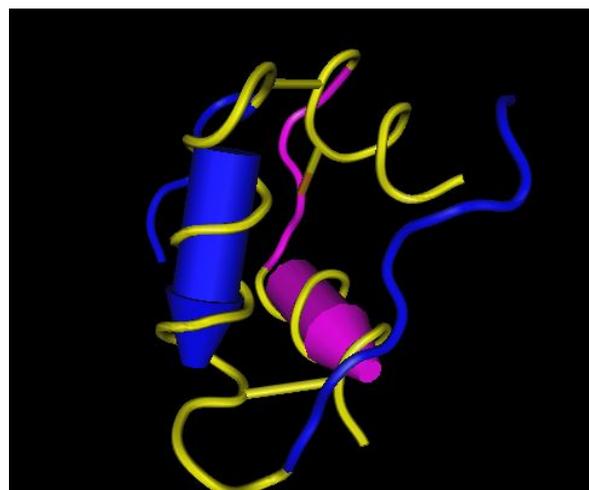
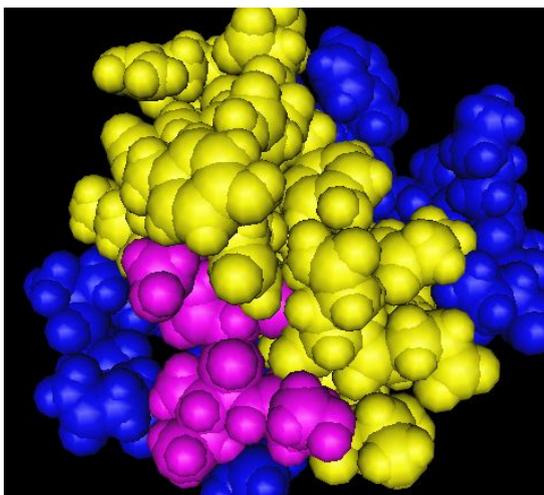
Primero, en la página de la base de datos NCBI, he buscado la estructura tridimensional, y con el programa Cn3D, que viene incluido en esta base y sirve como una aplicación de ayuda

para ver estructuras tridimensionales y mostrar simultáneamente la estructura, la secuencia y la alineación, he seleccionado la parte de la secuencia que más se conservaba fijándome en el gráfico anterior.



Este es un esquema de la estructura tridimensional de la forma activa (monómero) de la insulina de homo sapiens. Como se puede observar está formada por dos cadenas, la cadena A de 21 aminoácidos, de color rosa y la cadena B, de 30 aminoácidos en color azul y cuenta con tres puentes disulfuro en amarillo (unión covalente entre dos átomos de azufre). Además, la parte en hélice de las dos cadenas es la estructura secundaria.

Después, en color amarillo he seleccionado las partes que estaban más conservadas. En la cadena rosa, he seleccionado 17 aminoácidos y en la azul 16. En la cadena azul, en el inicio y el final es donde varía más la proteína y se producen más cambios a lo largo de la evolución. Esto significa, que en los extremos de las cadenas es donde se producen más variaciones y por lo tanto es la que está menos conservada entre especies.



6.- CONCLUSIONES

A partir de las relaciones taxonómicas actuales de los cordados, la hipótesis que quería demostrar es ver si las secuencias de aminoácidos de 15 especies de este filo siguen estas mismas relaciones. Y además, entre diferentes tipos de proteínas como la insulina (hormona), histona H1 (estructural) y citocromo c (transportadora) saber cuál de ellas cumple mejor con estos marcadores evolutivos.

La insulina ha sido el árbol que mejores resultados ha dado, con variaciones muy pequeñas. Ha tenido tres diferencias mínimas con el cladograma “base”, en las especies de *Xenopus tropicalis* (anfibio), *Ornithorhynchus anatinus* (ornitorrinco), *Alligator sinensis* (aligátor chino) y *Melopsittacus undulatus* (periquito común).

Al ser una hormona, lo esperable hubiera sido que fuera el árbol con más distinciones, ya que, en principio, varía más a lo largo de la evolución. Pero como es una proteína pequeña, y como se ha demostrado en este trabajo, está muy conservada en el grupo de los cordados, ha sido el árbol que mejores resultados ha dado.

El citocromo c, como se ha podido observar en el gráfico de semejanza era muy parecido en todas las secuencias, por lo que, un mínimo cambio en un aminoácido ha podido hacer que las especies más emparentadas aparezcan distantes en el cladograma. Las diferencias obtenidas en las ramas de este cladograma han sido en las especies *Python bivittatus* (pitón de Birmania), *Rhincodon typus* (tiburón ballena), *Ornithorhynchus anatinus* (ornitorrinco), y *Tursiops truncatus* (delfin mular).

Por ello, el árbol tiene bastantes diferencias con respecto al modelo y se puede decir que ha evolucionado muy poco a lo largo del tiempo.

Hipotéticamente, la histona H1 tendría que ser el árbol más parecido, ya que es una proteína estructural, que varía poco a lo largo de la evolución. Sin embargo, no ha sido así. Esto puede ser porque tiene varios subtipos, y no he escogido los mismos subtipos para todas las proteínas, ya que las diferencias entre ellas son mínimas y en las bases de datos no había de

todos los subtipos. Las divergencias en este árbol se han dado en las especies, *Ciona intestinalis* (urocordado), *Esox lucius* (lucio europeo) y *Alligator sinensis* (aligátor chino).

En la configuración espacial de la insulina, he encontrado que la región más conservada está en una única zona que debe cumplir funciones fundamentales para esta proteína.

Por todo esto, la conclusión final que he obtenido es que sí que se demuestra que las secuencias de las proteínas siguen las relaciones taxonómicas actuales con pocos cambios. La proteína que presenta mejor las relaciones de las especies en los árboles evolutivos ha sido la insulina, después el citocromo c y por último la histona H1.

Por lo que, las proteínas, al mostrar el grado de similitud entre las secuencias de aminoácidos de las especies, son un buen instrumento para reconstruir la filogenia de los seres vivos y especialmente proteínas como la insulina son mejores para conocer la historia evolutiva del grupo de los cordados.

7.- FUENTES CONSULTADAS

Reloj molecular. Disponible en:

- https://en.wikipedia.org/wiki/Emile_Zuckerland
- https://es.wikipedia.org/wiki/Reloj_molecular

Secuencias de proteínas y estructura. Disponible en:

- <http://www.bionova.org.es/biocast/tema08.htm>
- <https://es.wikipedia.org/wiki/Prote%C3%ADna>
- <http://iesmonre.educa.aragon.es/alumnos0607/websnov/proteinas/strucproteinas.htm>
- <http://tesisdeinvestigadores.blogspot.com/2011/05/las-proteinas.html>
- http://www.ibt.unam.mx/computo/pdfs/libro_25_aniv/capitulo_26.pdf
- <http://www.innovabiologia.com/biodiversidad/diversidad-animal/evolucion-electroforesis-proteinas/>
- <http://www.educa.madrid.org/web/ies.rayuela.mostoles/webrayuela/Publicaciones/Apuntes%20Biologia%20Bach2/Tema%20Bioquimica/Protidos.htm>
- <http://www.educa.madrid.org/web/ies.rayuela.mostoles/webrayuela/Publicaciones/Apuntes%20Biologia%20Bach2/Tema%20Bioquimica/AcidosNucleicos.htm>
- https://biologia-geologia.com/biologia2/444_funciones_de_las_proteinas.html

Enzimas. Disponible en:

- <https://es.wikipedia.org/wiki/Enzima>

Alostéricos. Disponible en:

- https://es.wikipedia.org/wiki/Interacci%C3%B3n_alost%C3%A9rica
- https://es.wikipedia.org/wiki/Regulaci%C3%B3n_alost%C3%A9rica

Centros activos. Disponible en:

- <https://quimica.laguia2000.com/conceptos-basicos/sitio-activo>
- https://es.wikipedia.org/wiki/Sitio_activo

Alineamientos de secuencias. Disponible en:

- http://www.pdg.cnb.uam.es/cursos/PVallarta_2004/03_Alignments/Alineamiento_secuencias/teoria.html
- https://bioinf.comav.upv.es/courses/intro_bioinf/alineamientos.html
- https://bioinf.comav.upv.es/courses/intro_bioinf/multiple.html
- <http://www.jalview.org/help/html/calculations/conservation.html>
- <http://www.jalview.org/help/html/calculations/quality.html>
- <http://www.uco.es/~bb1rofra/analisis/Tema4/Tema4.htm>
- <https://sites.google.com/a/agro.uba.ar/bioinformatica/Home/alineamientosmultiples>
- https://es.wikipedia.org/wiki/Alineamiento_m%C3%BAltiple_de_secuencias#:~:text=El%20alineamiento%20m%C3%BAltiple%20de%20secuencias,incluso%20amino%20Acidos%20nucle%C3%B3tidos%20individuales
- http://vis.usal.es/rodrigo/documentos/bioinfo/temas/5_Alineamientos%20m%C3%BAltiples.pdf

Material y métodos. Disponible en:

- http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352009000400003#:~:text=El%20NCBI%20agrupa%20sus%20bases,Data%20bases%20Molecular%20Data%20Genomes
- https://es.wikipedia.org/wiki/Centro_Nacional_para_la_Informaci%C3%B3n_Biotecnol%C3%B3gica
- <https://en.wikipedia.org/wiki/Clustal>
- https://en.wikipedia.org/wiki/List_of_sequence_alignment_software
- <https://en.wikipedia.org/wiki/Jalview>

- <https://www.genome.jp/tools-bin/clustalw>
- <https://www.ncbi.nlm.nih.gov/>
- <https://www.ebi.ac.uk/Tools/msa/muscle/>

Evolución. Disponible en:

- https://es.wikipedia.org/wiki/Evoluci%C3%B3n_de_los_peces#:~:text=Este%20grupo%20se%20divide%20en,pr%C3%B3ximos%20a%20los%20vertebrados%20terrestres.
- <https://es.wikipedia.org/wiki/Chordata>
- [https://es.wikipedia.org/wiki/Gnathostomata#:~:text=Los%20gnat%C3%B3stomos%20o%20gnatostomados%20\(Gnathostomata,stoma\)%20%22boca%22\).](https://es.wikipedia.org/wiki/Gnathostomata#:~:text=Los%20gnat%C3%B3stomos%20o%20gnatostomados%20(Gnathostomata,stoma)%20%22boca%22).)
- <https://es.wikipedia.org/wiki/Actinopterygii>
- <https://es.wikipedia.org/wiki/Vertebrata>

Proteínas escogidas. Disponible en:

- <https://es.wikipedia.org/wiki/Insulina>
- https://es.wikipedia.org/wiki/Citocromo_c
- <https://www.uab.cat/web/detalle-noticia/historia-evolutiva-de-la-familia-de-histonas-h1-1-h1-5-de-mamiferos-1345680342040.html?noticiaid=1345729229535>
- <https://es.wikipedia.org/wiki/Histona>
- <https://www.irbbarcelona.org/es/news/la-histona-1-el-guardian-de-la-estabilidad-del-genoma>
- <https://www.dciencia.es/la-insulina/>
- <http://proteinasestructurafuncion.usal.es/moleculas/Insulina/index.html>
- <https://dialnet.unirioja.es/servlet/articulo?codigo=6431287#:~:text=La%20insulina%20es%20una%20hormona,dentro%20de%20la%20cadena%20A.>
- <https://www.dciencia.es/la-insulina/>

Especies escogidas. Disponible en:

- https://es.wikipedia.org/wiki/Ciona_intestinalis
- https://es.wikipedia.org/wiki/Caenorhabditis_elegans
- https://es.wikipedia.org/wiki/Taeniopygia_guttata
- https://es.wikipedia.org/wiki/Alligator_sinensis
- https://es.wikipedia.org/wiki/Pseudonaja_textilis
- https://es.wikipedia.org/wiki/Pelodiscus_sinensis
- https://es.wikipedia.org/wiki/Tursiops_truncatus
- https://es.wikipedia.org/wiki/Homo_sapiens
- https://es.wikipedia.org/wiki/Ornithorhynchus_anatinus
- <https://amphibiaweb.org/species/5246>
- https://es.wikipedia.org/wiki/Xenopus_tropicalis
- https://es.wikipedia.org/wiki/Latimeria_chalumnae
- https://es.wikipedia.org/wiki/Esox_lucius
- https://es.wikipedia.org/wiki/Lepisosteus_oculatus
- https://es.wikipedia.org/wiki/Acipenser_ruthenus
- https://es.wikipedia.org/wiki/Amblyraja_radiata
- https://es.wikipedia.org/wiki/Rhincodon_typus
- <https://eu.oceana.org/es/eu/que-hacemos/fauna-y-flora-marina/tiburones/especies-en-peligro/tiburon-baillena>
- https://es.wikipedia.org/wiki/Petromyzon_marinus
- <https://en.wikipedia.org/wiki/Branchiostoma>
- <https://www.floridamuseum.ufl.edu/discover-fish/species-profiles/amblyraja-radiata/>
- <http://www.vertebradosibericos.org/peces/petmar.html>
- https://es.wikipedia.org/wiki/Python_bivittatus

Bases de datos de secuencias de la insulina. Disponible en:

- <https://www.ncbi.nlm.nih.gov/protein/1150763392>
- <https://www.ncbi.nlm.nih.gov/protein/1822396983>
- <https://www.ncbi.nlm.nih.gov/protein/1160096936>
- <https://www.ncbi.nlm.nih.gov/protein/1822491925>
- <https://www.ncbi.nlm.nih.gov/protein/1564225222>
- <https://www.ncbi.nlm.nih.gov/protein/742207409>
- <https://www.ncbi.nlm.nih.gov/protein/134254262>
- <https://www.ncbi.nlm.nih.gov/protein/386828>
- <https://www.ncbi.nlm.nih.gov/protein/1626067476>
- <https://www.ncbi.nlm.nih.gov/protein/470618679>
- <https://www.ncbi.nlm.nih.gov/protein/558217065>
- <https://www.ncbi.nlm.nih.gov/protein/602634847>
- <https://www.ncbi.nlm.nih.gov/protein/557321174>
- <https://www.ncbi.nlm.nih.gov/protein/527262035>
- <https://www.ncbi.nlm.nih.gov/protein/453231772>

Bases de datos de secuencias del citocromo c. Disponible en:

- <https://www.ncbi.nlm.nih.gov/protein/198438245>
- <https://www.ncbi.nlm.nih.gov/protein/126149239>
- <https://www.ncbi.nlm.nih.gov/protein/1160104945>
- <https://www.ncbi.nlm.nih.gov/protein/1836596819>
- <https://www.ncbi.nlm.nih.gov/protein/746816866>
- <https://www.ncbi.nlm.nih.gov/protein/56118488>
- <https://www.ncbi.nlm.nih.gov/protein/181242>
- <https://www.ncbi.nlm.nih.gov/protein/1626093538>
- <https://www.ncbi.nlm.nih.gov/protein/1835252533>
- <https://www.ncbi.nlm.nih.gov/protein/558229181>
- <https://www.ncbi.nlm.nih.gov/protein/602627881>
- <https://www.ncbi.nlm.nih.gov/protein/557288879>
- <https://www.ncbi.nlm.nih.gov/protein/527272765>
- <https://www.ncbi.nlm.nih.gov/protein/351060959>

Base de datos de secuencias de la histona H1. Disponible en:

- <https://www.ncbi.nlm.nih.gov/protein/198424338>
- <https://www.ncbi.nlm.nih.gov/protein/1160121848>
- <https://www.ncbi.nlm.nih.gov/protein/1564225402>
- <https://www.ncbi.nlm.nih.gov/protein/1634091342>
- <https://www.ncbi.nlm.nih.gov/protein/512876989>
- <https://www.ncbi.nlm.nih.gov/protein/184072>
- <https://www.ncbi.nlm.nih.gov/protein/1626109452>
- <https://www.ncbi.nlm.nih.gov/protein/1131241142>
- <https://www.ncbi.nlm.nih.gov/protein/557282692>
- <https://www.ncbi.nlm.nih.gov/protein/1838329885>
- <https://www.ncbi.nlm.nih.gov/protein/602652765>
- <https://www.ncbi.nlm.nih.gov/protein/558175042>
- <https://www.ncbi.nlm.nih.gov/protein/10886>

Estructura de la insulina de Homo sapiens. Disponible en:

- <https://www.ncbi.nlm.nih.gov/Structure/pdb/6X4X>