

Historia Evolutiva de la Ceruloplasmina a partir de una Multicobre Oxidasa Ancestral de 6  
Dominios

Isabella Beltrán Triana

Trabajo de grado para optar el título de Bióloga

Director  
Jorge Hernández Torres  
Ph.D. Biología Molecular

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS BÁSICAS  
ESCUELA DE BIOLOGÍA  
BUCARAMANGA  
2018

## Tabla de Contenido

|                                                                                       | Pág.      |
|---------------------------------------------------------------------------------------|-----------|
| <b>Introducción.....</b>                                                              | <b>9</b>  |
| <b>1. Objetivos.....</b>                                                              | <b>11</b> |
| 1.1 General.....                                                                      | 11        |
| 1.2 Específicos.....                                                                  | 11        |
| 1.3 Competencias que Desarrollará el Pasante. ....                                    | 11        |
| <b>2. Marco Teórico .....</b>                                                         | <b>12</b> |
| 2.1 De la Cupredoxina al Gen CP Eucariota.....                                        | 12        |
| 2.2 Origen Evolutivo de los Intrones.....                                             | 13        |
| 2.3 Pérdida y Ganancia de Intrones. ....                                              | 14        |
| 2.4 Reconstrucción de Relaciones Ancestrales. ....                                    | 15        |
| 2.4.1 Alineamientos Múltiples de Secuencias.....                                      | 16        |
| 2.4.2 Modelos Evolutivos.....                                                         | 16        |
| 2.4.3 Máxima Verosimilitud (MV). ....                                                 | 17        |
| <b>3. Materiales y Métodos.....</b>                                                   | <b>18</b> |
| 3.1 Captura de Secuencias.....                                                        | 18        |
| 3.2 Determinación de Organización y Límites de Intrón/Exón.....                       | 18        |
| 3.3 Análisis de la Fase de los Intrones .....                                         | 19        |
| 3.4 Alineamiento de Secuencias.....                                                   | 19        |
| 3.5 Reconstrucción de los Eventos de Inserción/Escisión de Intrones .....             | 20        |
| <b>4. Resultados .....</b>                                                            | <b>20</b> |
| 4.1 Captura de Secuencias.....                                                        | 20        |
| 4.2 Correlación de las Fases de Intrones con la Sucesión de Dominios Cupredoxina..... | 24        |
| 4.2.1 Intrones Ancestrales.....                                                       | 25        |
| 4.2.2 Cronología de la Deleción e Inserción de Intrones de la Ceruloplasmina. ....    | 27        |
| 4.3 Reconstrucción de Relaciones Ancestrales. ....                                    | 40        |
| <b>5. Discusión .....</b>                                                             | <b>43</b> |
| 5.1 Modelo Evolutivo de la Ceruloplasmina.....                                        | 45        |
| <b>6. Conclusiones .....</b>                                                          | <b>48</b> |
| <b>7. Recomendaciones.....</b>                                                        | <b>48</b> |
| <b>Referencias Bibliográficas.....</b>                                                | <b>49</b> |

## Lista de Figuras

|                                                                                                                                    | Pág. |
|------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Figura 1.</b> Resumen del esquema evolutivo de las MCOs de seis dominios.....                                                   | 13   |
| <b>Figura 2.</b> Organización de la estructura génica entre 12 especies de organismos eucariotas.                                  | 26   |
| <b>Figura 3.</b> Alineamiento de aminoácidos de 13 eucariotas de diferentes grupos.....                                            | 39   |
| <b>Figura 4.</b> Análisis de agrupamiento de aminoácidos hidrofóbicos entre los dominios de la ceruloplasmina en Homo sapiens..... | 40   |
| <b>Figura 5.</b> Reconstrucción de máxima verosimilitud de la ceruloplasmina en 93 especies de vertebrados.....                    | 42   |
| <b>Figura 6.</b> Reconstrucción de CP, HEPH, HEPH1 y F8 en 11 especies de diferentes linajes eucariotas. ....                      | 43   |
| <b>Figura 7.</b> Modelo evolutivo de la ceruloplasmina. ....                                                                       | 47   |

**Lista de Tablas****Pág.**

|                                                                                                                                                                  |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Tabla 1.</b> Códigos de acceso a las secuencias génicas obtenidas de las bases de datos. ....                                                                 | <b>21</b> |
| <b>Tabla 2.</b> Conservación de intrones entre los eucariotas implementados en el alineamiento<br>respecto a la estructura del gen CP en <i>H. sapiens</i> ..... | <b>27</b> |
| <b>Tabla 3.</b> Valores de RMSD entre los 6 dominios cupredoxina del gen CP. ....                                                                                | <b>39</b> |

## Resumen en Español

**TÍTULO:** HISTORIA EVOLUTIVA DE LA CERULOPLASMINA A PARTIR DE UNA MULTICOBRE OXISADA ANCESTRAL DE 6 DOMINIOS<sup>1</sup>

**AUTOR:** ISABELLA BELTRÁN TRIANA<sup>2</sup>

**PALABRAS CLAVES:** CERULOPLASMINA, EVOLUCIÓN, INTRONES, EXONES, OXIDASA MULTICOBRE.

### DESCRIPCION:

La evolución de la estructura de los genes ha sido materia de discusión entre los biólogos moleculares, asunto que permanece poco entendido en la actualidad (Roy *et al.*, 2006). Casos concretos como la evolución de las MCOs ha dejado abierto la discusión de los modelos evolutivos que podrían explicar su historia (Vasin *et al.*, 2013). Tal es el caso de la ceruloplasmina que, a partir de los análisis estructurales de la proteína, homología intra-dominio y análisis filogenéticos de los dominios, se ha sugerido un origen a partir de una sucesión duplicación de dominios cupredoxina (Terzulli *et al.*, 2009; Murphy *et al.*, 1984). No obstante, nuestras reconstrucciones de la estructura del gen y las relaciones filogenéticas revelan una amplia dinámica de inserciones y deleciones de intrones a lo largo de la historia evolutiva de eucariotas, los cuales han contribuido a la organización génica de la familia de las MCOs. El gen de la ceruloplasmina surge a partir de un dominio cupredoxina ancestral que sufre repetidas duplicaciones, permitiendo la formación de otras proteínas de 2 y 3 dominios.

---

1 Trabajo de Pasantía.

2 Facultad de Ciencias, Escuela de Biología, Director: Jorge Hernández Fuentes, Ph.D. Biología Molecular.

## Resumen en Ingles

**TITLE:** EVOLUTIONARY HISTORY OF CERULOPLASMIN FROM AN ANCESTRAL MULTICOPPER OXIDASE OF 6 DOMAINS.

**AUTHOR:** ISABELLA BELTRÁN TRIANA

**KEYWORDS:** EVOLUTION, CERULOPLASMIN, INTRONS, EXONS, MULTICOPPER OXIDASE.

### DESCRIPTION:

The evolution of the structure of genes has been the subject of discussion among molecular biologists, an issue that remains poorly understood at present (Roy *et al.*, 2006). Concrete cases such as the evolution of MCOs have left open the discussion of evolutionary models that could explain their history (Vasin *et al.*, 2013). Such is the case of ceruloplasmin that, from structural analyses of the protein, as well as intra-domain homology and phylogenetic analysis of the domains, an origin from a succession duplication of cupredoxin domains has been suggested (Terzulli *et al.*, 2009; Murphy *et al.*, 1984). However, our reconstructions of gene structure and phylogenetic relationships reveal a wide dynamic of introns insertions and deletions throughout the evolutionary history of eukaryotes, which have contributed to the gene organization of the MCO family. The ceruloplasmin gene arises from an ancestral cupredoxin domain that suffers repeated duplications, allowing the formation of other proteins of 2 and 3 domains.

---

1 Internship Research

2 Science Faculty, Department of Biology. Director: Jorge Hernández Fuentes, Ph.D. Molecular Biology.

## Introducción

Las multicobre oxidasas (MCOs) son una familia de enzimas que contienen como grupo prostético uno o más iones de cobre. Las MCOs oxidan una amplia gama de sustratos fenólicos y no fenólicos, como el hierro, cobre o ácido ascórbico (Baldrian *et al.*, 2006). Hacen parte de las MCOs las lacasas, ascorbato oxidasas, nitrito reductasas y la ferroxidasa ceruloplasmina (Sakurai *et al.*, 2007). Las MCOs evolucionaron a partir de la duplicación sucesiva de un dominio cupredoxina en proteínas de dos y tres dominios (lacasas y otras oxidasas) y seis dominios (ceruloplasmina).

La ceruloplasmina es una cuproproteína que contiene de 6 a 8 átomos de cobre en su estructura. Es la principal proteína de transporte de cobre en el plasma sanguíneo, ligando más del 95% del cobre (Koschinisky *et al.*, 1986). Además de regular la homeostasis del hierro, presenta actividad ferroxidasa del  $\text{Fe}^{2+}$  (ferroso) a  $\text{Fe}^{3+}$  (férrico), reacción que permite la formación de la hemoglobina (Nakamura *et al.*, 2005). También, es un agente antioxidante y participa en la oxidación de la serotonina, epinefrina y norepinefrina (Virit *et al.*, 2008). Esta enzima es sintetizada principalmente en el hígado —pero también en cerebro, pulmones, bazo y testículos— y su deficiencia es característica en la enfermedad de Wilson, rasgo que se hereda por autosomas recesivos (Hellman *et al.*, 2002; Messerschmidt *et al.*, 1990). Fue descrita por primera vez en 1948 por Holmberg y Laurell, y está distribuida en todos los vertebrados. Debido a su polifuncionalidad y los procesos que la originaron, su estudio es de gran interés tanto para la biología como para la medicina (Vasin *et al.*, 2013).

La ceruloplasmina, codificada por el gen *CP*, es una enzima de 6 dominios cupredoxina, que cubre alrededor de 50 kb con 19 exones y 18 intrones, de diferentes fases (0, 1 ó 2). Los modelos más recientes plantean eventos sucesivos de duplicación de un dominio cupredoxina primigenio, que finalmente dieron origen a la ceruloplasmina (Rydén *et al.*, 1993; Nakamura

*et al.*, 2003; Vasin *et al.*, 2013). En 1995, se publicó la organización de exones e intrones y fases del gen *CP* de *Homo sapiens* (Daimon *et al.*, 1995). Aunque se han reportado topologías génicas en algunos vertebrados (Doolittle *et al.*, 2015), no se han publicado las repercusiones de la inserción/escisión de los intrones en la evolución del gen *CP*. No obstante, en la actualidad hay sendos genomas completos disponibles que permiten reconstruir la dinámica de inserción/delección de intrones que dieron origen a la topología actual del gen *CP*.

Dado que en procariotas y Archaea también existen multicobre oxidasas de 6 dominios con función ferroxidasa o  $\text{Mn}^{2+}$  oxidasa (Vasin *et al.*, 2013), es indiscutible que todos los sextuplicados de la ceruloplasmina deben compartir un mismo origen evolutivo. De lo contrario, estaríamos frente a un escenario de convergencia evolutiva (Stern, 2013), un escenario muy poco probable, con base en la literatura científica vigente (Rydén *et al.*, 1993; Nakamura *et al.*, 2003; Vasin *et al.*, 2013). Aunque el origen evolutivo común es el más plausible, es probable que, en la reconstrucción de la organización de exones e intrones, aunada a análisis filogenéticos, encontremos indicios de eventos ancestrales mediados por las fases de los intrones, en las que, los intrones de fase 0 son los más antiguos en la evolución (De Souza *et al.*, 1998) y pudieron modelar el gen *CP* y secuencias ortólogas, de lo cual nada se ha publicado a la fecha.

Por lo anterior, en esta pasantía de investigación, nos proponemos reconstruir la historia evolutiva del gen *CP* nuclear, con base en la topología de intrones y exones en diferentes grupos de vertebrados y la fase de los intrones. La idea es escudriñar si el análisis de la multiplicidad de intrones y sus fases puede contribuir a perfeccionar los modelos vigentes del origen de la ceruloplasmina, desde una MCO ancestral de 6 dominios del procariota primitivo.

## **1. Objetivos**

### **1.1 General.**

Elucidar la dinámica de inserción/delección de intrones que dieron origen a la topología actual de los genes de la ceruloplasmina en vertebrados, a partir de una multicobre oxidasa ancestral de 6 dominios del procariota primitivo.

### **1.2 Específicos.**

1. Reconstruir la organización de intrones y exones del gen de la ceruloplasmina en diferentes grupos de vertebrados.
2. Determinar el probable ordenamiento cronológico de los eventos de recombinación con base en la fase de los intrones.
3. Contrastar el origen monofilético de las formas actuales de la ceruloplasmina con los eventos de sucesión inserción/delección de intrones.

### **1.3 Competencias que Desarrollará el Pasante.**

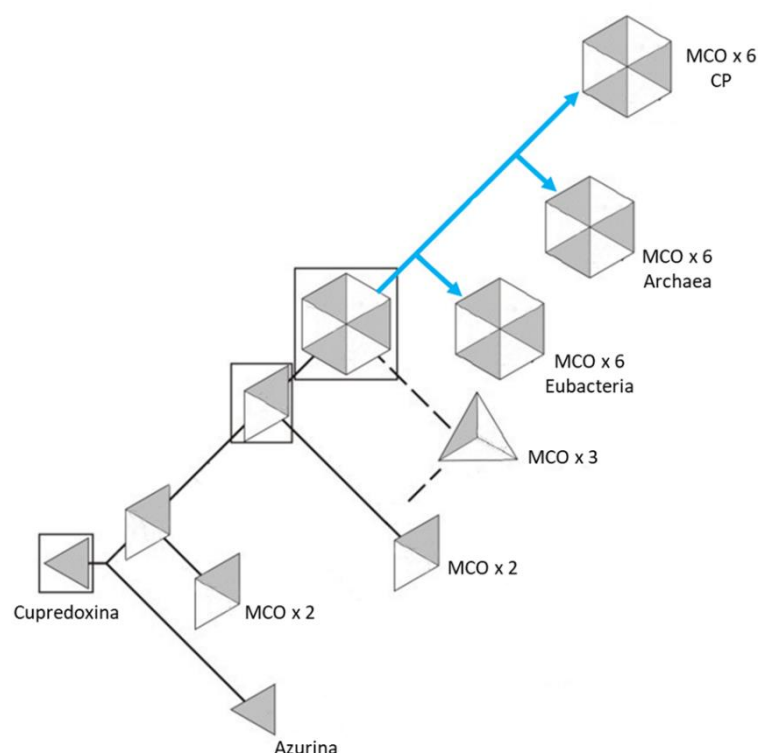
- Captura secuencias nucleotídicas de las bases de datos genómicas de vertebrados con la herramienta BlastN.
- Reconstruye la organización de intrones y exones del gen de la ceruloplasmina en diferentes grupos de vertebrados.

- Construye un modelo cronológico de los eventos de inserción/escisión con base en el análisis de la fase de los intrones.
- Demuestra la coherencia entre el origen monofilético de las formas actuales de la ceruloplasmina con los eventos de sucesión inserción/delección de intrones.

## 2. Marco Teórico

### 2.1 De la Cupredoxina al Gen *CP* Eucariota.

La topología del gen de la ceruloplasmina de *Homo sapiens* fue publicado por Daimon *et al.*, en 1995. El gen cubre alrededor de 50 kb con 19 exones y 18 intrones, de diferentes fases (0, 1 y 2). Desde entonces, nada más se ha publicado sobre este gen en otros grupos taxonómicos, ni sobre la influencia de los intrones en su evolución. En cambio, se han propuesto algunos modelos de la evolución de las oxidasas multicobre (MCOs). En la **Figura 1** se muestra un resumen del modelo propuesto por (Vasin *et al.*, 2013).



**Figura 1.** Resumen del esquema evolutivo de las MCOs de seis dominios.

**Nota:** El dominio cupredoxina originó las MCOs por duplicación sucesiva. Inicialmente, surgió un duplicado (OMC x 2), conocidos hoy como lacasas de dos dominios. Posteriormente, una triplicación del duplicado dio lugar a una OMC de 6 dominios cupredoxina. En ese proceso, surgieron de manera colateral las MCOs de tres dominios (lacasas y ascorbato oxidasas). La OMC ancestral de 6 dominios dio lugar a las isoformas de Eubacteria, Archaea y finalmente la ceruloplasmina (CP) junto con otras secuencias ortólogas de eucariotas (*Blood coagulation factor* (BCF); Fox - multicobre ferroxidasa de algas; Heph1 - hefastina 1). Esquema redibujado a partir del modelo *in extenso* de (Vasin *et al.*, 2013).

## 2.2 Origen Evolutivo de los Intrones.

El descubrimiento de la estructura discontinua génica en 1977 (Berget *et al.*, 1977; Brack *et al.*, 1977; Breathnach *et al.*, 1977; Chow *et al.*, 1977) y en conjunto con el mecanismo del *splicing* alternativo, en el que participa la maquinaria enzimática del spliceosoma, deja abierto el desarrollo de diferentes hipótesis que intentan aclarar la evolución y origen inicial de los intrones nucleares. Dado que esta es una característica distintiva de los eucariotas y al no encontrarse en las bacterias o arqueobacterias, las pistas sobre la pérdida y ganancia de intrones se debe rastrear a partir de los análisis genómicos de Eucaria (Jeffares *et al.*, 2006; Rogozin *et al.*, 2005).

Las dos principales alternativas que pretenden explicar el contexto evolutivo de los intrones proponen: i) el modelo de los ‘intrones tempranos’ sugiere que los intrones son en extremo ancestrales y numerosos desde su aparición (Jeffares *et al.*, 1998; Poole *et al.*, 1998). Han sido heredados del progenote, cuyas diferencias estructurales entre genes homólogos eucariotas se debe a la pérdida de intrones en el curso de la evolución. Así, la pérdida progresiva deja sin remanentes primordiales de intrones nucleares a los ancestros procariotas, en un proceso de ‘racionalización genómica’. Debido a las presiones evolutivas, el tamaño del genoma se reduce por la replicación rápida del ADN o también ‘evolución reductiva’ (Vibrantovski *et al.*, 2003; Rogozin *et al.*, 2012; Penny *et al.*, 2009). En contraste, ii) el modelo de los ‘intrones tardíos’, argumenta que los intrones se insertan continuamente en la evolución de los eucariotas después de su divergencia. Bajo este escenario, es posible que los intrones autocatalíticos (Tipos I y II), encontrados en los genomas procariotas, hayan invadido los genomas eucariotas desde su radiación (Hickey *et al.*, 1992; Stoltzfus *et al.*, 1994; Lynch *et al.*, 2002). Otros autores

### 2.3 Pérdida y Ganancia de Intrones.

El origen de los intrones nucleares y su función en la evolución de los genes es uno de los mayores retos que se sigue debatiendo en la evolución molecular. Diferentes aspectos y evidencias se tienen en cuenta para evaluar su impacto: el momento de su aparición y posterior propagación; la posición conservada de los intrones; las fuerzas evolutivas y mecanismos que han marcado su historia evolutiva: densidad de intrones en diferentes linajes y pérdida y ganancia de intrones. En conjunto, estos factores revelan la complejidad de los genes actuales (Roy *et al.*, 2006; Rogozin *et al.*, 2005).

Aunque uno de los mecanismos más importantes en la evolución de la arquitectura de los genes eucariotas son los eventos de pérdida y ganancia de intrones, sigue siendo un fenómeno poco entendido para la comunidad científica (Rogozin *et al.*, 2012). Sin embargo, análisis en

sistemática sugieren que, en genes de organismos eucariotas altamente conservados, la pérdida y adquisición de nuevos intrones ha actuado durante la evolución de la estructura génica de este grupo, lo que ha indicado que al menos un 25% de los intrones han sido heredados del ancestro común de plantas y animales (Babenko *et al.*, 2014). Además, la misma conservación de las posiciones intrónicas indican incluso una ancestría más lejana en protistas, permitiendo deducir la posibilidad de que estos eventos sucedan desde el inicio de la evolución eucariota (Rogozin *et al.*, 2005; Jeffares *et al.*, 2006; Fedorov *et al.*, 2002; Rogozin *et al.*, 2003).

## 2.4 Reconstrucción de Relaciones Ancestrales.

Dado que la evolución implica que los organismos se modifiquen —sufran mutaciones— en el tiempo y que estas nuevas características adquiridas sean transmitidas a las siguientes generaciones, la información entre ancestro-descendiente se preserva en el genoma, lo que permite representar los parentescos en árboles filogenéticos (Abascal *et al.*, 2014).

Los árboles filogenéticos se pueden construir con diferentes métodos -Parsimonia, Inferencia Bayesiana o Máxima Versomilitud- y son el producto de inferencias matemáticas, las cuales describen patrones o comportamientos de las mutaciones (datos) que se mantienen o no conservadas en las secuencias. Los análisis filogenéticos a nivel molecular, usualmente se soportan en las secuencias de ADN o proteínas, en el que las sustituciones de los nucleótidos o aminoácidos son la materia prima para rastrear el origen evolutivo en común (homología) entre diferentes organismos (Li, 1997). Así, secuencias homólogas que han surgido a partir de eventos de especiación (divergencia entre dos o más especies) se conocen como genes ortólogos. Por otra parte, las secuencias que han surgido por eventos de duplicación (el gen ocupa un espacio adicional en el genoma del individuo), se conocen como genes parálogos (Dufayard *et al.*, 2005).

**2.4.1 Alineamientos Múltiples de Secuencias.** Identificar las regiones homólogas entre las posiciones de las secuencias a partir de alineamientos múltiples, es lo que permite generar las hipótesis de los procesos evolutivos. Debido a que las comparaciones entre diferentes secuencias varían dependiendo de su similitud, es posible identificar las relaciones funcionales o estructurales de las proteínas en diferentes organismos o especies (Chenna *et al.*, 2003). Algunos algoritmos desarrollados para estos fines, como la serie de programas Clustal, facilitan la interpretación al generar matrices de sustituciones que van variando conforme se generan los alineamientos hasta la obtención global del mismo (Thompson *et al.*, 1994).

**2.4.2 Modelos Evolutivos.** La conexión entre los alineamientos y los métodos de reconstrucción filogenéticos son los modelos de evolución, los cuales describen matemáticamente la evolución. Ya sea que la unidad de estudio sean codones, aminoácidos o nucleótidos, los modelos dependen del cálculo probabilístico de que las unidades cambien de un estado a otro y de la frecuencia con la que lo hacen (Yang, 2006; Abascal *et al.*, 2014).

Desde muy simples hasta muy complejos, varían los modelos, así, el modelo más simple se conoce como Jukes Cantor (JC) (Jukes *et al.*, 1969), el cual asume que cualquier cambio en los nucleótidos o frecuencia tienen la misma probabilidad de ocurrir. En contraste, el modelo más complejo se conoce por sus siglas en inglés como GTR (General Time Reversible) (Tavaré, 1986). Este modelo asume parámetros diferentes para cada tipo de sustitución y frecuencia del nucleótido. Algunos programas como JModelTest (Posada *et al.*, 2008) o PartitionFinder 2 (Lanfear *et al.*, 2016) se usan para hallar el modelo que mejor se adapte a las secuencias de estudio. Sin embargo, programas de cómputo como Randomized Axelerated Maximum Likelihood, RAxML para estimar filogenias con el método de Máxima Verosimilitud por

defecto tienen predeterminado un modelo evolutivo, el cual corresponde a GTR (Stamatakis *et al.*, 2016).

**2.4.3 Máxima Verosimilitud (MV).** Para reconstruir los sucesos evolutivos, se pueden aplicar diferentes métodos que pueden ser calculados con las distancias génicas o caracteres. Uno de los métodos que se basa en la reconstrucción ancestral a partir de caracteres y modelos es el de Máxima Verosimilitud (MV) (Felsenstein *et al.*, 1973). MV busca el árbol filogenético y la longitud de las ramas con la mayor probabilidad de ser explicados por los datos o lo que es, que refleje las secuencias de ADN observadas (Felsenstein *et al.*, 1981; Strimmer, 1997). Adicionalmente, los métodos de reconstrucción cuentan con un procedimiento importante que demuestra qué tan altas están correlacionadas las probabilidades (*bootstrap*) y refleja la confianza de las relaciones entre los clados (Efron *et al.*, 1996). Valores por encima del 50% son un buen soporte de *bootstrap* en los árboles (Margush *et al.*, 1981). Programas como *Phylogenies by Maximum Likelihood*, PhyML (Guindon *et al.*, 2010), *Genetic Algorithm for Rapid Likelihood Inference*, GARLI (Bazinnet *et al.*, 2014) ó RAxML (Stamatakis *et al.*, 2014) son los más usados para obtener filogenias con el método de MV.

### 3. Materiales y Métodos.

#### 3.1 Captura de Secuencias

Se utilizó la herramienta BlastN (Altschul *et al.*, 1990) para la búsqueda de contigs cromosómicos que albergaran el gen *CP* en las bases de datos Genbank y Ensembl. Para la captura de secuencias se utilizó como referencia la secuencia del gen de la ceruloplasmina (Genbank: D45045.2) y del mRNA (M13699.1). Para la captura de las secuencias de los genes HEPH, HEPH1, F8, FOX1 y FOX, se realizó una búsqueda en las bases de datos *Kyoto Encyclopedia of Genes and Genomes*, KEGG, no redundantes (NR) de Genbank y otras fueron encontradas en Ensembl. En conjunto, se seleccionaron diferentes grupos representativos de eucariotas: peces, reptiles, aves, anfibios, mamíferos, tunicados, moluscos, artrópodos, equinodermos, hemicordados, cnidarios y algas. Todos los organismos se clasificaron de acuerdo con la Taxonomía del *National Center for Biotechnology Information*, NCBI (<http://www.ncbi.nlm.nih.gov/Taxonomy>).

#### 3.2 Determinación de Organización y Límites de Intrón/Exón

Para la construcción de la topología del gen *CP*, inicialmente se mapeó el gen de *Homo sapiens* (Daimon *et al.*, 1995); posteriormente, se reconstruyeron los demás grupos taxonómicos. La base de datos del NCBI tiene incorporado el algoritmo de predicción génica, GENSCAN (Wu *et al.*, 2008) de manera que en la sección *Nucleotide* se puede acceder a la delimitación de los exones. Por otra parte, ENSEMBL tiene incluido de manera automática un proceso de anotación génica (Aken *et al.*, 20016) y es posible descargar las secuencias de los

genes con las secuencias discriminadas por exones e intrones. Así, se pudieron contrastar las secuencias genómicas con los cDNAs de los mismos grupos.

### 3.3 Análisis de la Fase de los Intrones

La fase de los intrones se estableció con base en las recomendaciones de Patthy *et al.*, (1987) así: intrones fase 0: aquellos que se ubican entre dos codones e.g., 5'GAC **CAG**:GT—intrón—AG:**GTC** ATG 3'. Intrones fase 1: aquellos que se ubican entre el primer y segundo nucleótidos de un codón e.g., 5'GA CCA **G**:GT—intrón—AG:**GT** CAT G3'. intrones fase 2: aquellos que se ubican entre el segundo y tercer nucleótidos de un codón e.g., 5'G ACC **AG**:GT—intrón—AG:**G** TCA TG3'. Se correlacionó la presencia de intrones y su fase con la sucesión de dominios cupredoxina.

### 3.4 Alineamiento de Secuencias

Se realizaron alineamientos múltiples de secuencias de nucleótidos y aminoácidos con la herramienta ClustalO (Li *et al.*, 2015). Se empleó el método de análisis de los agregados hidrofóbicos (HCA, *Hydrophobic Cluster Analysis*, Callebaut *et al.*, 1997) para determinar intrones ancestrales. El método se centra en el examen de la conservación de los agregados de aminoácidos hidrofóbicos (clusters), que surgen en una representación bidimensional de la secuencia de las proteínas. Mediante el método HCA se ha podido analizar individualmente los dominios de una misma proteína y relacionarlos con los de otras proteínas de función conocida (Callebaut *et al.*, 1997; Woodcock *et al.*, 1992). Se compararon los valores de RMSD, Root Mean Square Deviation (Kufareva *et al.*, 2012) para conocer la similitud estructural entre los dominios de CP.

### 3.5 Reconstrucción de los Eventos de Inserción/Escisión de Intrones

Con base en el análisis de la fase de los intrones y de presencia/ausencia, pudimos deducir un ordenamiento lógico de la cronología de inserción/escisión de intrones. Para demostrar las relaciones monofiléticas, se recurrió a la construcción de árboles filogenéticos con el método de Máxima Verosimilitud (MV) con el programa RAxML versión 8 (Stamatakis *et al.*, 2014). Los parámetros se siguieron por defecto, de acuerdo con el manual (Stamatakis *et al.*, 2016).

## 4. Resultados

### 4.1 Captura de Secuencias

La primera topología del gen *CP* se realizó en 1995 para *H. sapiens*, con 3.321 pb que codifican la proteína (Genbank número M13699.1). La base de datos GenBank aloja el gen secuenciado en 94 especies de vertebrados, de las cuales 93 se usaron para resolver las relaciones ancestrales del gen *CP*. No obstante, en la base de datos genómica Ensembl Metazoa del Instituto Europeo de Bioinformática (EMBL-EBI), se alojada la secuencia del gen *CP* para una especie de eucariota no vertebrado: el ostión o *Crassostrea gigas*. Por otra parte, los genes de otras MCOs de 6 dominios, i.e., (Heph y Heph1- Hefaestina y Hefaestina 1, FVIII- Factor de coagulación VIII y FOX1- Ferroxidasa de algas) se pueden localizar en ambas bases de datos, aunque la cantidad de secuencias es restringida.

Las secuencias de los 5 genes ortólogos se utilizaron para ampliar las relaciones planteadas por Vasin *et al.*, (2013) de las MCOs de 6 dominios. La **Tabla 1** proporciona la información de la clase, orden, género y especie de cada individuo con los respectivos accesos a las bases de datos que se utilizaron en las diferentes etapas de este trabajo.

**Tabla 1.**

Códigos de acceso a las secuencias génicas obtenidas de las bases de datos.

| Orden           | Especies                          | Número de acceso               | Base de datos |
|-----------------|-----------------------------------|--------------------------------|---------------|
| <b>MAMMALIA</b> |                                   |                                |               |
| Primates        | <i>Homo sapiens</i>               | D45045.2; M13699.1             | Genbank       |
|                 | <i>Rhinopithecus bieti</i>        | XM_017856546.1                 | Genbank       |
|                 | <i>Otolemur garnettii</i>         | XM_003789471.3                 | Genbank       |
|                 | <i>Carlito syrichta</i>           | XM_021706408.1                 | Genbank       |
| Rodentia        | <i>Mus musculus</i>               | BC062957.1                     | Genbank       |
|                 | <i>Mus pahari</i>                 | XM_021197165.1                 | Genbank       |
|                 | <i>Castor canadensis</i>          | XM_020158342.1                 | Genbank       |
|                 | <i>Ictidomys tridecemlineatus</i> | XM_013358823.2                 | Genbank       |
|                 | <i>Jaculus jaculus</i>            | XM_012948190.1                 | Genbank       |
|                 | <i>Meriones unguiculatus</i>      | XM_021637856.1                 | Genbank       |
| Artiodactyla    | <i>Ovis aries</i>                 | NM_001009733.1                 | Genbank       |
|                 | <i>Bos taurus</i>                 | NM_001256556.1                 | Genbank       |
|                 | <i>Sus scrofa</i>                 | NM_001267694.2                 | Genbank       |
|                 | <i>Pantholops hodgsonii</i>       | XM_005958414.1                 | Genbank       |
| Perissodactyla  | <i>Equus asinus</i>               | XM_014838548.1                 | Genbank       |
| Chiroptera      | <i>Miniopterus natalensis</i>     | XM_016221503.1                 | Genbank       |
| Scandentia      | <i>Tupaia chinensis</i>           | XM_006149715.2                 | Genbank       |
| Carnivora       | <i>Ailuropoda melanoleuca</i>     | XM_002927039.3                 | Genbank       |
|                 | <i>Leptonychotes weddellii</i>    | XM_006745439.1                 | Genbank       |
| Insectivora     | <i>Erinaceus europaeus</i>        | XM_016185995.1                 | Genbank       |
|                 | <i>Sorex araneus</i>              | XM_012936180.1                 | Genbank       |
| Cetacea         | <i>Physeter catodon</i>           | XM_024131860.1                 | Genbank       |
| Monotremata     | <i>Ornithorhynchus anatinus</i>   | XM_016226163.1                 | Genbank       |
| Afrosoricida    | <i>Echinops telfairi</i>          | XM_013006124.1                 | Genbank       |
|                 | <i>Chrysochloris asiatica</i>     | XM_006876360.1                 | Genbank       |
| <b>AVES</b>     |                                   |                                |               |
| Galliformes     | <i>Numida meleagris</i>           | XM_021394168.1                 | Genbank       |
|                 | <i>Meleagris gallopavo</i>        | XM_003209288.3                 | Genbank       |
|                 | <i>Coturnix japonica</i>          | XM_015872122.1                 | Genbank       |
|                 | <i>Gallus gallus</i>              | NC_006096.5;<br>XM_015291853.1 |               |
| Sphenisciformes | <i>Aptenodytes forsteri</i>       | XM_009284331.2                 | Genbank       |

| Orden             | Especies                               | Número de acceso | Base de datos |
|-------------------|----------------------------------------|------------------|---------------|
|                   | <i>Pygoscelis adeliae</i>              | XM_009334012.1   | Genbank       |
| Charadriiformes   | <i>Calidris pugnax</i>                 | XM_014945801.1   | Genbank       |
|                   | <i>Charadrius vociferus</i>            | XM_009884464.1   | Genbank       |
| Falconiformes     | <i>Haliaeetus leucocephalus</i>        | XM_010584891.1   | Genbank       |
|                   | <i>Haliaeetus albicilla</i>            | XM_009929892.1   | Genbank       |
|                   | <i>Falco cherrug</i>                   | XM_005434308.1   | Genbank       |
|                   | <i>Falco peregrinus</i>                | XM_005233762.1   | Genbank       |
| Passeriformes     | <i>Ficedula albicollis</i>             | XM_016300504.1   | Genbank       |
|                   | <i>Geospiza fortis</i>                 | XM_014312327.1   | Genbank       |
|                   | <i>Sturnus vulgaris</i>                | XM_014886123.1   | Genbank       |
| Accipitriformes   | <i>Aquila chrysaetos canadensis</i>    | XM_011592460.1   | Genbank       |
| Gruiformes        | <i>Balearica regulorum gibbericeps</i> | XM_010307786.1   | Genbank       |
| Pelecaniformes    | <i>Phaethon lepturus</i>               | XM_010293904.1   | Genbank       |
|                   | <i>Egretta garzetta</i>                | XM_009634338.1   | Genbank       |
|                   | <i>Nipponia nippon</i>                 | XM_009468481.1   | Genbank       |
|                   | <i>Phalacrocorax carbo</i>             | XM_009515272.1   | Genbank       |
| Caprimulgiformes  | <i>Caprimulgus carolinensis</i>        | XM_010163603.1   | Genbank       |
| Strigiformes      | <i>Tyto alba</i>                       | XM_009966336.1   | Genbank       |
| Coraciiformes     | <i>Leptosomus discolor</i>             | XM_009946809.1   | Genbank       |
|                   | <i>Merops nubicus</i>                  | XM_008941388.1   | Genbank       |
| Piciformes        | <i>Picoides pubescens</i>              | XM_009902382.1   | Genbank       |
| Procellariiformes | <i>Fulmarus glacialis</i>              | XM_009583250.1   | Genbank       |
| Columbiformes     | <i>Columba livia</i>                   | XM_013369752.2   | Genbank       |
| Tinamiformes      | <i>Tinamus guttatus</i>                | XM_010216892.1   | Genbank       |
| Coliiformes       | <i>Colius striatus</i>                 | XM_010204851.1   | Genbank       |
| Musophagiformes   | <i>Tauraco erythrolophus</i>           | XM_009990571.1   | Genbank       |
| Opisthocomiformes | <i>Opisthocomus hoazin</i>             | XM_009942421.1   | Genbank       |
| Cariamiformes     | <i>Cariama cristata</i>                | XM_009697479.1   | Genbank       |
| Cuculiformes      | <i>Cuculus canorus</i>                 | XM_009567740.1   | Genbank       |
| Apodiformes       | <i>Calypte anna</i>                    | XM_008500229.1   | Genbank       |
|                   | <i>Chaetura pelagica</i>               | XM_010004753.1   | Genbank       |
| Trogoniformes     | <i>Apaloderma vittatum</i>             | XM_009865155.1   | Genbank       |
| <b>REPTILIA</b>   |                                        |                  |               |

| Orden                 | Especies                            | Número de acceso               | Base de datos |
|-----------------------|-------------------------------------|--------------------------------|---------------|
| Testudines            | <i>Chrysemys picta bellii</i>       | XM_005312912.1                 | Genbank       |
|                       | <i>Terrapene mexicana triunguis</i> | XM_024208569.1                 | Genbank       |
| Crocodylia            | <i>Alligator sinensis</i>           | XM_006032943.3                 | Genbank       |
|                       | <i>Alligator mississippiensis</i>   | XM_006259175.3                 | Genbank       |
|                       | <i>Crocodylus porosus</i>           | XM_019532566.1                 | Genbank       |
|                       | <i>Gavialis gangeticus</i>          | XM_019503116.1                 | Genbank       |
| Squamata              | <i>Anolis carolinensis</i>          | XM_003218193.3;<br>NC_014778.1 | Genbank       |
|                       | <i>Pogona vitticeps</i>             | XM_020781856.1                 | Genbank       |
| AMPHIBIA              |                                     |                                |               |
| Anura                 | <i>Xenopus tropicalis</i>           | XM_002937329.3;<br>NC_030681.1 | Genbank       |
|                       | <i>Nanorana parkeri</i>             | XM_018574049.1                 | Genbank       |
| <b>ACTINOPTERYGII</b> |                                     |                                |               |
| Salmoniformes         | <i>Salmo salar</i>                  | XM_014123624.1                 | Genbank       |
|                       | <i>Oncorhynchus tshawytscha</i>     | XM_024422392.1                 | Genbank       |
| Characiformes         | <i>Astyanax mexicanus</i>           | XM_022669471.1                 | Genbank       |
|                       | <i>Pygocentrus nattereri</i>        | XM_017717849.1                 | Genbank       |
| Perciformes           | <i>Amphiprion ocellaris</i>         | XM_023291362.1                 | Genbank       |
|                       | <i>Chionodraco rastrospinosus</i>   | AM419432.1                     | Genbank       |
|                       | <i>Acanthochromis polyacanthus</i>  | XM_022206751.1                 | Genbank       |
| Cypriniformes         | <i>Danio rerio</i>                  | NM_131802.1;<br>NC_007133.7    | Genbank       |
|                       | <i>Sinocyclocheilus rhinoceros</i>  | XM_016569640.1                 | Genbank       |
|                       | <i>Sinocyclocheilus anshuiensis</i> | XM_016453081.1                 | Genbank       |
|                       | <i>Sinocyclocheilus grahami</i>     | XM_016276515.1                 | Genbank       |
| Cichliformes          | <i>Pundamilia nyererei</i>          | XM_005736946.1                 | Genbank       |
| Semionotiformes       | <i>Lepisosteus oculatus</i>         | XM_015360752.1                 | Genbank       |
| Cyprinodontiformes    | <i>Kryptolebias marmoratus</i>      | XM_017420559.2                 | Genbank       |
| Osteoglossiformes     | <i>Paramormyrops kingsleyae</i>     | XM_023843865.1                 | Genbank       |
|                       | <i>Scleropages formosus</i>         | XM_018761899.1                 | Genbank       |
| Carangiformes         | <i>Seriola lalandi dorsalis</i>     | XM_023428685.1                 | Genbank       |
| Clupeiformes          | <i>Clupea harengus</i>              | XM_012825353.1                 | Genbank       |
| Siluriformes          | <i>Ictalurus punctatus</i>          | NM_001278815.1                 | Genbank       |
| <b>CHONDRICHTHYES</b> |                                     |                                |               |
| Orectolobiformes      | <i>Rhincodon typus</i>              | XM_020523081.1                 | Genbank       |

| Orden                | Especies                             | Número de acceso                  | Base de datos |
|----------------------|--------------------------------------|-----------------------------------|---------------|
| Chimaeriformes       | <i>Callorhinchus milii</i>           | XM_007902443.1;<br>NW_006890148.1 | Genbank       |
| <b>BIVALVIA</b>      |                                      |                                   |               |
| Ostreoida            | <i>Crassostrea gigas</i>             | CGI_10006923                      | Ensembl       |
| <b>ENTEROPNEUSTA</b> |                                      |                                   |               |
|                      | <i>Saccoglossus kowalevskii</i>      | XM_006823210.1                    | Genbank       |
| <b>ARACHNIDA</b>     |                                      |                                   |               |
| Araneae              | <i>Stegodyphus mimosarum</i>         | X975_00230                        | Ensembl       |
| <b>ANTHOZOA</b>      |                                      |                                   |               |
| Scleractinia         | <i>Acropora digitifera</i>           | XM_015904053.1                    | Genbank       |
| <b>ECHINODERMATA</b> |                                      |                                   |               |
| Echinoida            | <i>Strongylocentrotus purpuratus</i> | XP_791569.2                       | GenBank       |
| <b>ASCIDIACEA</b>    |                                      |                                   |               |
| Enterogona           | <i>Ciona intestinalis</i>            | XM_009859184.2                    | Genbank       |
| <b>CHLOROPHYCEAE</b> |                                      |                                   |               |
| Chlamydomonadales    | <i>Chlamydomonas reinhardtii</i>     | XM_001694533.1;<br>XP_001694585.1 | Genbank       |

#### 4.2 Correlación de las Fases de Intrones con la Sucesión de Dominios Cupredoxina.

El análisis de las secuencias en la delimitación de los intrones permitió la construcción de los mapas génicos que se muestran en la **Figura 2**. En cordados, 7 especies fueron seleccionadas para la realización de los mapas: *Homo sapiens*, *Gallus gallus*, *Anolis carolinensis*, *Xenopus tropicalis*, *Callorhinchus milii*, *Danio rerio* y *Ciona intestinalis*. En hemicordados, la especie *Saccoglossus kowalevskii* como único representante; en moluscos, artrópodos y cnidarios, las especies *Crassostrea gigas*, *Stegodyphus mimosarum* y *Acropora digitifera*, respectivamente. Como organismo unicelular, el alga verde *Chlamydomonas reinhardtii*. Los genes ortólogos de Cp- Ceruloplasmina: Heph y Heph1- Hefaestina y Hefaestina 1, FVIII- Factor de coagulación VIII y FOX1- Ferroxidasa de algas proporcionan

una ancestría en eucariotas que se corroboró en los alineamientos múltiples de aminoácidos y la construcción de las relaciones filogenéticas entre ellos.

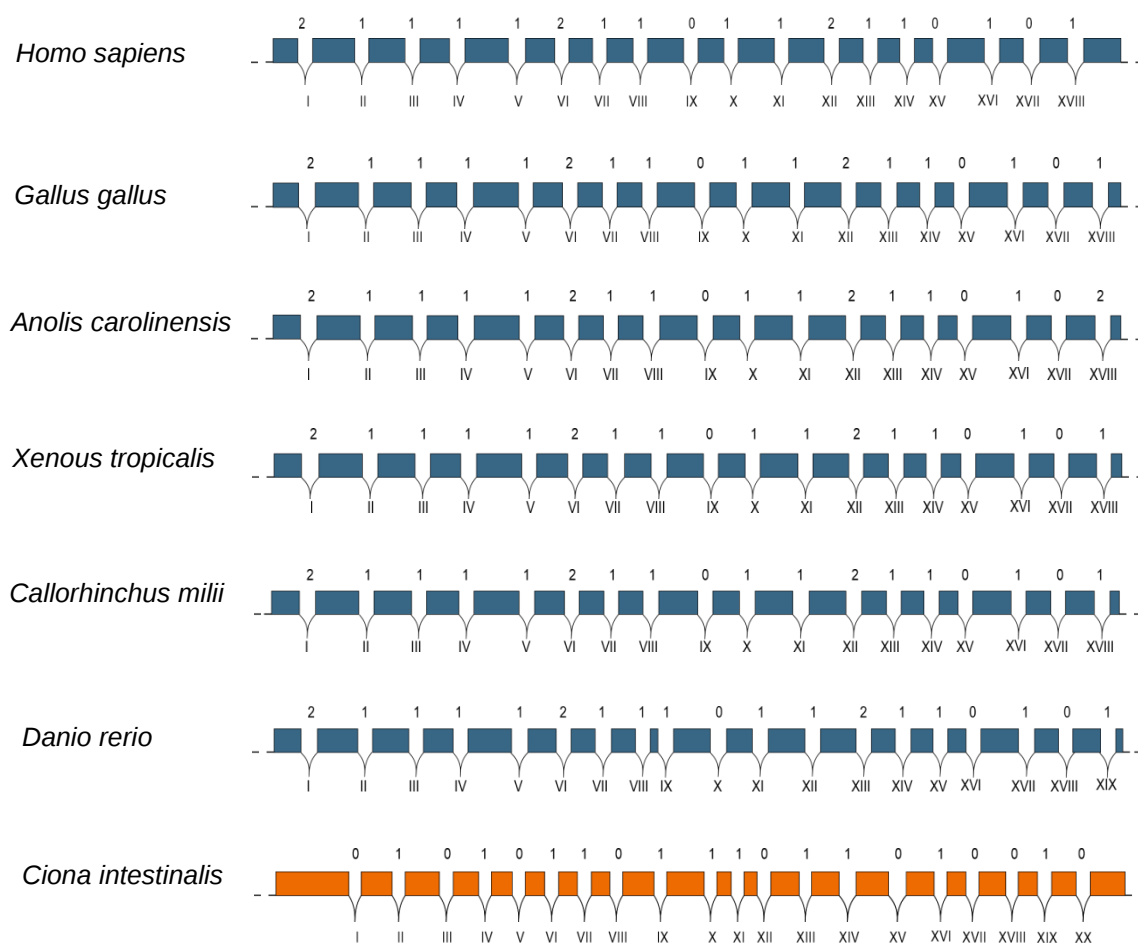
Como puede apreciarse en la **Figura 2**, el gen de la ceruloplasmina en los vertebrados presenta 19 exones y 18 intrones, manteniéndose de manera estricta el número y la posición de los intrones. Sin embargo, se encontró que *D. rerio* y *C. milii* han conservado el intrón VIII. Por otra parte, *C. mili* ha sido el único en perder el intrón IX que se encuentra en *D. rerio*, el cual pasó a ser el VIII de los demás vertebrados después de la pérdida del primer intrón en el linaje. Aunque el alineamiento múltiple evidencia un grado de conservación de intrones con los grupos no vertebrados eucariotas, la disposición varía conforme disminuyen las relaciones evolutivas (**Figura 3**), producto de las especiaciones y las dinámicas de inserción/delección que debió sufrir el ancestro eucariota durante la radiación del linaje.

En el alineamiento múltiple de los residuos de aminoácidos que se exponen en la **Figura 3**, se observan los intrones conservados de cada gen usado en el trabajo y se incluye el gen ferroxidasa (FOX) de la especie *Strongylocentrotus purpuratus* para el análisis. Las especies representan los grupos y los genes de la **Figura 2**.

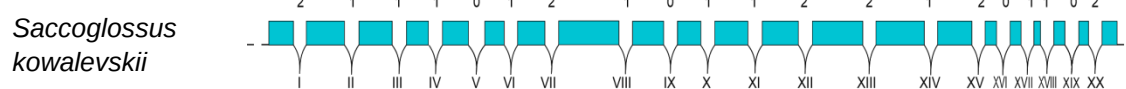
**4.2.1 Intrones Ancestrales.** Los intrones conservados entre los diferentes grupos corresponden a los intrones compartidos en cada ancestro eucariota de CP en vertebrados (**Tabla 2**). El 71% de los eucariotas que se usaron en el alineamiento múltiple, comparten con vertebrados los intrones I, II, III, XI, XII; el 57% conservan los intrones VI, VIII, IX, X, XVI; 43% conservan los intrones V, XVIII; 29% conservan el intrón XVII y el 14% conservan los intrones IV, VIII y XIV. La estructura de todos los genes a excepción del gen FOX, se encuentran esquematizados en la **Figura 2**.

## DIVISIÓN GÉNERO - ESPECIE

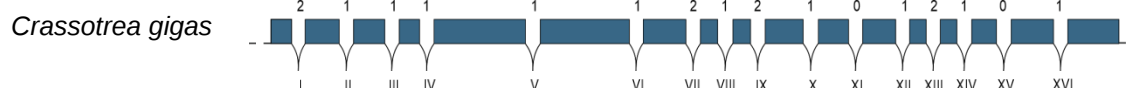
## CHORDATA



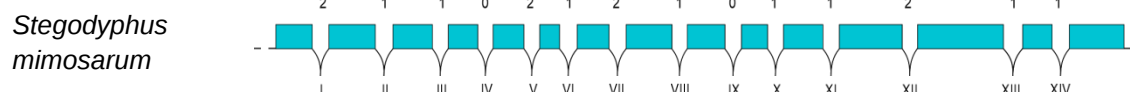
## HEMICHORDATA



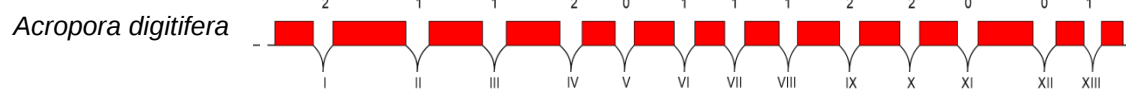
## MOLLUSCA



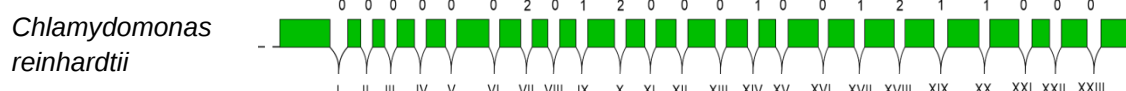
## ARTHROPODA



## CNIDARIA



## CHLOROPHYTA



■ Ceruloplasmina ■ Hefastina 1 ■ Factor VIII ■ Hefastina ■ FOX1

Figura 2. Organización de la estructura génica entre 12 especies de organismos eucariotas.

**Nota:** Genes pertenecientes a la familia de las proteínas Multicobre Oxidasas, MCOs: CP; ceruloplasmina, Heph1; Hefastina 1, FVIII; Factor de coagulación VIII, Heph; Hefastina, FOX1; multicobre ferroxidasa. En números arábigos: fase del intrón (0,1 ó 2). En números romanos: sucesión de intrones (I, I + n). Los exones están dibujados como casillas, cuyos colores representan el gen respectivo. Los tamaños de los intrones no se encuentran escalados, pero el tamaño de los exones es proporcional a su secuencia.

**Tabla 2.**

Conservación de intrones entre los eucariotas implementados en el alineamiento respecto a la estructura del gen CP en *H. sapiens*.

| CP_ <i>H. sapiens</i>           |   |   |   |   |   |   |   |      |    |    |    |    |    |    |    |    |    |    |
|---------------------------------|---|---|---|---|---|---|---|------|----|----|----|----|----|----|----|----|----|----|
| Número del intrón               | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8;8* | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| FOX1_C.<br><i>reinhardtii</i>   | - | - | - | - | - | - | - | -    | -  | -  | 17 | 18 | -  | -  | -  | -  | 23 | -  |
| FOX_S.<br><i>purpuratus</i>     | 1 | 2 | 3 | - | - | - | - | 7    | 8  | 9  | 10 | -  | -  | -  | -  | 12 | -  | 14 |
| FVIII_C.<br><i>intestinalis</i> | - | 2 | - | 4 | 5 | - | - | 8    | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  |
| HEPH_A.<br><i>digitifera</i>    | 1 | - | 3 | - | - | 4 | - | -    | -  | -  | -  | 9  | -  | 10 | -  | -  | 12 | 13 |
| HEPH1_S.<br><i>mimosarum</i>    | 1 | 2 | 3 | - | 6 | 7 | - | 8    | 9  | 10 | 11 | 12 | -  | -  | -  | 14 | -  | -  |
| HEPH1_S.<br><i>kowalevskii</i>  | 1 | 2 | 3 | - | - | 6 | - | 7    | 8  | 9  | 10 | 11 | -  | -  | -  | 13 | -  | 15 |
| CP_C. <i>gigas</i>              | 1 | 2 | 3 | - | 8 | 9 | - | 10   | 11 | 15 | 12 | 13 | -  | -  | -  | 16 | -  | -  |

**Nota:** En amarillo se indica el intrón conservado en peces y perdido en los demás vertebrados. En rojo se indica el intrón en todos los vertebrados pero que ha perdido *C. mili*.

En efecto, el alineamiento, permite rastrear procesos evolutivos que han sido dominados mayormente por la pérdida de intrones. En nuestro caso, aquellos intrones que con el tiempo han permanecido en el gen, representan las ganancias de estas recombinaciones.

**4.2.2 Cronología de la Delección e Inserción de Intrones de la Ceruloplasmina.** La ceruloplasmina posee dos tipos de dominios: la clase IV (dominios 1, 3 y 5) con *Identidad*, ID hasta de 53% y la clase V (dominios 2, 4 y 6) ID menores a 40% (**Tabla 3**), siendo los dominios de clase V los más ancestrales (Murphy *et al.*, 1997). Los dominios delimitados con sus respectivos intrones conservados entre dominios del gen CP en *H. sapiens* se encuentran en la **Figura 4**.

Han pasado aproximadamente 1082-1781 millones de años (M.A) desde el surgimiento de las algas verdes hasta la aparición de los primeros vertebrados (Hedges *et al.*, 2004; Parfrey *et al.*, 2011). Durante este tiempo eventos de ganancia y pérdida de intrones se han heredado en los eucariotas, dando lugar a la recombinación génica y diversificación de su genoma actual (Rogozin *et al.*, 2005).

Desde la divergencia en algas, han permanecido conservados en los vertebrados 3 intrones en los dominios 4 (XI), 5 (XII) y 6 (XVII). Alrededor de 611-1035 M.A (Lynch *et al.*, 1999; Peterson *et al.*, 2004) con la aparición de los cnidarios, la ceruloplasmina sufre inserciones en los dominios 1 (I, III), 3 (VI), 5 (XIV) y 6 (XVIII), ganando al menos 1 intrón en cada dominio.

|                    |                                                            |
|--------------------|------------------------------------------------------------|
| Chlamydomonas      | MDSKEAAEPASVHVNVDEAQAQ-----AQAEAAAKGGACATSGMSKGGKIIIVTS    |
| Acropora           | ----MYFAPAGVALQCHFQFSLATSLFPAVLQTDNRPQVDNLQQRQPPLSSGAIMLLY |
| Strongylocentrotus | -----MAPRV                                                 |
| Ciona              | -----MDIV                                                  |
| Crassostrea        | -----MTMHGIQLRLYLHLSKYVFM---                               |
| Stegodyphus        | -----MSST-                                                 |
| Saccoglossus       | -----MKGLQ                                                 |
| Danio              | -----MKVT                                                  |
| Xenopus            | -----MKIL                                                  |
| Homo               | -----MKLY                                                  |
| Gallus             | -----MKLV                                                  |
| Anolis             | -----MTRPV                                                 |
| Callorhinchus      | -----                                                      |

|                    |                                                                |
|--------------------|----------------------------------------------------------------|
| Chlamydomonas      | LVIFLGVAVGVLGVGLGVLKKDDGSSAYTSLDLGTSGGGNTYFVAADKIQWNYAPSG<br>0 |
| Acropora           | LAFSVAV-----LDSCLPVVLSITREYYIAAMVREWNYPATG                     |
| Strongylocentrotus | VCIIATLA-----LIGCMFQGACAVDRVYIIAAVEEWDYAPSG                    |
| Ciona              | -----                                                          |
| Crassostrea        | WSSVLC-----LSAIIITISQGREVDLYVGIDEVEWYKQTD                      |
| Stegodyphus        | VCFFACL-----VETQRIPEEDRPRTIYIGAVEVDWDYAPSG                     |
| Saccoglossus       | --TVSAL-----LFLALCRASQAIIRTHYIACVEEHWNYPATG                    |
| Danio              | --FALL-----GVLCCAGIASSITREYFFAIKEIQWNYAPSG                     |
| Xenopus            | --ALCL-----LCLLHLGFAAAEEERIFYIGIRETNWDYAPSG                    |
| Homo               | --ILGI-----FLFLCSTPAWAKEKHYYIGIETTDWDYASDH                     |
| Gallus             | --LLNV-----VLFFCCRTGAVTREYYIGITETAWDYAPGS                      |
| Anolis             | --VLSH-----LLACFFAVAWAVNKEYYLGIVETDWNYPATG                     |
| Callorhinchus      | --VLCL-----IVLHLTASVRGMTREHYIAIKEIENWYAPSN                     |

|                    |                                                                     |
|--------------------|---------------------------------------------------------------------|
| Chlamydomonas      | RNKCFF----PDLAAYKYLAMPQGITRVGGTFAKAIYRAYTDSSFNILATTPAEWQHGLN<br>0 0 |
| Acropora           | LNLIKGVPLNDSSEAKPFT--QGGAGRIATVYKKILYREYTDGTFTEQIHPH--GHLGI<br>2    |
| Strongylocentrotus | MNNITGAAFQPNTEETDKM--VRSATRIGRVYKKALFREYTDATFTERSPRP--VWLGM<br>2    |
| Ciona              | -----MKAHYVEYTDNFIKTKPA--SNSHHGA<br>0                               |
| Crassostrea        | N-----SDVVRKAVYQEYTDQSFSNRKPVV--SQQGL<br>2                          |
| Stegodyphus        | NLLGPGA----DEYKDLYLSQSSDPLKLGKVKYRKAFLREYTDFTTEQKPRD--PWMGI<br>2    |
| Saccoglossus       | LNVTGLPLEEDEYARRNM--VTDTRIGAKYKKALYREFTDDFTVQKVPV--GWLGA<br>2       |
| Danio              | KNLIQNKTVQQDEAARVFL--ERGEQRIGRVYKKAVYHQYTDATYRQEIDKP--KWLGY         |

|                    |                                                                                                                                                                |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Xenopus            | 2<br>NNIISGTSIKDD <sup>2</sup> ED <sup>2</sup> AQTFL--VGGPDRIGRVYKKAVYLQYSDDSYSKEIKKP---QWLGH                                                                  |
| Homo               | 2<br>GEKK--LISVD <sup>2</sup> TEHSNIYL--QNGPDRIGRLYKKALYLQYTDETFRTTIEKP---VWLGF                                                                                |
| Gallus             | 2<br>TDLISGQRF <sup>2</sup> AEE <sup>2</sup> EQ <sup>2</sup> AEVFL--KRG <sup>2</sup> PQRIGSTYKKAVYTQYTSTLYDVIVEKP---SWLGF                                      |
| Anolis             | 2<br>KDKISGKLFSEE <sup>2</sup> EH <sup>2</sup> ASVFV--QQGPDRIGSTYKKAVYVQFTDGSYKTAVEKP---SWLGL                                                                  |
| Callorhinchus      | 2<br>KNVITGTTVEAD <sup>2</sup> PH <sup>2</sup> ASIYL--KRG <sup>2</sup> PTRIGGVYKKAVFLEYTDGSYNKYVPKA---NWLGF                                                    |
|                    | 2<br>* :. : *                                                                                                                                                  |
| Chlamydomonas      | VGPVMYGAVGQVIRVVFKN <sup>0</sup> N <sup>0</sup> LD <sup>0</sup> FPVNMAPSGGLIA-----WDGNRRSARIDPVKPG                                                             |
| Acropora           | LGPVVRGEVGD <sup>0</sup> TIKVHFKNFAKRPLTI <sup>0</sup> HPH <sup>0</sup> GVLYSKANEGAIYADQNSNTDQKRDDSVQTN                                                        |
| Strongylocentrotus | MGPVIKAEVGDKVIVHFRN <sup>1</sup> MASRPYSINPGLGYTKSAE <sup>1</sup> GA <sup>1</sup> LYNDGTSGNQKADDTVQPN                                                          |
| Ciona              | LGPVIRVEIGEVLRI <sup>1</sup> VFKNSASRN <sup>1</sup> NFL <sup>1</sup> LDGFPFNRT <sup>1</sup> NL <sup>1</sup> LGVS <sup>1</sup> PD-----G                         |
| Crassostrea        | LGPVLRASVGETL <sup>1</sup> KIHVRNNAKRSFSFYTDGLKTQKNI-----T <sup>1</sup> GS <sup>1</sup> ATLGSPVNPG                                                             |
| Stegodyphus        | LGPI <sup>1</sup> LQGEVDILNIVFYNMASH <sup>1</sup> PVSI <sup>1</sup> HPH <sup>1</sup> GVRYQKIHE <sup>1</sup> GA <sup>1</sup> LYDDNTDAEKIDDRVQPG                 |
| Saccoglossus       | LGPVLKGEVGDEI <sup>1</sup> HVYFNNQCGRGYSM <sup>1</sup> HPH <sup>1</sup> GVMYTKDSE <sup>1</sup> GA <sup>1</sup> IYHDNTKGADKADDNVLPG                             |
| Danio              | LGPLISAEEDTVI <sup>1</sup> VHLKNMARRAYS <sup>1</sup> L <sup>1</sup> HAH <sup>1</sup> GLSYNKTNE <sup>1</sup> GA <sup>1</sup> LYPDSSEKVEKHDDSVAPG                |
| Xenopus            | LGPI <sup>1</sup> LKAEIGDIIIVHLKNFASRPYSI <sup>1</sup> HPH <sup>1</sup> GVQYTKENE <sup>1</sup> GA <sup>1</sup> LYPDNASSG-HKDDAIEPG                             |
| Homo               | LGPI <sup>1</sup> IKAETGDKVYVHLKNLASRPYTF <sup>1</sup> HSH <sup>1</sup> GITYYKEHE <sup>1</sup> GA <sup>1</sup> IYPDNTDFQRADDKVYPG                              |
| Gallus             | LGPI <sup>1</sup> IKGEVGDSIVIHLKNFASRNYTL <sup>1</sup> HPH <sup>1</sup> GVKYTKENE <sup>1</sup> GA <sup>1</sup> FYPDNTKGFEKRDDAVKPG                             |
| Anolis             | LGPVLKAEVGDSFTVHLKNFASRAYTL <sup>1</sup> HPH <sup>1</sup> GVKYTKENE <sup>1</sup> GA <sup>1</sup> LYPDMTSGLLKKDDIVQPG                                           |
| Callorhinchus      | LGPVLKAEQGD <sup>1</sup> KMIIHLKNFASRCYSI <sup>1</sup> HPH <sup>1</sup> GLTYLKAKE <sup>1</sup> GA <sup>1</sup> LYPDNSTDVEKLDDCVQPG                             |
|                    | 1<br>:**: .: . : . * .: *                                                                                                                                      |
| Chlamydomonas      | QTVTYLW <sup>0</sup> QI <sup>0</sup> PEDAGPVANATVTSRLWLYRSSVDPQKHDNAGLVGP <sup>0</sup> IIIVTSAANADAN---                                                        |
| Acropora           | QVQVYTW <sup>0</sup> VFVNEEHAPTAA-DEDCVTRMYH <sup>0</sup> SH <sup>0</sup> VISSQDVNSGLI <sup>0</sup> GP <sup>0</sup> LLICKRGTLIENS--                            |
| Strongylocentrotus | TTFIYRWNTGKAGPK <sup>1</sup> QQQ-TENCIGWQYT <sup>1</sup> SH <sup>1</sup> RFPSQDQ <sup>1</sup> NAGLVGITI <sup>1</sup> ICRO <sup>1</sup> GF <sup>1</sup> LAS---- |
| Ciona              | NITIYQWMITQDN <sup>1</sup> IPTLD-NIQCLTRMYYSNVYPIRDISTGLVGPFLICRSGSLGEN---                                                                                     |
| Crassostrea        | TTLTYYIYIVIDNGPRSGSSAGDCVTS <sup>1</sup> LYYSDVNFLQDTNSGLVGM <sup>1</sup> LIVCQ <sup>1</sup> PV <sup>1</sup> SAIKAEFD                                          |
| Stegodyphus        | SRHVYRWLMNEHDAPTDS-DPDCLTW <sup>1</sup> MYH <sup>1</sup> SH <sup>1</sup> RHGETELHAGLLG <sup>1</sup> PLLVCKK <sup>1</sup> GA <sup>1</sup> LTN----               |
| Saccoglossus       | ETHLYEWIVDGHSAPQED-DPNCLTW <sup>1</sup> AY <sup>1</sup> SH <sup>1</sup> FDEPKDANAGLYGALLTCKP <sup>1</sup> NT <sup>1</sup> LDAE---                              |
| Danio              | TTFTYIWTLPASHAPGKD-DTNCLTRI <sup>1</sup> Y <sup>1</sup> SH <sup>1</sup> VNAPKDIASGLIGPLIVCKK <sup>1</sup> GS <sup>1</sup> LDV----                              |
| Xenopus            | QSYTYRWEAVSEQGPTVQ-DEDCITRI <sup>1</sup> Y <sup>1</sup> SH <sup>1</sup> IDGPKDVSSGLIGPLI <sup>1</sup> ICKK <sup>1</sup> GS <sup>1</sup> INA----                |
| Homo               | EQYTYMLLATEEQSPGEG-DGNCVTRI <sup>1</sup> Y <sup>1</sup> SH <sup>1</sup> IDAPKDIASGLIGPLI <sup>1</sup> ICKK <sup>1</sup> DS <sup>1</sup> TK----                 |
| Gallus             | GQYTYTWDVTEDQGPAGK-DADCITRVY <sup>1</sup> SH <sup>1</sup> IDAPRDVASGLVGP <sup>1</sup> LI <sup>1</sup> ICKK <sup>1</sup> GA <sup>1</sup> MNK----                |
| Anolis             | EQYVYTWEVTEDHGPAGG-DNNCVTRI <sup>1</sup> Y <sup>1</sup> SH <sup>1</sup> IDAPKDIASGLIGPLITCKK <sup>1</sup> GT <sup>1</sup> LQD----                              |
| Callorhinchus      | SNFSYIWCLTEDHAPTIA-ETDCVTRI <sup>1</sup> Y <sup>1</sup> SH <sup>1</sup> THATKDISTGLIGPL <sup>1</sup> LICKK <sup>1</sup> GT <sup>1</sup> LTGKLK-                |
|                    | 1<br>* . * * . : * * :                                                                                                                                         |
| Chlamydomonas      | GRARD-----VDRD-----                                                                                                                                            |
| Acropora           | NMNRN-----VDKE-----                                                                                                                                            |
| Strongylocentrotus | -----TDKD-----                                                                                                                                                 |
| Ciona              | GE <sup>0</sup> QRN-----VDKE-----                                                                                                                              |

|                    |                                                                |
|--------------------|----------------------------------------------------------------|
|                    | 0                                                              |
| Crassostrea        | GTCRIIKDLEEFTHVLTLPQSGKDIIYKFQMTGKYPNAKPNVIIRCPALTEGQIEEMVKC   |
|                    | 1                                                              |
| Stegodyphus        | ---GQ-----IDRR-----                                            |
| Saccoglossus       | GNRLD-----VDHD-----                                            |
| Danio              | HGDKT-----GDYL-----                                            |
| Xenopus            | V-EQH-----KYKE-----                                            |
| Homo               | EKEKH-----IDRE-----                                            |
| Gallus             | DNDKY-----VDAE-----                                            |
| Anolis             | GKDKD-----VDRE-----                                            |
| Callorhinchus      | ASKKE-----VDEE-----                                            |
| Chlamydomonas      | -----                                                          |
| Acropora           | -----                                                          |
| Strongylocentrotus | -----                                                          |
| Ciona              | -----                                                          |
| Crassostrea        | MDTEADGTPMIQTMLTKAEQYLSANNINSSSKPKSGSGGNHENTGRKNKEKRKRKEKQE    |
| Stegodyphus        | -----                                                          |
| Saccoglossus       | -----                                                          |
| Danio              | -----                                                          |
| Xenopus            | -----                                                          |
| Homo               | -----                                                          |
| Gallus             | -----                                                          |
| Anolis             | -----                                                          |
| Callorhinchus      | -----                                                          |
| Chlamydomonas      | -----VVA---IFQIVQE---RASPL-----                                |
|                    | 0                                                              |
| Acropora           | -----FVI---LFTEFDE---NKS WL-----                               |
| Strongylocentrotus | -----IVL---NYGHIDE---NQSWY-----                                |
| Ciona              | -----IFL---HFSTINE---NLNWF-----                                |
| Crassostrea        | EEEVKDEKKASMKTTADDVIKRIQWDAELPQDEFLVG YIDRFLGIQEKYFTSFSWEDIASV |
| Stegodyphus        | -----IVL---VFMNFDE---NKSTY-----                                |
| Saccoglossus       | -----FTL---LLTMMDE---NQSWY-----                                |
| Danio              | -----YTL---MFTVSDE---NLSWY-----                                |
| Xenopus            | -----FVL---MFSVIDE---NLSWY-----                                |
| Homo               | -----FVV---MFSV VDE---NFSWY-----                               |
| Gallus             | -----FIL---MFSVMDE---NLSWY-----                                |
| Anolis             | -----FIL---MFSVMDE---NFSWY-----                                |
| Callorhinchus      | -----FAL---MFSV VDE---NNSWY-----                               |
|                    | : *                                                            |
| Chlamydomonas      | -----LFQEDT-----                                               |
| Acropora           | -----LDDNIE-----                                               |
| Strongylocentrotus | -----IDENIR-----                                               |
| Ciona              | -----INETIS-----                                               |
| Crassostrea        | DYSVLAVPKHRIEYFKYKDIKVWDKPRRIDNVFGSTGNPDVAQTTEGIQDTILEHESRFE   |
|                    | 1                                                              |
| Stegodyphus        | -----IDENIN-----                                               |
| Saccoglossus       | -----LDENIE-----                                               |
| Danio              | -----LDENIK-----                                               |
| Xenopus            | -----LDENIN-----                                               |
| Homo               | -----LEDNIK-----                                               |
| Gallus             | -----LDDNIR-----                                               |
| Anolis             | -----LDENIK-----                                               |
| Callorhinchus      | -----LDQNIK-----                                               |
|                    | .                                                              |
| Chlamydomonas      | -----SLSAG-----                                                |
| Acropora           | -----KYANN-----PGAI-----                                       |
| Strongylocentrotus | -----TYASG-----VTP-----                                        |
| Ciona              | -----ATTIG-----SM-----                                         |
| Crassostrea        | SCCMPLHCLHMTLCTIGLDTPEQVEHCVSVLNRLKPELASMAPKNIP IILKGIGHFFNRA  |
| Stegodyphus        | -----RHAPA-----MPMEER-----                                     |
| Saccoglossus       | -----TYCLV-----PESV-----                                       |
| Danio              | -----TYCSA-----PAKV-----                                       |
| Xenopus            | -----TFCTE-----PSSV-----                                       |
| Homo               | -----TYCSE-----PEKV-----                                       |
| Gallus             | -----MYCSE-----PSKV-----                                       |
| Anolis             | -----KYCSE-----PDKV-----                                       |
| Callorhinchus      | -----KFCTT-----PSKV-----                                       |

|                    |                                                               |   |
|--------------------|---------------------------------------------------------------|---|
| Chlamydomonas      | -----                                                         |   |
| Acropora           | -----KAIKSSPE-----                                            |   |
| Strongylocentrotus | -----AMKDSKE-----                                             |   |
|                    | 1                                                             |   |
| Ciona              | -----VKLKQSPT-----                                            |   |
| Crassostrea        | LYAKIEGAKELYEFVDHIKLCFKNEGIEIRDNHEFVPHMTIIKVTRPVGNSTGQKYIPPW  |   |
| Stegodyphus        | -----ESIKKDAE-----                                            |   |
| Saccoglossus       | -----N--PDDEN-----                                            |   |
|                    | 1                                                             |   |
| Danio              | -----N--KDDEA-----                                            |   |
| Xenopus            | -----D--KNDTD-----                                            |   |
| Homo               | -----D--KDNED-----                                            |   |
| Gallus             | -----D--KDDED-----                                            |   |
| Anolis             | -----D--KENED-----                                            |   |
| Callorhinchus      | -----N--RDDED-----                                            |   |
| <hr/>              |                                                               |   |
| Chlamydomonas      | -----                                                         |   |
| Acropora           | -----F--                                                      |   |
| Strongylocentrotus | -----F--                                                      |   |
| Ciona              | -----F--                                                      |   |
| Crassostrea        | LYSHKTDVDFGQQVLDNIYLCKMTHERRREDGFYVTPAYEKAKSTIFLHFGTFYEEESWYS |   |
|                    | 1                                                             |   |
| Stegodyphus        | -----F--                                                      |   |
| Saccoglossus       | -----F--                                                      |   |
| Danio              | -----F--                                                      |   |
| Xenopus            | -----F--                                                      |   |
| Homo               | -----F--                                                      |   |
| Gallus             | -----F--                                                      |   |
| Anolis             | -----F--                                                      |   |
| Callorhinchus      | -----F--                                                      |   |
| <hr/>              |                                                               |   |
| Chlamydomonas      | -----TSYTKMAINGYTWCNMPDGAITTKTGERVRWHVASIGSSSESLNHFHWCHV      |   |
|                    | 2                                                             |   |
| Acropora           | -----IASNKMASINGYIYGNLPG--LDMKQGDKISWHIIGVGWSDIHSPSFCHTV      |   |
|                    | 1                                                             |   |
| Strongylocentrotus | -----IGSNLMYAVNGYAYGNLPG--LEVCGQDQIRWFIFNFGGETGHSVFFNGNPL     |   |
| Ciona              | -----YQSNLMSVNGFGYGNLMH--VNAHINDHVSLHVMCTGSEHDVHNIRIGNQTF     |   |
|                    | 1                                                             | 0 |
| Crassostrea        | SANLSASLRQTTFSTINGYTGGNLDP--IEVCLDRNVRLHMTSLGGQNDIHTVVLYGHQF  |   |
| Stegodyphus        | -----RASNLMHTINGYAFGNLPG--LELCLNDRVTWHLVGMGDVVDVHTASFCHV      |   |
|                    | 0                                                             |   |
| Saccoglossus       | -----IKSNMYSINGYMYGNMPS--LKTCAGMNISWHLIGMGTEVDHPIYFCHGNM      |   |
|                    |                                                               | 0 |
| Danio              | -----QESNKMHSINGYVYGNLDP--LSMCIGNKIHWHLFGMGNEVDIHSAFFHGGIIL   |   |
|                    | 1                                                             |   |
| Xenopus            | -----QESNKMHSINGYMFGNLPG--LSMCEGEHIHWYLFGMGNEADIHSAYFHGGVVL   |   |
|                    | 1                                                             |   |
| Homo               | -----QESNRMYSVNGYTFGSLPG--LSMCAEDRVKWLFGMGNEVDVHAAFFHGGAL     |   |
|                    | 1                                                             |   |
| Gallus             | -----QESNKMHSINGYMYGYLPD--LTMCVEDKVKWHLFGMGNEADIHSAYFHGGTL    |   |
|                    | 1                                                             |   |
| Anolis             | -----QESNKMHSINGYMYGYLPD--LKMCTQDKVKWYLFVGNEADIHSAYFHGGTL     |   |
|                    | 1                                                             |   |
| Callorhinchus      | -----QESNKMHSINGYVYGNLPG--LSMFMGDKVMWHLFGMGNEVDIHSAYFHGNML    |   |
|                    | 1                                                             |   |
|                    | . :*: : : .: .: * ..* ..                                      |   |
| <hr/>              |                                                               |   |
| Chlamydomonas      | ELN-----GHHVDQFTAIPATYSVNMVPDEPGTWMFHCNVNFMMDGGMVALYTVT       |   |
|                    | 0                                                             | 1 |
| Acropora           | NIS-----SHRNTATLLPATFLTAYMTAWNPGTWLLNCMTSQQFDKCMNSALFRVH      |   |
| Strongylocentrotus | ELD-----GKRADNIPPIPGPVNVVTMEAMNPGVWLLSSLSTTDALNNGSALYNVR      |   |
|                    | 0                                                             | 1 |
| Ciona              | FSEISALDSSSGHSGIGNIEIVPLSGKTFYFKPAVPKWKISTDNLKLFEINGILAYFKVE  |   |
|                    |                                                               | 1 |
| Crassostrea        | EVK-----SORIDSLSLYPGAAVQANFRTLEEGTQQLTTQHTSDNLHGRVIVS---      |   |
|                    | 2                                                             | 1 |
| Stegodyphus        | QWE-----GHKTDISISLLPAKFTTSYMVARRPGVWPIRCVEGSHAHSGMEAFFRVR     |   |
|                    | 2                                                             | 1 |
| Saccoglossus       | NYA-----HHTTDTIGLFPATFATGLMTPHNPLWMSKCANTHHYDYCMMAVLFDIQ      |   |

|                    |                                                      |                      |                       |                                          |                            |
|--------------------|------------------------------------------------------|----------------------|-----------------------|------------------------------------------|----------------------------|
| Danio              | TDK-----RQHVDTVSLFPATFVNVMVADNPGQWLLS                | CQVNDHLE             | AGM                   | 1                                        | QAFFEIK                    |
| Xenopus            | TFQ-----MYRVDTLSLFPATMIEGHMVTKNPGKWLTT               | CQVNDHLE             | AGM                   | 1                                        | QAIYEVK                    |
| Homo               | TNK-----NYRIDTINLFPATLFDAYMVAQNPGEWMLSC              | QNLNHLK              | AGL                   | 1                                        | QAFFQVQ                    |
| Gallus             | IER-----HHRVDITINLFPATFIDALMIPRNPGEWLLSC             | QVNDHIE              | GGM                   | 1                                        | QALFKVE                    |
| Anolis             | IER-----NHRVDTISLFPATFIDALMVPENDGEWLLSC              | QVNDHIE              | GGM                   | 1                                        | QAIFEVK                    |
| Callorhinchus      | IEK-----HHHADTFSLFSASFITAVMVPTKKGKWLSC               | QVNDHLE              | AGM                   | 1                                        | QALFEVK                    |
|                    | .                                                    | :                    | * *                   | *                                        |                            |
| Chlamydomonas      | GDPA-----PLPTGGVERVYYVRAQE                           | EVESYSGPNNTQAC       | AVP                   | 2                                        |                            |
| Acropora           | R-RKIGVAM-----PTISGGMTKVYYLAAEEIMWDY                 | GPSGINNIDGTP         |                       |                                          |                            |
| Strongylocentrotus | SDCGTTGGSGGGTTGGSGGGTTGGSGGGTASGKVVEHFIAIQEVDWNYAPFA | SK                   |                       | 2                                        |                            |
| Ciona              | GNTTNTDHT-----STNIHD                                 | DA                   | 1                     | TVRWYFIAAIEQTWYVSGENQG                   |                            |
| Crassostrea        | -TCNSSRPA-----QQLSGNIRYYYYIAAEREQ                    |                      |                       |                                          |                            |
| Stegodyphus        | Q-CNNSRPD-----EKEALKGILREVFIAAEDVEWNYAPSGFNFLTGTN    |                      |                       |                                          |                            |
| Saccoglossus       | D-CGLADAP-----KLTGLTDAVREYYIQAEVDVYWDY               | GPSGMNIFYGGP         |                       |                                          |                            |
| Danio              | K-CFPNVHK-----PRPFGEVRQYYIAAEEI                      | IWDYGPTEINQYTMKK     |                       |                                          |                            |
| Xenopus            | S-CPAQAE-----RESSRIGRERRYFIAAEEAMWNYAPTDTNQFTGQK     |                      |                       |                                          |                            |
| Homo               | E-CNKSSSK-----DNIRGK                                 | VRHYIIAAEEI          | IWNYAPSGIDIFTKEN      |                                          |                            |
| Gallus             | G-CRKSTID-----HNETTKIRQYFIAAEEI                      | IWNYGPSALNHFTGQE     |                       |                                          |                            |
| Anolis             | N-CKHHV-P-----SSRNTKIRRYIIAAEEI                      | IWNYAPSGLNFTGEE      |                       |                                          |                            |
| Callorhinchus      | D-DGTQHKR-----PTFSQKERHYIIAAEEI                      | IWDYGPTGRDQFTEKP     |                       |                                          |                            |
|                    |                                                      | :                    | :                     |                                          |                            |
| Chlamydomonas      | ELQFSSEPGSEEVNGNVFLEGSPSTDVRLGHIYT                   | KT                   | 0                     | LLIEYTDASFTTVKPRPADEQYLG                 |                            |
| Acropora           | LT---SPGS                                            | 2                    | ESAVHFT               | ---QADDRIGGRYKKALYVEYTDFTIKANRSYSELHLG   |                            |
| Strongylocentrotus | -----SPTGSGPVFLKAMFVG                                | YTDATYQRFNVRSQDEQYMG |                       |                                          |                            |
| Ciona              | -----G-----QMRGNNTVYK                                | KV                   | 0                     | LYEEFTDATFTHRKPVP---YNG                  |                            |
| Crassostrea        | -----GS                                              | 2                    | TPGPVIFQEYTGPTFKTKKPS | ---NLQING                                |                            |
| Stegodyphus        | LT---D-DS                                            | 2                    | YSAPFFV               | ---KGPHRIGGIYRKAVYREYTDATFTIQKPRSPEDLHLA |                            |
| Saccoglossus       | LN---EPGS                                            | 2                    | ESGTFE                | ---RGDYRIGGVYKKTIFVEYTDATFNTRKDRTVDDXHMG |                            |
| Danio              | LV---D-DI                                            | 2                    | VSDTFFD               | ---NRNDRIGGKYKKVQYVEYTDFTTKRKEKTPEEQHLG  |                            |
| Xenopus            | ID---HPDT                                            | 2                    | ESAIFFQ               | ---RNEVRIGGTYTKAVYVEYTDATFTTRKQRSKDEEHLG |                            |
| Homo               | LT---APGS                                            | 2                    | DSAVFFE               | ---QGTTRIGGSYKKLVYREYTDASFTRNKERGPREEHLG |                            |
| Gallus             | LI---A-DS                                            | 2                    | ESVFFE                | ---RSETRIGGSYKKAIYKEYTDGFTAHKKRLPEEEHLG  |                            |
| Anolis             | VI---H-DE                                            | 2                    | DAKTFFE               | ---QGDKTIGGTYKKALFYEYLSEAFKTRRRRLPIEEHLG |                            |
| Callorhinchus      | LD---YND                                             | 2                    | DSAVYFQ               | ---QGEKRIGGRYKKVVYQEYTDMMFTRQKERSPEQQHLG |                            |
|                    |                                                      | .                    | :                     | :                                        | .                          |
| Chlamydomonas      | LLGPVMRANVGDTIKV                                     | 0                    | VLKNDAKIDVSL          | HP                                       | GVRYSKANEGTLYEDGTSGA-----  |
| Acropora           | FLGPVIRAEVGDEINVFFKN                                 | QG                   | SRNYSI                | HP                                       | PLGVFYDKANEGAVYRDGTSG----- |

|                    |                                                               |
|--------------------|---------------------------------------------------------------|
|                    | 0                                                             |
| Strongylocentrotus | LIGPVIRAEVGDNIRVTLNMGATRAYNIHPNNVMYTKANEGYPYADGIPANQVSGDSVAP  |
| Ciona              | ILGPVKAIEYGDIIKVVFNHQAQCRPYSIHHPVKLMHSVEDNGLVWLEGDGIQ-----1   |
| Crassostrea        | RLGPLIYAETKDMVKVVFYNRADRNLTLYPHGLRVGKMFEGVLYDDVTGK-----1      |
| Stegodyphus        | IMGPTLRAEVDVLKITFKNLAAFPYTIDTHGAYYLRNGAA-----1                |
| Saccoglossus       | YLGPPIRAEVGDQIHVTFRNMANRSHSIHHPHGVFYDKANEGAVTRMELKVSRT-----1  |
| Danio              | ILGPVIRAEEDTIKVTFRNKASRPFSIQPHGVQYNIDMDGTLYHNVLVESYT-----1    |
| Xenopus            | ILGPVMIAQVGDSMKVTFKNKASRPYSIQAHGVSYSKNMEGASYRTGN--L-----1     |
| Homo               | ILGPVIWAEVGDITRVTFHNKGAYPLSIEPIGVRFNKNNEGTYYSNPYNPQ-----1     |
| Gallus             | LLGPVKAIEVGESIRVTFRNNASRPFSIQPHGVSYHKSNEGALYRT---A-----1      |
| Anolis             | ILGPIIRAIEVGETIQVTFRNKASYPFSIQAHGLSFSKDNESFFNT---S-----1      |
| Callorhinchus      | LLGPVKAIEVGDRVKVTFWNASHSYSIQAHGVAYEKSSEGAKYPTDTGPA-----1      |
|                    | : ** : * : : . : : * . : .                                    |
| Chlamydomonas      | -----                                                         |
| Acropora           | -----                                                         |
| Strongylocentrotus | GGTYTYNWTIPSSYGPTASDPDCITGIYYSTAGGDRDLHAGLVGPLIVCKPGKLSSNGTQ  |
| Ciona              | -----TD                                                       |
| Crassostrea        | -----                                                         |
| Stegodyphus        | -----                                                         |
| Saccoglossus       | -----IPS-----PSPDLNT-----ADILQ                                |
| Danio              | -----DK-----                                                  |
| Xenopus            | -----                                                         |
| Homo               | -----                                                         |
| Gallus             | -----                                                         |
| Anolis             | -----                                                         |
| Callorhinchus      | -----                                                         |
| Chlamydomonas      | -----DKADDVVAPGTTYTYV                                         |
| Acropora           | -----NDTADDAIMPNQTFLYR                                        |
| Strongylocentrotus | QRHATPATFDLTRYYVIGEVHRLSTVSQSRGSRERLPLLSPPNQVSGDSVAPGTRYTYN   |
| Ciona              | FRH-----RDP-----IFVNRGSVFTGRQTQYT                             |
| Crassostrea        | -----GAVPPNRIKTYF                                             |
| Stegodyphus        | -----KANETQVIY                                                |
| Saccoglossus       | YRHST-VVFFI-----SRWP--YLRAADKLDDAVPPGGVYTYM                   |
| Danio              | -----KLREL-----KKQP--RVIEPLPAAMVRPDTTYKYE                     |
| Xenopus            | -----TNDDPSPGSHVAPGDTFTYE                                     |
| Homo               | -----SRSVPPSASHVAPTETFTYE                                     |
| Gallus             | -----SRDTESPAHVSPGTTFTYE                                      |
| Anolis             | -----SGGNTRPSSHVNPGDTFYFE                                     |
| Callorhinchus      | -----SMS--PRSPLPVPPGQTFVYE                                    |
| Chlamydomonas      | WNVDPDR--AGPGPCDPSSMLWMYHSHIDETAETYAGVAGGIIVTAKDMARSTADLTPKDV |
| Acropora           | WTVPEA--VAPTEKDPPCITWMYHSNVEKTKDIESGLIGPLIVCKKGILNASRPWRKDV   |
| Strongylocentrotus | WTIPSS--YGPTASDPDCITGIYYSTAGGDRDLHAGLVGPLIVCKPGKLSSNG--TQASV  |
|                    | 0 1 1 0                                                       |

|                    |                                                                                                                                                        |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ciona              | WKVPPVADYDSSLNPSECAVSYYYSNADATRDVHSGLVG <sup>1</sup> PLLCLNQVLRTHN--IRYDI                                                                              |
| Crassostrea        | WSVPED--VSVSFLDDDCLVRYHTSGIGHGDD--SFLFGPLKICYDGF <sup>0</sup> SVS-----T                                                                                |
| Stegodyphus        | WTFLLQ--YGPSDLDPDCLPWRYYSQDFKRDAYAGLNGILLVCRKGTLDKNG--R <sup>0</sup> QKNV                                                                              |
| Saccoglossus       | WSVPTE--WAPLAGDGPCLTYVYYS <sup>0</sup> VHPTKDIYSGLIGPLIVCNQGG-----I <sup>0</sup> GVDF                                                                  |
| Danio              | WVVPKD--GGPTEKDPDCLTYLYYSAVDPIRDTNSGLVGPLLICKPKTL-KSG--K <sup>0</sup> QKNV                                                                             |
| Xenopus            | WEVSSA--VGSTNQNLNCLVWLYFSSVDAIKDTNSGLVGPLLICKS--IDANN--H <sup>0</sup> QGI                                                                              |
| Homo               | WTVPKE--VGPTNADPVCLAKMYSAVDPTKDIFTGLIGPMKICKKGS <sup>0</sup> L <sup>0</sup> ANG--R <sup>0</sup> QKDV                                                   |
| Gallus             | WTVPED--VGPTDQDPDCLTWLYYSAVNSVRDTSSGLVGPLMVCRKGALLPSA--K <sup>0</sup> QKTV                                                                             |
| Anolis             | WEVPET--VGPTPEDPDCLTRMYYSASDPIKDTSAGLVGPLLVCRKGALLPSG--R <sup>0</sup> QKNV                                                                             |
| Callorhinchus      | WRIPED--YGPSRSEGDCLTWLYFSAVDPVKDTNSGLVGPLLICKPNTLDKKG--K <sup>0</sup> QKNV                                                                             |
|                    | * . . * * : : : * : :                                                                                                                                  |
| Chlamydomonas      | DREIVIFFTVVDEIKSSN <sup>0</sup> F <sup>0</sup> MENLANKLGDGGALAAQLAANATEMTALVTDVPVMEHMLKH                                                               |
| Acropora           | QHDFFLRFGLNDESNWYFNENKNL-AGNAST-----INE <sup>1</sup> VDPGFSASNKMR                                                                                      |
| Strongylocentrotus | DQEVVTLILSMNEERSWYWPRNRNV-----RYNSDFPV-EELR                                                                                                            |
| Ciona              | DREFYLLFK <sup>1</sup> E <sup>1</sup> FDENQSWYASENQHYL-----NQWNLNGS <sup>0</sup> N <sup>0</sup> IFH                                                    |
| Crassostrea        | EKELFVIHT-----KTSTDIE                                                                                                                                  |
| Stegodyphus        | DREYSLMFSALDENESHYIDDSVLAYTSDPYG-----VDFYDRDFRLANRRY                                                                                                   |
| Saccoglossus       | NREVFLLHFGVTDENQSWHIESNIQKFCGLPDD-----VDRQNEDFMESNLMH                                                                                                  |
| Danio              | DKEFHLLATVFDENLSWYLDNINRYAKQPKS-----VNKEDADFQESNKM                                                                                                     |
| Xenopus            | AHNYFLLPTVFDENESWYLDENIQQTGTGPAN-----VNKEDPDFQESNMMH                                                                                                   |
| Homo               | DKEYFLFPTVFDENESLLEDNIRMFTTAPDQ-----VDKEDEDFQESNKM                                                                                                     |
| Gallus             | TREFFLLATVFDENLSWYLDNINRMFTAKPNE-----IDKDNEDFQESNKM                                                                                                    |
| Anolis             | DKEYFLLATVFDENLSWYLDENIKHYTNTPE-----IDKEDEDFQESNMMH                                                                                                    |
| Callorhinchus      | DKEMYMLVTVFDENQSWYLEENIQMFTGRPRA-----VDKEDEDFQESNKM                                                                                                    |
|                    | ::                                                                                                                                                     |
| Chlamydomonas      | GINGHMYCHMPRLTFEQGD <sup>0</sup> K <sup>0</sup> YRLHVMVLGTLED <sup>0</sup> M <sup>0</sup> HTPNMGGPRFDYNG-----                                          |
| Acropora           | NINGYMYGNLPGMRMCKGDNVSWHLIGG--PGGM <sup>0</sup> HTATF <sup>0</sup> H <sup>0</sup> GN <sup>0</sup> TFLRNG-----                                          |
| Strongylocentrotus | <sup>1</sup> T <sup>1</sup> INGYVSENLP <sup>1</sup> TLGLCVGDHVRWHAPCIGNEETYMGL <sup>1</sup> H <sup>1</sup> IPYQTFEYKG-----                             |
| Ciona              | TINGLTHGGLVTLETETEPGERIAWHVINMDS--QP <sup>1</sup> H <sup>1</sup> SINFEGNSVLTYMNSASLLS <sup>1</sup> D <sup>1</sup> Y <sup>1</sup> LST                   |
| Crassostrea        | SINGVPSESLIPGPVCLKERLTFNLMTYGTREP-QSFVVGGSVIRQKG-----                                                                                                  |
| Stegodyphus        | <sup>1</sup> S <sup>1</sup> INGRMYGNLEGLSMCKGDRVVWNIMGVGRSIDL <sup>1</sup> H <sup>1</sup> TAYF <sup>1</sup> H <sup>1</sup> GNNFLHRG-----               |
| Saccoglossus       | <sup>1</sup> G <sup>1</sup> INGYLYANIPSLAFCTSDTIMFHLMTMGDAVDI <sup>1</sup> H <sup>1</sup> GAHF <sup>1</sup> H <sup>1</sup> GN <sup>1</sup> TLTWS-----  |
| Danio              | <sup>1</sup> S <sup>1</sup> INGYMYGNLKG <sup>1</sup> NLMCKGDKVSWHLSGLGSETDI <sup>1</sup> H <sup>1</sup> GLYFEGNRFYLYKE-----                            |
| Xenopus            | <sup>1</sup> S <sup>1</sup> INGYMYGNQ <sup>1</sup> PGLDMCAGDTVLWHLMGLGNEVD <sup>1</sup> M <sup>1</sup> H <sup>1</sup> GIHFSGNTIVTHY-----               |
| Homo               | <sup>1</sup> SM <sup>1</sup> NGFMYGNQ <sup>1</sup> PGLTMCKGDSVVWYLFSAAGNEADV <sup>1</sup> H <sup>1</sup> GIYFSGNTYLWRG-----                            |
| Gallus             | <sup>1</sup> S <sup>1</sup> INGYMYGNQ <sup>1</sup> PGLMCKGDVISWHLMGLGSEVD <sup>1</sup> M <sup>1</sup> H <sup>1</sup> GIYFSQNTFVTKG-----                |
| Anolis             | <sup>1</sup> S <sup>1</sup> INGYMYGNL <sup>1</sup> RGLMCSGETVSWHLIGLGSEVDI <sup>1</sup> H <sup>1</sup> GI <sup>1</sup> H <sup>1</sup> FSGNTFLTQG-----  |
| Callorhinchus      | <sup>1</sup> A <sup>1</sup> INGYLFGNQ <sup>1</sup> PGLMKCKGDLVSWHMF <sup>1</sup> GI <sup>1</sup> GSEVDI <sup>1</sup> H <sup>1</sup> GIYFNGNTFLSKE----- |

|                    |                                                        |                      |                     |                 |                     |                      |                      |   |   |
|--------------------|--------------------------------------------------------|----------------------|---------------------|-----------------|---------------------|----------------------|----------------------|---|---|
|                    | 1                                                      | : **                 | :                   | :               | .                   |                      |                      |   |   |
| Chlamydomonas      | -MHTDSIQISPGGMVSADVQMTSP                               | GD                   | YELO                | RVAD            | VMAG                | RAKYTVTA-NASRMVNV--  |                      |   |   |
| Acropora           | -NNHDTLGMSE                                            | GS                   | LLTMQMTPDNPGEWKL    | RVNSNLRGG       | VAKYIVQE-DCGTAYTVNQ | 1                    |                      |   |   |
| Strongylocentrotus | -RRANTLPLVPGQTTEEVDFTVGSP                              | GT                   | WIINGN--            | VYNYDLEVNYTVS-- | DCYRTLDSQK          | 1                    |                      |   |   |
| Ciona              | VSSSVSRPLLPGASTTVMVPTNPGAWMVQDFASQSSVMG                | MYCIYHVRPK           | EC                  | RKPLGLES        | 1                   |                      |                      |   |   |
| Crassostrea        | -HVLNVLAAEENSMSYKVTfVKt                                | GN                   | WSVRSFRNENV----     | SFIFSVD--       | DCGETIDFV-          | 1                    |                      |   |   |
| Stegodyphus        | -IYRDTVAVVPSILYSVSMIAENE                               | GD                   | WIE                 | CRTNS           | HLTGG               | MRALYRVE--DCTNKRQKRQ |                      |   |   |
| Saccoglossus       | -NHVDTIPLTPSILQHAIMQPKNI                               | GD                   | WAFV                | CRTND           | HYRAG               | MGLYSVN--DCAKANSIES  |                      |   |   |
| Danio              | -TRRDTINVFPHISHTVIMEPDSM                               | GQ                   | FEVN                | CKTTD           | YHGG                | MRANYTVE--KCRFWNRQSE |                      |   |   |
| Xenopus            | -TRADTASLFPHISYSVMNPDNE                                | GV                   | FNVE                | CMTTD           | HYTAG               | MRQHYRVK--SCGKQSF--S |                      |   |   |
| Homo               | -ERRDTANLFPQTS                                         | LTLMHWPDE            | GT                  | FNVE            | CLTTD               | HYTGG                | MKQKYTVN--QCRRQSEDST |   |   |
| Gallus             | -ARKDTANLFPHTFVTAIMKPDSE                               | GI                   | FEVS                | CLTTD           | HYRGG               | MKQKYVKV--PCHWRNVDP  |                      |   |   |
| Anolis             | -TRKDTTNLFPHTSVTAIMKPDSE                               | GI                   | FEVA                | CLTTD           | HYTAG               | MRQKYEVR--RCDTSPENV  |                      |   |   |
| Callorhinchus      | -TRRDTVNIFPHISLTAIMQPDST                               | GV                   | FDVA                | CQTTD           | HFQGG               | MKQHYTVD--KCSFWSGFPT |                      |   |   |
|                    | .                                                      | .                    | .                   | *               | :                   | .                    | :                    | * | . |
| Chlamydomonas      | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Acropora           | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Strongylocentrotus | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Ciona              | GLIADHQITSSSVYITKAAGFGIMGKWSARYARLNTHGHINAWMPSRSDRRPWI | QV                   | DLGS                | 0               |                     |                      |                      |   |   |
| Crassostrea        | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Stegodyphus        | VFNDINEISSHVEYYND-----                                 |                      |                     |                 |                     |                      |                      |   |   |
| Saccoglossus       | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Danio              | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Xenopus            | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Homo               | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Gallus             | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Anolis             | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Callorhinchus      | -----                                                  |                      |                     |                 |                     |                      |                      |   |   |
| Chlamydomonas      | -----PSGVTRTYI-----                                    | QAEAVNWDYAPAGYQ      | KC                  | TDT-D--         | FSYQSSVY            | 2                    |                      |   |   |
| Acropora           | -----MKGRKRQYFI-----                                   | AAVEETWDYAPSGQDL     | IEGK-QLEYS          | VF              | ASRY                | 2                    |                      |   |   |
| Strongylocentrotus | -----AMKKERTIFL-----                                   | AAEEVIWDFSPT         | EYDYIRNQSL-----     |                 |                     |                      |                      |   |   |
| Ciona              | PMRITGITVQGTMDGHVRKYVTRFKILYSVNNREWAFYSRQGSNV-----     | DMEFK                | GNLF                | 1               |                     |                      |                      |   |   |
| Crassostrea        | -----YSSKEYSSFI-----                                   | AIGAGKWSYSKVT        | DWAS                | SK              | LP-----             | 2                    |                      |   |   |
| Stegodyphus        | PLADAKTKLQPAMGGKIREYFV-----                            | AAVDEWEYAPRDRHLVVG   | G-SLMDN             | EE              | SKTF                | 2                    |                      |   |   |
| Saccoglossus       | -----AKHVDAIPQAMTRTYI-----                             | AAVEEWDYAPNGYDRITN   | SSLYNP              | NS              | GYIF                | 2                    |                      |   |   |
| Danio              | -----TMLHQKKYI-----                                    | AAVEMDWDYSPTRTWEDKM  | H-NGLKE             | SP              | GNEF                | 2                    |                      |   |   |
| Xenopus            | -----LLFRNTTYI-----                                    | AAEEIEWDYSPNRTWEHERH | -IHHGE              | SP              | GDAF                | 2                    |                      |   |   |
| Homo               | -----FYL                                               | ERTYI-----           | AAVEEWDYSPQREWEKELH | -HLQE           | QNV                 | SNAF                 | 2                    |   |   |
| Gallus             | -----LYLHEKTHYI-----                                   | AAVEEWDYSPNRTWEFERH  | -QYHKE              | SP              | GNPF                | 2                    |                      |   |   |
| Anolis             | -----GYLQEKTYI-----                                    | AAVEIEWDYSPNRTWEHELH | -QFHDE              | SP              | GNAF                | 2                    |                      |   |   |
| Callorhinchus      | -----FTWHTKTYI-----                                    | AAVEIEWDYSPSRTWELEMY | -HSSEE              | SP              | GTIY                | 2                    |                      |   |   |
|                    |                                                        |                      |                     | *               | :                   |                      |                      |   |   |

|                    |                                                                                                                                                              |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Chlamydomonas      | LRRTSYTIGSRYRKAVYRAYTDATFSTRV---PTPAYYGTMGPMIIAEV <sup>1</sup> GVDRIVVHFKN                                                                                   |
| Acropora           | TIPGPHYIGRRNKKAVYREYTDSTFTKQKQRSKDEEYLGILGPLIKAEVGDIIIEVVFKNL                                                                                                |
| Strongylocentrotus | -----                                                                                                                                                        |
| Ciona              | -----FNPVTRT-----FSPQLLARYIRLVPIRWQNO                                                                                                                        |
| Crassostrea        | ----RFQMGPAPQKAMFEEYYSNFFLFAKT-PPNVADKVLHGPQLYAKVGDKIKL <sup>1</sup> DVNNK                                                                                   |
| Stegodyphus        | VERGNDRIGTRYKKAVFREFTDVTFTTRKCETPQTRHLAVLGPMIKAEVGDLLRVTFRNM                                                                                                 |
| Saccoglossus       | CYRGNTSIGHQYKKAVYKEYEDKDFLIEKPRTPEESHLHMLGPYIKAEIGDTEVVVFKNM                                                                                                 |
| Danio              | LKKEGFIGSKYKKVLYKEYTDEFTTKLKERTADMEHLGIM <sup>1</sup> GFMIHGKVGKEVKVIFKNM                                                                                    |
| Xenopus            | LNKGEKFIGSTYKKAVYRQYTDRTFTKRKERNEEEKHLGIL <sup>1</sup> GFLLIARTHEKVKIVFKNM                                                                                   |
| Homo               | LDKGEFYIGSKYKKVVYRQYTDSTFRVPVERKAEEHLGIL <sup>1</sup> GFQLHADVGDKVKIIFKNM                                                                                    |
| Gallus             | LNKDDKFIGSKYKKVVYREYTDQTFSSIPKSRAEEHQHLQIQ <sup>1</sup> GFLLIMSSIGDKINIVFKNL                                                                                 |
| Anolis             | LNKEDKFIGSKYKKAVYQEYTDQTFSPKERKEKEEHLGLI <sup>1</sup> GFLLVADVRDVKIVFKNM                                                                                     |
| Callorhinchus      | VVRGNTTIGSRYKKVVYREYTDQTFTKQKIRPAGEEHLEIQ <sup>1</sup> GFLLHAKVSDRIKIVFKNL                                                                                   |
| Chlamydomonas      | VTD-LEEYPLNISPGGGLLVEG-----AADENCAEVAAGETCVYRWIVPDSSGP                                                                                                       |
| Acropora           | ASR-----QYSI <sup>2</sup> HPGLFY <sup>2</sup> RKQDEGSQYKDSTAGADIVDNAIAPGKMYNEYEWEPHRAGP                                                                      |
| Strongylocentrotus | -----SDPN <sup>2</sup> RWTMLEKAGP                                                                                                                            |
| Ciona              | PAVRM <sup>0</sup> ELLGCSI-----EGRCEEPQGIGS                                                                                                                  |
| Crassostrea        | EKR-----QHSL <sup>1</sup> HVDGLS-----DQTNIAPGGSLQITWEINEAIGP                                                                                                 |
| Stegodyphus        | ASR-----PYSF <sup>1</sup> HPGILHSQDKNHRNF-----GSEDDPVFPGLDLYVEWLVPETAGP                                                                                      |
| Saccoglossus       | ASR-----SYSL <sup>2</sup> HPGLAYSKSSEGSFYMDGNVGGG---DAVQPGEVYNY <sup>2</sup> LWTVPASSGP                                                                      |
| Danio              | ARR-----PYSI <sup>1</sup> HA <sup>1</sup> GVKTDSP-----QVALTRP <sup>1</sup> GE <sup>1</sup> TQTYTWYLPKNSGP                                                    |
| Xenopus            | ASR-----PYSIYA <sup>1</sup> GVKTDGP-----V-VQTEP <sup>1</sup> GQ <sup>1</sup> TQTYIWKIPERSGP                                                                  |
| Homo               | ATR-----PYSI <sup>1</sup> HA <sup>1</sup> GVQTESS-----TVTPTLP <sup>1</sup> GE <sup>1</sup> TLTYVWKIPERSGA                                                    |
| Gallus             | ASR-----PYSI <sup>1</sup> HA <sup>1</sup> GVKTDSS-----AVAITNP <sup>1</sup> GE <sup>1</sup> TKMYVWKISARSGS                                                    |
| Anolis             | ASR-----PYSI <sup>1</sup> HA <sup>1</sup> GVKTDSP-----LVVETSP <sup>1</sup> GA <sup>1</sup> IREYIWKIPKRS GP                                                   |
| Callorhinchus      | ATR-----PYSI <sup>1</sup> HANGVKTN SQ-----EVKPTLP <sup>1</sup> GS <sup>1</sup> TETYIWEIPERSGP                                                                |
|                    | *                                                                                                                                                            |
| Chlamydomonas      | GT- <sup>1</sup> AD <sup>1</sup> FNTA---VGYTSSVDVATAPSAGLAGALVVAGRQQLVAGPDGSLLPRGVDLMVP                                                                      |
| Acropora           | GP-KD <sup>0</sup> TDCV---SWGYSFSDVNPVKDTNSGLVGPLIICGK <sup>0</sup> GTLD <sup>0</sup> DKD---KIKSV <sup>0</sup> DREYV                                         |
| Strongylocentrotus | TD-SQSN <sup>0</sup> CV---VWPYYSSI <sup>0</sup> VT <sup>0</sup> KDTNSGLIAPFIVCRKGT <sup>0</sup> LNSQ <sup>0</sup> G---QRTD <sup>0</sup> VDH <sup>0</sup> DFA |
| Ciona              | GEVPDTSITASSEYLYKRR-----ILWEARFGR <sup>0</sup> LN <sup>0</sup> YR-----                                                                                       |
| Crassostrea        | GS-ADPSCI---MK <sup>0</sup> P <sup>0</sup> YYSRTS-QRDIPSGLFGPLIICSSTVPNDV-----GSNLQ <sup>0</sup> QKV                                                         |
| Stegodyphus        | GP-NGFNCT---PWAYYS <sup>1</sup> AVDPVKDTNAGLIGPLVICRRGILTES <sup>1</sup> C---A <sup>1</sup> RTDAEQEFV                                                        |
| Saccoglossus       | SD-YDSNCI---TWAYYS <sup>1</sup> AVNP <sup>1</sup> IKDTNSGLIGPLIVCRPGILDAEG---TRNDV <sup>1</sup> DREFA                                                        |
| Danio              | TE-EQE <sup>0</sup> ECS---VGAYYSTVDVI <sup>0</sup> KD <sup>0</sup> MYSGLVGPLVICKKSLART-L---GLKKEVEEFA                                                        |
| Xenopus            | GYHEDVNCV---TWAYYSTVDQV <sup>0</sup> KD <sup>0</sup> TYSGLIGPLVICKRSLPL-D---ML-SKPKRFA                                                                       |

|               |                                                                                        |
|---------------|----------------------------------------------------------------------------------------|
| Homo          | GT-EDSACI---PWAYYSTVDQV <sup>0</sup> KDLYSGLIGPLIVCRPYLK <sup>0</sup> V-F---NP-RRKLEFA |
| Gallus        | ER-GDLHCT---AWAYHSMVDII <sup>0</sup> KDTYSGLIGTLVVCPRHYLPS-S---HT-RKKVRFA              |
| Anolis        | GE-GESTCI---PWAYYSTVDQV <sup>0</sup> KDVYSGLIGTLLVCRRELLPL-H---PP-PKRVEFA              |
| Callorhinchus | SP-GDSCCI---SWAYYSQVDQV <sup>0</sup> KDLFSGLIGTLVICRKSFLGL-S---KT-HDLREFA              |
|               | * : .                                                                                  |

|                    |                                                                                                     |
|--------------------|-----------------------------------------------------------------------------------------------------|
| Chlamydomonas      | LYW <sup>0</sup> QVVDENSSPFLDLNVEAAQLNVTK-FENDAVLSADF <sup>0</sup> DEGNR-----MHSINGYVYC             |
| Acropora           | MLFTVFDENQSWYIDENIKTYLTAKNPGPLDQLKATFDFWESNL-----KGTINGYMFG                                         |
| Strongylocentrotus | LLYTIFDENDDWYLDNIRMFTTQ---PNQVDKESEDFEGSND-----AFT <sup>1</sup> INGYIFG                             |
| Ciona              | -----PP <sup>0</sup> NS <sup>0</sup> ESGRDLFTATR-----DHVL-----S                                     |
| Crassostrea        | FIVGSVDETQSWYLEDNLKTYTGI-----VDTDDPRFAEANAIRGKCSYI <sup>1</sup> TVNGYSDG                            |
| Stegodyphus        | VLYTIFDENESWYLQENIMRYTRQ---PYSVNPNDNRNFRTSNR-----MH <sup>1</sup> AVNGRIYG                           |
| Saccoglossus       | LLFTVTDENESWYFDENIQTYCYT---PGEVDKDDGDFQESNL-----MH <sup>1</sup> GINGKIYG                            |
| Danio              | LLFMVFDENESWYLEENIKTHVKN---PPKALTEDQQFIESNK-----MH <sup>1</sup> GINGLVFG                            |
| Xenopus            | LLFLIFDENESWYLDENIKTYS <sup>1</sup> LH---PDQVQKEDEDFIESNK-----MH <sup>1</sup> GINGKLYA              |
| Homo               | LLFLVFDENESWYLDN <sup>1</sup> IKTYS <sup>1</sup> DH---PEKVNKDDEEFIESNK-----MH <sup>1</sup> AINGRMFG |
| Gallus             | LLFMVFDENESWYLDENIETYSAN---PHLV <sup>1</sup> DKENE <sup>1</sup> EFLESNK-----MH <sup>1</sup> AINGVFG |
| Anolis             | LLFMVFDENESWYLDENIKTYSAN---PEDVNKEDEDFIESNK-----MH <sup>1</sup> VINGRVFG                            |
| Callorhinchus      | LLFMVFDENESWYLDHNIQTYCGQ---PRLV <sup>1</sup> NTEDEDFIESNK-----MH <sup>1</sup> AINKLYG               |
|                    | : . :                                                                                               |

|                    |                                                                                                                            |
|--------------------|----------------------------------------------------------------------------------------------------------------------------|
| Chlamydomonas      | NQPLVTIAKGKKLRWVLVAYGTEGDF <sup>0</sup> HS <sup>0</sup> PQFTGQSL-----EAD <sup>0</sup> --K-----                             |
| Acropora           | NLPMPKMYIGE <sup>0</sup> QVAWYLLQVGGNTEM <sup>0</sup> HTV <sup>0</sup> HF <sup>0</sup> HGQSI-----LYK <sup>0</sup> --S----- |
| Strongylocentrotus | HPLNISMAVGDNIAFYIIIVVGGADV <sup>0</sup> HSIN <sup>0</sup> FG <sup>0</sup> HGTV-----TEK--T-----                             |
| Ciona              | AAPAWSPSRINRRQWIQVDF <sup>1</sup> GR <sup>1</sup> IIKVAGVITQGA <sup>1</sup> KYGRWKLYVKS <sup>1</sup> FALKYSSDGLRWEVY       |
| Crassostrea        | SMKNMTVPMGEDVYFHF <sup>1</sup> FLNVGNDL--LTI <sup>1</sup> EVYGLNTSVSPQTRV--GLS-TVTLYPWDS-                                  |
| Stegodyphus        | TLEGLVARNGTKTAWYLLGMGSEEDL <sup>1</sup> HTAH <sup>1</sup> FH <sup>1</sup> GQTF-----LIR--S-----                             |
| Saccoglossus       | NAYGVEMYSGERIDWYLMAMGSEVDM <sup>1</sup> HTV <sup>1</sup> HF <sup>1</sup> HGTTF-----LHE--H-----                             |
| Danio              | NLKGLNMKVGDKVYWYLMGLGNEVDI <sup>0</sup> HTAH <sup>0</sup> FH <sup>0</sup> GHSF-----DYK <sup>0</sup> --V-----               |
| Xenopus            | NLEGLTMHVGDKIYWH <sup>0</sup> LIGLNEADI <sup>0</sup> HTM <sup>0</sup> HF <sup>0</sup> HAHSF-----TYV <sup>0</sup> --R-----  |
| Homo               | NLQGLTMHVGDEVNWYLMGMGNEIDL <sup>0</sup> HTV <sup>0</sup> HF <sup>0</sup> HGHSF-----QYK <sup>0</sup> --H-----               |
| Gallus             | NLHGLTMHVGDEVSWYLMAMGNEIDI <sup>0</sup> HTAH <sup>0</sup> FH <sup>0</sup> GHSF-----DYK <sup>0</sup> --Q-----               |
| Anolis             | NLHGLTMHVGDNVWYLLGMGNEVDM <sup>0</sup> HTAH <sup>0</sup> FH <sup>0</sup> GHSF-----DYK <sup>0</sup> --Q-----                |
| Callorhinchus      | NLHGLEMVEGERAFWYLMGMGNEVDI <sup>0</sup> HTAH <sup>0</sup> FH <sup>0</sup> GHSF-----DYK <sup>0</sup> --M-----               |
|                    | : : * . .                                                                                                                  |

|                    |                                                           |
|--------------------|-----------------------------------------------------------|
| Chlamydomonas      | --SG--YSTLAS--LMPSIARVA-----DMTAADVGTWL-----              |
| Acropora           | --MSYHRGDVYD--LLPESFATV-----LMVPEAVGTWL-----              |
| Strongylocentrotus | --RVIKNLDSVA--TLP <sup>0</sup> GFITV-----ELEATSRGTWL----- |

|                    |                                                              |
|--------------------|--------------------------------------------------------------|
|                    | 1                                                            |
| Ciona              | EDPVSHRAKVFRGNIDPDTAVTQTLQNPVQTRFIRIQPRSWRS----RIALRFEI----- |
| Crassostrea        | -----KSIVTRFDTPGSFYQEEL-----TSQGHTTQ-----                    |
| Stegodyphus        | --DTVRRGDVVD--LFPGYFETV-----EMINDNPGTWI-----                 |
| Saccoglossus       | -GIGVHRADVVD--LFPGKFDTSD-----SKMHR SINMMMFVKISLITLVVVIAGVSA  |
| Danio              | --SGTHRADVVD--LVPGTFTQTV-----TMRPLYSGTWL-----                |
| Xenopus            | --GATFHSDVVD--LFPGTFTQTV-----EMIARVPGTWL-----                |
| Homo               | --RGVYSSDVVD--IFPGTYQTL-----EMFPRTPGIWL-----                 |
| Gallus             | --TEVYRADVVD--LFPGTFTQTV-----EMIPQNPGVWL-----                |
| Anolis             | --TRNFRNDVVD--LFPGTFTQTV-----EMVPNNPGTWL-----                |
| Callorhinchus      | --KETYRGDVVD--LFPGTFTQTL-----EMYPQHIGTWL-----                |
|                    | *                                                            |
| Chlamydomonas      | -----LYCDVHDHYMAGMMSQFAVTAA-----                             |
| Acropora           | -----LHCHVNVHINKGCMNGLFTVMDPTGTPPTNQA---PPKGE-----SGP        |
| Strongylocentrotus | -----LHDAVNKHSDLGAETRYTVYAKKSDIPCPECATAAR--PPTGA-----AG      |
| Ciona              | -----LGCQSVDAESRDLLDATP-----F-----PDNCR-----                 |
| Crassostrea        | -----IHGFYTVLSNTKTTE-----                                    |
| Stegodyphus        | -----LHCHVDDHMYRYCMVAVTFTVTA-----                            |
| Saccoglossus       | KEPLSFEFPLNCRVLIHR-----YRLQPRGGGGGGGGGASSGLKCPDPNGKLKRSQS    |
| Danio              | -----LHCHVTDHIQACMETTYTVLEKDGKRGFLGIFGSG-----                |
| Xenopus            | -----LHCHVTDHIHACMEVLYTVLETMQEKKIR-----                      |
| Homo               | -----LHCHVTDHIHACMETTYTVLQNEDTKS-----                        |
| Gallus             | -----LHCHVTDHIHACMEATYTVLPKEDKWSLFPRLFSQSKCK-----GTP         |
| Anolis             | -----LHCHVTDHIHACMETTYTVLPKANADIERGQVKM-SIGE-----QSQ         |
| Callorhinchus      | -----LHCHVVDHIEGGMETTYTVHKKPEKKGIFHGIKTLFGSS-----G--         |
| Chlamydomonas      | -----                                                        |
| Acropora           | YLRASFIPILFFLF-VAVMI-ANHNSDQ-----                            |
| Strongylocentrotus | RLESTICSLVAALIIGALMSFE-----                                  |
| Ciona              | -----NSE-----                                                |
| Crassostrea        | -----                                                        |
| Stegodyphus        | -----                                                        |
| Saccoglossus       | YTDDDGCTVCDCDGTGVFDCDNSPCDTSGFSGSSNSGCLYNTDSYKVGAVWEIDQCGNS  |
| Danio              | -----                                                        |
| Xenopus            | -----                                                        |
| Homo               | -----                                                        |
| Gallus             | NMRD-----                                                    |
| Anolis             | N-----                                                       |
| Callorhinchus      | -----                                                        |
| Chlamydomonas      | -                                                            |
| Acropora           | -                                                            |
| Strongylocentrotus | -                                                            |
| Ciona              | -                                                            |
| Crassostrea        | -                                                            |
| Stegodyphus        | -                                                            |
| Saccoglossus       | W                                                            |

|               |   |
|---------------|---|
| Danio         | - |
| Xenopus       | - |
| Homo          | - |
| Gallus        | - |
| Anolis        | - |
| Callorhinchus | - |

**Figura 3.** Alineamiento de aminoácidos de 13 eucariotas de diferentes grupos (Tabla 1 y Figura 2)

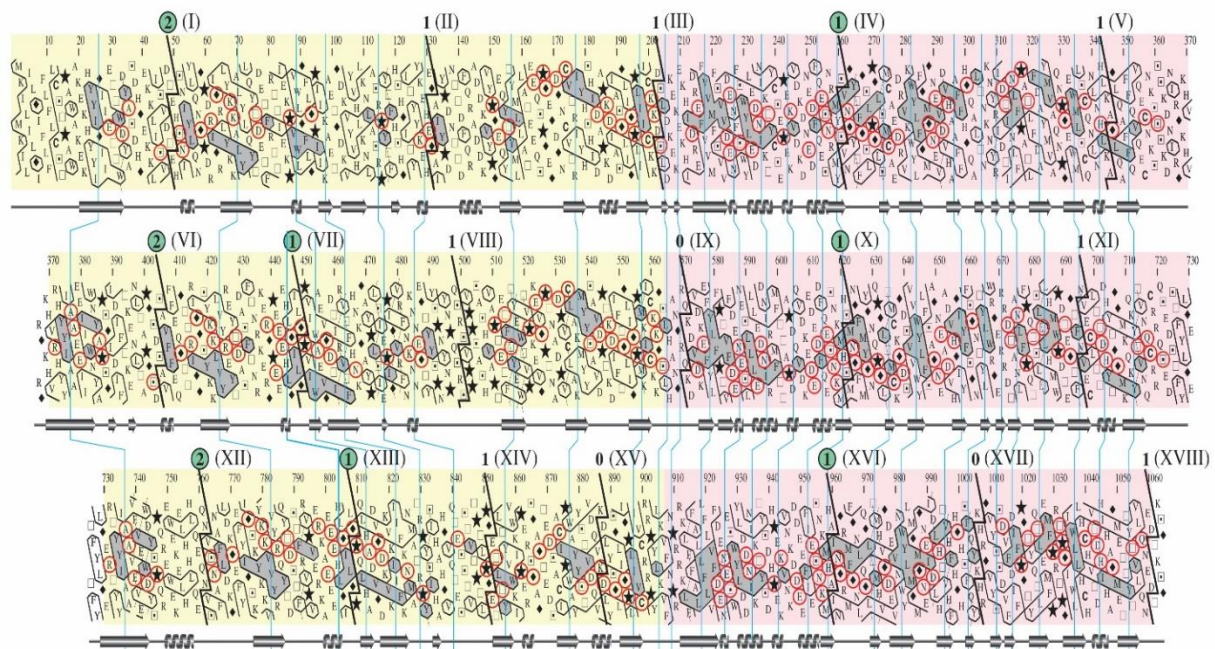
**Nota:** En orden descendente, los genes de cada individuo son: *Chlamydomonas*, FOX1 (Ferroxidasa 1); *Acropora*, HEPH (Hefastina); *Strongylocentrotus*, FOX (Ferroxidasa); *Ciona*, FVIII (Factor VIII); *Crassostrea*, CP (Ceruloplasmina); *Stegodhypyus* y *Saccoglossus*, HEPH1 (Hefastina 1); *Danio*, *Xenopus*, *Homo*, *Gallus*, *Anolis* y *Callorhinchus*, CP. En color azul se resaltan los centros de unión al cobre. Los números representan la posición y la fase de los intrones.

Con el surgimiento de los moluscos y artrópodos, alrededor de 678-916 M.A. (Cartwright *et al.*, 2007; Nei *et al.*, 2001; Peterson *et al.*, 2008; Hausdorf *et al.*, 2000), el gen es invadido por nuevos intrones, de los cuales 5 permanecen en los dominios 1(II), 2 (V), 4 (IX, X) y 6 (XVI). El intrón VIII de osteíctios y condriictios, que está presente en el dominio 3, se conserva en el ancestro de molusco y artrópodos, lo que sugiere que la pérdida del intrón en los demás vertebrados sucedió durante esta divergencia. En la separación de los hemicordados a cordados, se presentó una ganancia en el dominio 2 (IV), hace aproximadamente 601-860 M.A. (Blair *et al.*, 2005; Peterson *et al.*, 2008).

La aparición de los vertebrados permite la inserción de 3 nuevos intrones en los dominios 3(VII) y 5 (XIII, XV). Los intrones del gen CP se encuentran altamente conservados en todos los vertebrados, a pesar de la pérdida del intrón VIII que permanece en osteíctios y condriictios (**Figura 3**). Finalmente, el intrón VIII que se ha perdido en el grupo de los condriictios, sugiere que este evento de delección se haya dado hace 450-497 M.A. (Betancur-R *et al.*, 2013) en la divergencia del ancestro común con osteíctios.

**Tabla 3.**  
Valores de RMSD entre los 6 dominios cupredoxina del gen CP.

|   | 1    | 2    | 3    | 4    | 5    | 6    |
|---|------|------|------|------|------|------|
| 1 | 0.00 | 3.55 | 1.69 | 3.55 | 1.46 | 3.51 |
| 2 | 3.55 | 0.00 | 3.49 | 0.01 | 3.68 | 0.63 |
| 3 | 1.69 | 3.49 | 0.00 | 3.49 | 1.63 | 3.47 |
| 4 | 3.55 | 0.01 | 3.49 | 0.00 | 3.68 | 0.63 |
| 5 | 1.46 | 3.68 | 1.63 | 3.68 | 0.00 | 3.67 |
| 6 | 3.51 | 0.63 | 3.47 | 0.63 | 3.67 | 0.00 |



**Figura 4.** Análisis de agrupamiento de aminoácidos hidrofóbicos entre los dominios de la ceruloplasmina en *Homo sapiens*.

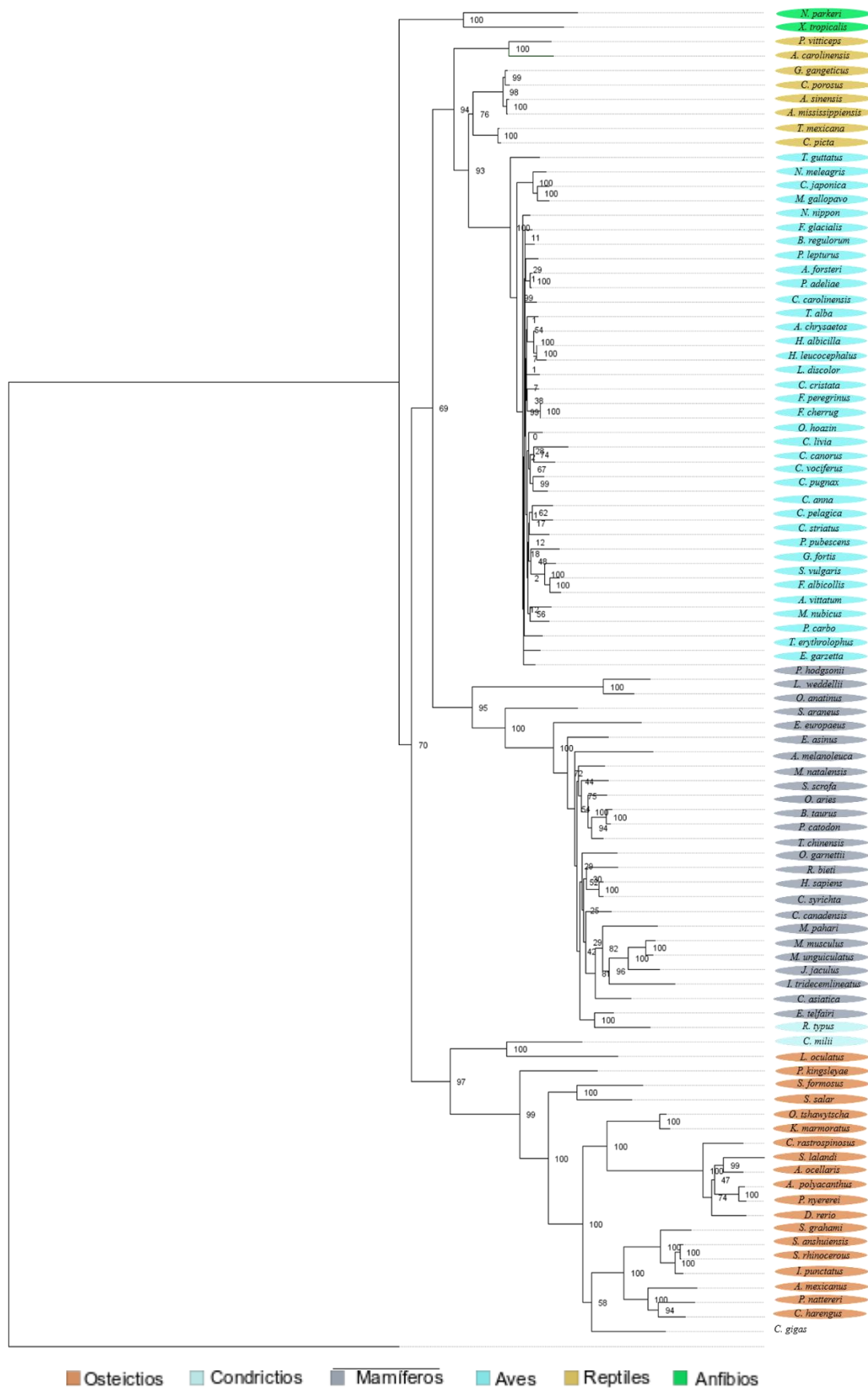
**Nota:** Los colores amarillo y rosa delimitan los 6 dominios cupredoxina. En líneas en negro tipo 'zig-zag' se indican los intrones y encima de éstas se encuentran en números arábigos (0, 1 y 2) las fases respectivas. El orden de las fases se denota en números romanos (I, I + n). Los números dentro de los círculos representan los intrones conservados entre los dominios. Las líneas azules corresponden a los residuos conservados entre los dominios. Las estructuras secundarias en forma de hojas beta y alfa hélices, se indican debajo de los dominios.

### 4.3 Reconstrucción de Relaciones Ancestrales.

Las reconstrucciones filogenéticas de los genes MCOs en los diferentes clados principales de la historia evolutiva de los seres vivos, se realizaron con el método de Máxima Verosimilitud (**Figura 5**, **Figura 6** y **Figura 7**). En el árbol filogenético, construido con 93 especies de vertebrados (**Figura 5**) se observan 6 agrupamientos que corresponden a cada grupo del clado: anfibios, reptiles, aves, mamíferos, peces óseos y cartilaginosos. Se usaron alineamientos sin la presencia de intrones en las secuencias (mRNA). Los nodos se encuentran bien soportados, con valores mayores al 50%. Anfibios presenta un soporte del 100%, nodo que además agrupa los peces y los separa de mamíferos y saurópsidos (aves y reptiles). Por otra parte, el nodo de los saurópsidos presenta un soporte del 94% y 95% para mamíferos. Peces óseos se encuentra

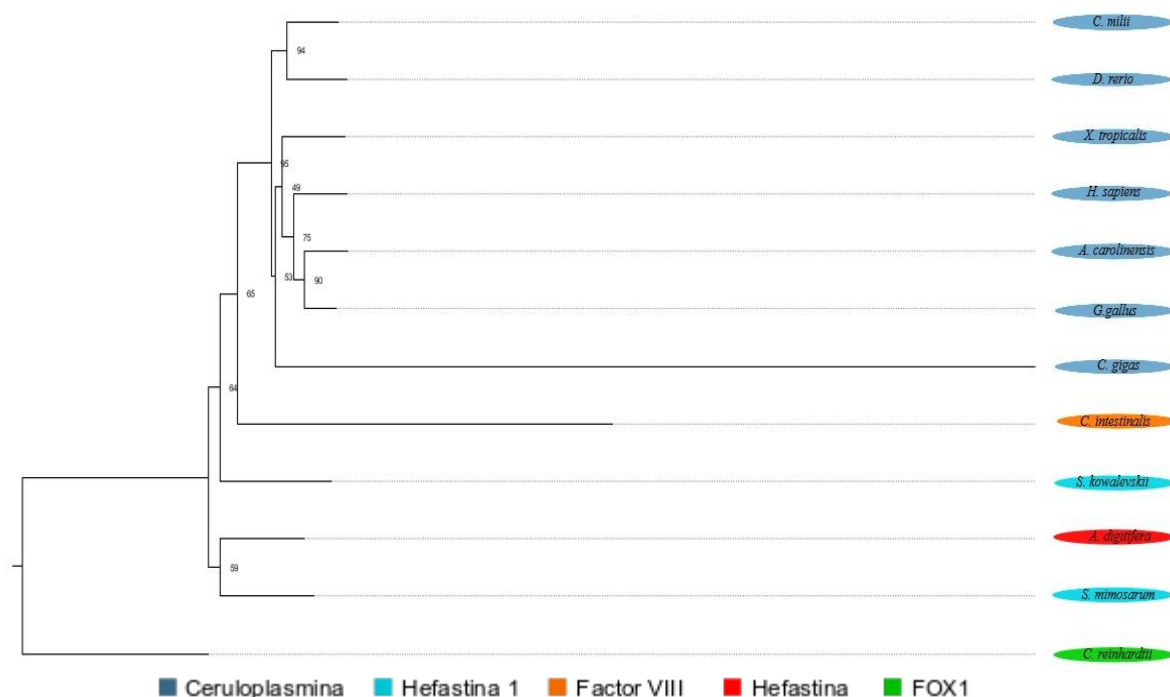
separado de peces cartilaginosos con un soporte del 97%. Los patrones de las ramas demuestran que todas las terminales proceden de un ancestro común y los agrupamientos concuerdan con la taxonomía de vertebrados. Para enraizar los vertebrados se usó el molusco *C. gigas* como el ancestro del gen más cercano.

La **Figura 6** representa las relaciones ancestrales de 5 genes ortólogos en especies de cordados, hemicordados, moluscos, artrópodos y cnidarios. El gen *CP* se encuentra agrupado con F8 con un soporte *bootstrap* de 65%. *S. kowalevskii* con HEHP1 se relaciona con los genes CP y F8 con un soporte de 64%. Finalmente, *S. mimosarum* y *A. digitifera* que corresponden a los genes HEPH1 y HEPH respectivamente, se encuentran agrupado en el mismo nodo con un soporte del 59%. Aunque las dos especies del gen HEPH1 no se agrupan dentro del mismo nodo, se mantiene la agrupación taxonómica. El árbol se enraizó con *C. reinhardtii*.



**Figura 5.** Reconstrucción de máxima verosimilitud de la ceruloplasmina en 93 especies de vertebrados

**Nota:** Los valores de Bootstrap se muestran entre los nodos. Escala de distancia: 0.3.



**Figura 6.** Reconstrucción de CP, HEPH, HEPH1 y F8 en 11 especies de diferentes linajes eucariotas.  
**Nota:** *C. reinhardtii* con el gen FOX1 como raíz del árbol filogenético. Escala de distancia: 0.8.

## 5. Discusión

Los eucariotas evaluados en nuestros resultados revelan una ancestría en la disposición de sus intrones. De acuerdo con Fedorov *et al.*, (2002) y Rogozin *et al.*, (2003) en los análisis estructurales de los genes en familias de homología ortóloga, del 10% al 25 % de los intrones se comparten entre animales y plantas, indicando que los intrones presentes en la misma posición son intrones conservados desde el ancestro del gen. Por otra parte, Roy *et al.*, (2003) describe una adquisición de intrones en genes ortólogos de mamíferos, aunque sin eventos de pérdidas. Sin embargo, los últimos sucesos de inserción-delección en los genes que evaluamos, se demuestran ambos fenómenos en la especiación del ancestro vertebrado. La última pérdida de intrones que se observó en vertebrados ocurre en la especiación de peces.

Rzhetsky *et al.*, (1997) reporta casos de inserciones y deleciones, las cuales han existido durante la evolución de Eucaria; mientras que Rogozin *et al.*, (2003) en análisis de genes

ortólogos de vertebrados, demuestra pérdidas pero ninguna ganancia. Sin embargo, con una cantidad menor de genes, nuestros resultados evidencian ambos eventos. No obstante, Rogozin *et al.*, (2003) sugiere que las tasas de pérdidas de intrones varían entre los diferentes linajes.

La relación de la distribución de las fases en los metazoos principales del estudio parece obedecer a una mayor proporción de presencia de intrones en fase 1, contrario a Csuros *et al.*, (2011), que revela una correlación de las fases en proporción 2:1:1, siendo la fase 0 la más predominante. Aunque es posible que las proporciones no tengan ninguna dependencia con la misma cantidad de intrones en el ancestro más reciente, en el gen FOX1 de *Chlamydomonas reinhardtii* se observa una alta población de intrones en fase 0, rasgo que según Csuros *et al.*, (2011) puede variar entre los diferentes linajes y además puede indicar una característica compartida del ancestro común de Eucaria.

Según Roy *et al.*, (2005), el 37% de los intrones en humanos y plantas o grupos no pertenecientes al clado de los metazoos, se debieron adquirir en grandes cantidades temprano en la evolución, heredados a través de los ancestros comunes en cada grupo. En un porcentaje mayor, el 80% de los intrones del gen CP en vertebrados se han heredado a partir de un ancestro común desde los Eucaria unicelulares, contra un 66% conservados desde el ancestro metazoa. En ambos casos, los sucesos de inserciones y deleciones han favorecido la prevalencia de algunos de los intrones dispuestos en posiciones únicas de las secuencias génicas.

La reconstrucción de nuestro árbol filogenético bien soportado del gen CP, es consistente con la filogenias de Evan *et al.*, (2009) y Yamamoto *et al.*, (2017), manteniéndose las relaciones de las ramas. El registro fósil del ancestro cordado de vertebrados, surgió alrededor de 570 M.A. Hace 440 M.A diverge el ancestro de peces cartilaginosos, continua a la especiación de los tetrápodos hace 420-400 MA. Los amniotas comparten ancestro con los anfibios hace aproximadamente 325 MA. De esta manera, el árbol filogenético revela una coherencia con las

apariciones y pérdidas de intrones en los grupos eucariotas, lo que permite contrastar la cronología de los intrones.

La topología de nuestro árbol filogenético hecho a partir de 5 genes ortólogos, concuerda con la filogenia metazoa presentada por Dunn *et al.*, (2008). Cnidarios soportado por valores de *bootstrap* superiores al 80%, se separa de artrópodos. Con relaciones más cercanas, hemicordados y cordados se separan con valores de *bootstrap* mayores al 98%. Aunque con valores más bajos, los patrones de nuestras ramas se soportan en valores de *bootstrap* del 59% entre cnidarios y artrópodos y 64% entre hemicordados y cordados.

El gen CP agrupa todos los miembros que lo poseen, entre ellos, el molusco *C. gigas* que aunque según Dunn *et al.*, (2008) se encuentra alejado de cordados. Sin embargo, el gen HEPH1 no se agrupa en un mismo clado sino que diverge abedeciendo las distancias evolutivas entre ellos. No obstante, se concuerda con la presencia de ramas largas terminales y cortas con Dunn *et al.*, (2008), Nosenko *et al.*, (2013) y Erwin *et al.*, (2011) que sugieren señales de radiaciones rápidas en los ancestros, consistentes con el registro fósil, hace aproximadamente 700 M.A. Rokas *et al.*, (2003) sugiere que las reconstrucciones filogenéticas con pocos genes no son adecuadas para resolver relaciones entre los nodos más alejados de metazoa, lo que sin duda puede afectar la resolución de la topología de nuestro árbol. Además, es posible que la limitación de los modelos evolutivos de sustitución de MV y el tamaño del muestreo de los taxones jueguen un papel importante en la estimación de los soportes entre los nodos.

## 5.1 Modelo Evolutivo de la Ceruloplasmina.

Observamos en nuestros resultados patrones en la disposición de los intrones de la ceruloplasmina en *Homo sapiens*, tanto en el gen CP de moluscos como en los demás genes ortólogos (**Tabla 3**). El intrón I, VI, y XII que se encuentran en fase 2 en los dominios 1, 3 y

5, mantiene la misma relación de disposición en los genes HEPH, HEPH1 y CP de no vertebrados. Mientras tanto, el intrón V, XI y XVIII que se encuentra en fase 1 en los dominios 2, 4 y 6 se presentan en los mismos genes, algunos conservados con FOX1.

Durante los eventos de especiación de los eucariotas, sucesos de masivas invasiones de intrones permitieron colonizar el gen nuclear. La hipótesis de los intrones tardíos concuerda con la alta disposición de intrones. Además, la estructura de los genes presenta inserciones a lo largo de su historia evolutiva. Algunos autores como Logsdon *et al.*, (1998), Stoltzfus *et al.*, (1994) y (1997), presentan evidencias de intrones tardíos como una nueva adquisición en Eucaria. Por lo tanto, al no encontrarse intrones en arqueas o bacterias, estos nunca se han podido heredar del ancestro universal.

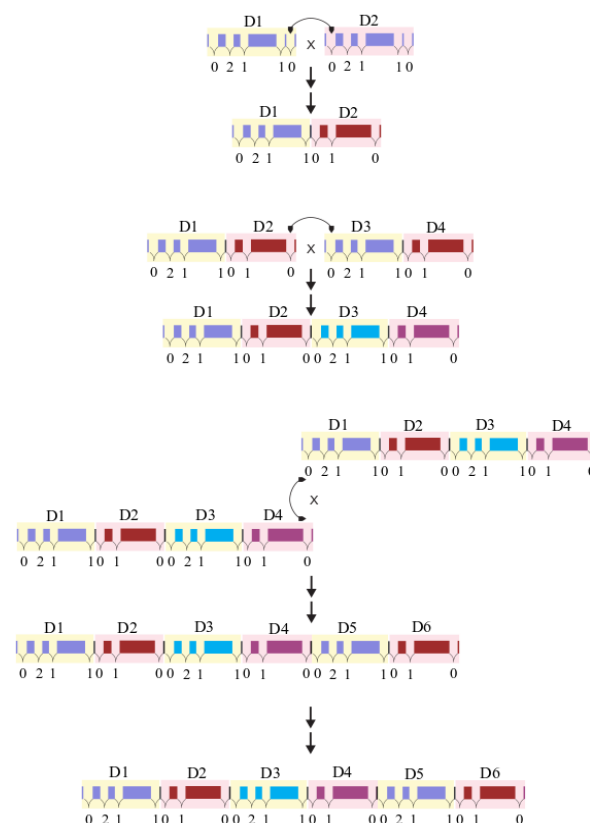
Por otra parte, los porcentajes de identidad o los valores de RMSD que concuerdan con el autor Murphy *et al.*, (1997), sugieren que el dominio 1 es el dominio primigenio precursor, posteriormente este se duplica, formándose el dominio 3. Finalmente, una nueva duplicación del dominio 1 y 2 produce el dominio 5 y 6, dando origen a una proteína de 6 dominios. Los dominios cupredoxina de la ceruloplasmina presentan una homología con los dominios de la ferroxidasa de algas Terzulli *et al.*, (2009), por lo que, en esencia, la identidad de la proteína con genes ortólogos corrobora su origen monofilético.

Así, proponemos un modelo que sugiere una ruta alternativa obedeciendo a una recombinación mediada por intrones del gen CP, en el que el ancestro de Eucaria a partir de un ancestro cupredoxina se duplica y genera eventualmente proteínas de 2, 3 y 6 dominios. Durante la invasión de los intrones en los eucariotas se pudieron conservar algunos intrones, manteniendo la función de la proteína, pero además, la presencia de intrones en fase 0 debieron ser necesarios al inicio de la secuencia del dominio 1 y final del dominio 2.

Sin embargo, es posible que los intrones en fase 0 necesarios para no truncar la proteína ancestral y que esta especiara en los eucariotas, se han perdido por los procesos de

diversificación, mientras las mutaciones que producen *indels*, desplazan nucleótidos que paulatinamente cambian el marco de lectura abierto (Li, 1997). Entonces, las fases 1 y 2 (**Figura 4**) delimitadas por una fase 0 permiten discriminar unidades de duplicación, en el que la unidad 1 representaría los dominios 1 y 2, la unidad 2 los dominios 3 y 4 y la unidad 3 los dominios 5 y 6, por lo que se concordaría con Ortel *et al.*, (1984).

Las recombinaciones homólogas despareadas entre cromosomas durante la división celular, son eventos mediados por el grado de homología de las secuencias (Tayebi *et al.*, 2003), como se observa en la homología intra-dominio de CP. Por lo tanto, al favorecer las deleciones y duplicaciones, los dominios 5 y 6 se pudieron adquirir en la proteína de 4 dominios ancestral. Por otro lado, la copia extra pudo perder la funcionalidad o pudo continuar un linaje de proteínas de 2 dominios.



**Figura 7.** Modelo evolutivo de la ceruloplasmina.

**Nota:** La estructura del gen pertenece a *Homo sapiens*. Los 6 dominios estructurales están denotados como: D1, dominio 1; D2, dominio 2; D. 3, dominio 3; D4, dominio 4; D5, dominio 5; D6: dominio 6. Las fases de los intrones se encuentran modificadas del gen original y están denotadas en números arábigos (0, 1 y 2). La 'x' representa la recombinación despareada del cromosoma. Los exones del gen se proporcionados a su secuencia. Las flechas indican el orden de aparición de los dominios.

## **6. Conclusiones**

La disponibilidad de las secuencias en las bases de datos, permitieron reconstruir la historia de la estructura del gen de la ceruloplasmina en vertebrados. Los intrones del gen se han conservado casi en su totalidad en todo el grupo.

El gen de la ceruloplasmina data desde el último ancestro común eucariota. Las adquisiciones de los intrones surgen en el último ancestro común de Eucaria, evidencia que aporta en el entendimiento de la hipótesis de los intrones tardíos.

Durante la radiación eucariota, la ceruloplasmina sufrió continuos eventos de pérdida e inserción de intrones, que dejaron rastro en cada linaje. La ceruloplasmina presenta en las reconstrucciones filogenéticas un origen monofilético, el cual se comprueba en la presencia de los intrones conservados.

## **7. Recomendaciones**

Dado que la ceruloplasmina presenta un ancestro muy antiguo, se sugiere realizar un trabajo exclusivo para caracterizar el gen en otros linajes y mejorar el modelo de sucesión de dominios.

Si se requiere hacer un análisis más profundo de las relaciones monofiléticas, se recomienda usar otro método de reconstrucción como el de MrBayes, dado que las opciones para elegir el modelo que mejor se ajuste a los datos es mucho más amplia.

### Referencias Bibliográficas

- Abascal, F., Irisarri, I., & Zardoya, R. (2014). Filogenia y evolución molecular. *Bioinformática con Ñ*, 1, 231-257.
- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., ... & Howe, K. (2016). The Ensembl gene annotation system. *Database*, 2016.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.
- Babenko, V. N., Rogozin, I. B., Mekhedov, S. L., & Koonin, E. V. (2004). Prevalence of intron gain over intron loss in the evolution of paralogous gene families. *Nucleic Acids Research*, 32(12), 3724-3733.
- Baldrian, P. (2006). Fungal laccases—occurrence and properties. *FEMS microbiology reviews*, 30(2), 215-242.
- Bazinet, A. L., Zwickl, D. J., & Cummings, M. P. (2014). A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Systematic biology*, 63(5), 812-818.
- Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, 74(8), 3171-3175.
- Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., ... & Zhang, F. (2013). The tree of life and a new classification of bony fishes. *PLoS currents*, 5.
- Blair, J. E., & Hedges, S. B. (2005). Molecular phylogeny and divergence times of deuterostome animals. *Molecular biology and evolution*, 22(11), 2275-2284.
- Brack and Tonegawa,: Brack, C., & Tonegawa, S. (1977). Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. *Proceedings of the National Academy of Sciences*, 74(12), 5652-5656.
- Breathnach et al. : Breathnach, R., Mandel, J. L., & Chambon, P. (1977). Ovalbumin gene is split in chicken DNA. *Nature*, 270(5635), 314.
- Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat., & Mornon, J. P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and Molecular Life Sciences*, 53(8), 621-645.

- Cartwright, P., & Collins, A. (2007). Fossils and phylogenies: integrating multiple lines of evidence to investigate the origin of early major metazoan lineages. *Integrative and Comparative Biology*, 47(5), 744-751.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research*, 31(13), 3497-3500.
- Chow et al., 1977: Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1), 1-8.
- Csuros, M., Rogozin, I. B., & Koonin, E. V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS computational biology*, 7(9), e1002150.
- Daimon, M., Yamatani, K., Igarashi, M., Fukase, N., Kawanami, T., Kato, Tominaga, H., & Sasaki, H. (1995). Fine structure of the human ceruloplasmin gene. *Biochemical and biophysical research communications*, 208(3), 1028-1035.
- De Souza, S. J., Long, M., Klein, R. J., Roy, S., Lin, S., & Gilbert, W. (1998). Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proceedings of the National Academy of Sciences*, 95(9), 5094-5099.
- Doolittle, R. F. (2015). Bioinformatic characterization of genes and proteins involved in blood clotting in lampreys. *Journal of molecular evolution*, 81(3-4), 121-130.
- Dufayard, J. F., Duret, L., Penel, S., Gouy, M., Rechenmann, F., & Perrière, G. (2005). Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics*, 21(11), 2596-2603.
- Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., ... & Sørensen, M. V. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452(7188), 745.
- Efron, B., Halloran, E., & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences*, 93(23), 13429-13429.
- Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., & Peterson, K. J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *science*, 334(6059), 1091-1097.
- Evans, S. E. (2009). Evolution and Phylogeny of Vertebrates. In *Encyclopedia of Neuroscience* (pp. 1194-1197). Springer, Berlin, Heidelberg.

- Fedorov, A., Merican, A. F., & Gilbert, W. (2002). Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proceedings of the National Academy of Sciences*, 99(25), 16128-16133.
- Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology*, 22(3), 240-249.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6), 368-376.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
- Hausdorf, B. (2000). Early evolution of the bilateria. *Systematic biology*, 49(1), 130-142.
- Hedges, S. B., Blair, J. E., Venturi, M. L., & Shoe, J. L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC evolutionary biology*, 4(1), 2.
- Hellman, N. E., & Gitlin, J. D. (2002). Ceruloplasmin metabolism and function. Annual review of nutrition, 22(1), 439-458.
- Hickey, D. A. (1992). Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. *Genetica*, 86(1-3), 269-274.
- Holmberg, C. G., & Laurell, C. B. (1948). Investigations in serum copper. *Acta chem scand*, 2, 550-556.
- Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., ... & Durbin, R. (2002). The Ensembl genome database project. *Nucleic acids research*, 30(1), 38-41.
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., ... & Suzuki, Y. (2016). A new view of the tree of life. *Nature microbiology*, 1(5), 16048.
- Jeffares, D. C., Mourier, T., & Penny, D. (2006). The biology of intron gain and loss. *TRENDS in Genetics*, 22(1), 16-22.
- Jeffares, D. C., Poole, A. M., & Penny, D. (1998). Relics from the RNA world. *Journal of Molecular Evolution*, 46(1), 18-36.
- Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. *Mammalian protein metabolism*, 3(21), 132.

- Koschinsky, M. L., Funk, W. D., Van Oost, B. A., & MacGillivray, R. T. (1986). Complete cDNA sequence of human preceruloplasmin. *Proceedings of the National Academy of Sciences*, 83(14), 5086-5090.
- Kufareva, I., & Abagyan, R. (2011). Methods of protein structure comparison. In *Homology Modeling* (pp. 231-257). Humana Press.
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), 772-773.
- Li, W. (1997). *Molecular evolution*. Sinauer associates incorporated.
- Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Mi Park, Y., Buso, N., & Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic acids research*, 43(W1), W580-W584.
- Logsdon Jr, J. M., Stoltzfus, A., & Doolittle, W. F. (1998). Molecular evolution: recent cases of spliceosomal intron gain?. *Current Biology*, 8(16), R560-R563.
- Lynch, M. (1999). The age and relationships of the major animal phyla. *Evolution*, 53(2), 319-325.
- Lynch, M., & Richardson, A. O. (2002). The evolution of spliceosomal introns. *Current opinion in genetics & development*, 12(6), 701-710.
- Margush, T., & McMorris, F. R. (1981). Consensus n-trees. *Bulletin of Mathematical Biology*, 43(2), 239-244.
- Messerschmidt, A., & Huber, R. (1990). The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin Modelling and structural relationships. *European journal of Biochemistry*, 187(2), 341-352.
- Murphy, M. E., Lindley, P. F., & Adman, E. T. (1997). Structural comparison of cupredoxin domains: domain recycling to construct proteins with novel functions. *Protein Science*, 6(4), 761-770.
- Nakamura, K., & Go, N. (2005). Function and molecular evolution of multicopper blue proteins. *Cellular and Molecular Life Sciences*, 62(18), 2050-2066.
- Nei, M., Xu, P., & Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proceedings of the National Academy of Sciences*, 98(5), 2497-2502.

- Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., ... & Vacelet, J. (2013). Deep metazoan phylogeny: when different genes tell different stories. *Molecular phylogenetics and evolution*, 67(1), 223-233.
- Ortel, T. L., Takahashi, N., & Putnam, F. W. (1984). Structural model of human ceruloplasmin based on internal triplication, hydrophilic/hydrophobic character, and secondary structure of domains. *Proceedings of the National Academy of Sciences*, 81(15), 4761-4765.
- Parfrey, L. W., Lahr, D. J., Knoll, A. H., & Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, 108(33), 13624-13629.
- Pathy, L. (1987). Intron-dependent evolution: preferred types of exons and introns. *FEBS letters*, 214(1), 1-7.
- Peterson, K. J., Cotton, J. A., Gehling, J. G., & Pisani, D. (2008). The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1496), 1435-1443.
- Peterson, K. J., Lyons, J. B., Nowak, K. S., Takacs, C. M., Wargo, M. J., & McPeck, M. A. (2004). Estimating metazoan divergence times with a molecular clock. *Proceedings of the National Academy of Sciences*, 101(17), 6536-6541.
- Poole, A. M., Jeffares, D. C., & Penny, D. (1998). The path from the RNA world. *Journal of Molecular Evolution*, 46(1), 1-17.
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular biology and evolution*, 25(7), 1253-1256.
- Rogozin, I. B., Carmel, L., Csuros, M., & Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biology direct*, 7(1), 11.
- Rogozin, I. B., Sverdlov, A. V., Babenko, V. N., & Koonin, E. V. (2005). Analysis of evolution of exon-intron structure of eukaryotic genes. *Briefings in bioinformatics*, 6(2), 118-134.
- Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. G., & Koonin, E. V. (2003). Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Current Biology*, 13(17), 1512-1517.
- Rokas, A., King, N., Finnerty, J., & Carroll, S. B. (2003). Conflicting phylogenetic signals at the base of the metazoan tree. *Evolution & development*, 5(4), 346-359.

- Roy, S. W., Fedorov, A., & Gilbert, W. (2003). Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proceedings of the National Academy of Sciences*, 100(12), 7158-7162.
- Rydén, L. G., & Hunt, L. T. (1993). Evolution of protein complexity: the blue copper-containing oxidases and related proteins. *Journal of molecular evolution*, 36(1), 41-66.
- Rzhetsky, A., Ayala, F. J., Hsu, L. C., Chang, C., & Yoshida, A. (1997). Exon/intron structure of aldehyde dehydrogenase genes supports the “introns-late” theory. *Proceedings of the National Academy of Sciences*, 94(13), 6820-6825.
- Sakurai, T., & Kataoka, K. (2007). Basic and applied features of multicopper oxidases, CueO, bilirubin oxidase, and laccase. *The Chemical Record*, 7(4), 220-229.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
- Stamatakis, A. (2016). The RAxML v8. 2. X Manual. *Heidelberg Institute for Theoretical Studies*. Available from: <http://sco. h-its. org/exelixis/web/software/raxml/#documentation> (accessed 20 July 2016).
- Stern, D (2013). The genetic causes of convergent evolution. *Nature reviews. Genetics* 14(11), 751-764.
- Stoltzfus, A., Spencer, D. F., Zuker, M., Logsdon, J. M., & Doolittle, W. F. (1994). Testing the exon theory of genes: the evidence from protein structure. *Science*, 265(5169), 202-207.
- Strimmer, K. S. (1997). *Maximum likelihood methods in molecular phylogenetics*. Herbert Utz Verlag.
- Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*, 17(2), 57-86.
- Tayebi, N., Stubblefield, B. K., Park, J. K., Orvisky, E., Walker, J. M., LaMarca, M. E., & Sidransky, E. (2003). Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease. *The American Journal of Human Genetics*, 72(3), 519-534.
- Terzulli, A. J., & Kosman, D. J. (2009). The Fox1 ferroxidase of *Chlamydomonas reinhardtii*: a new multicopper oxidase structural paradigm. *JBIC Journal of Biological Inorganic Chemistry*, 14(2), 315-325.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

- position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680.
- Vasin, A., Klotchenko, S., & Puchkova, L. (2013). Phylogenetic analysis of sixdomain multi-copper blue proteins. *PLOS Currents Tree of Life*.
- Vibranovski, M. D., & de Souza, S. J. (2003). Insights on the Origin and Evolution of Introns. In *Proc. of the Third Brazilian Symp. on Mathematical and Computational Biology-v1* (Vol. 1, p. 254). Editora E-papers.
- Virit, O., Altindag, A., Selek, S., Yumru, M., Bulut, M., Erel, O., Asuman, H., & Herken, H. (2008). Increased plasma ceruloplasmin levels in schizophrenia. *Bulletin of Clinical Psychopharmacology*, 18, 282-287.
- Woodcock, S., Mornon, J. P., & Henrissat, B. (1992). Detection of secondary structure elements in proteins by hydrophobic cluster analysis. *Protein engineering*, 5(7), 629-635.
- Wu, J. (2008). Improving the specificity of exon prediction using comparative genomics. *BMC genomics*, 9(2), S13.
- Yamamoto, K., Bloch, S., & Vernier, P. (2017). New perspective on the regionalization of the anterior forebrain in Osteichthyes. *Development, growth & differentiation*, 59(4), 175-187.
- Yang, Z. (2006). *Computational molecular evolution*. Oxford University Press.