

**Accepted Manuscript / Manuscrito aceptado**

Análisis de redes complejas para genes ortólogos en vertebrados: un estudio a nivel genético de la evolución

Doi: <https://doi.org/10.15446/abc.v30n1.111588>

To appear in / Para aparecer en: *Acta Biológica Colombiana*

Received Date / Fecha de recibido: 22<sup>nd</sup> January 2024 / 22 de enero de 2024

Revised Date / Fecha de revisado: 07<sup>th</sup> May 2024 / 07 de mayo de 2024

Accepted Date / Fecha de aceptado: 29<sup>th</sup> August 2024 / 29 de agosto de 2024

**Please cite this article as / Cite así:** Análisis de redes complejas para genes ortólogos. (2025).

Análisis de redes complejas para genes ortólogos en vertebrados: un estudio a nivel genético de la evolución. *Acta Biol Colomb.*, 30(1), XX-XX. <https://doi.org/10.15446/abc.v30n1.111588>

This is a PDF file of an unedited manuscript accepted for publication. As a service to our readers, we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Este es un archivo PDF de un manuscrito sin editar que ha sido aceptado para su publicación. Como servicio a nuestros lectores, ponemos a su disposición esta versión preliminar del manuscrito. Este manuscrito se someterá a corrección de estilo, composición tipográfica y revisión de la prueba resultante antes de ser publicado en su forma definitiva. Tenga en cuenta que durante el proceso de producción pueden descubrirse errores que podrían afectar al contenido, y que se aplican todas las exenciones de responsabilidad legales que se aplican a la revista.

Acta Biol Colomb (2025); **30(1)**: pag–pag

Doi: <https://doi.org/10.15446/abc.v30n1.111588>

Available online at: <http://www.revistas.unal.edu.co/index.php/actabiol>



Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the [Attribution-NonCommercial-ShareAlike 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) international (CC BY-NC-SA 4.0)

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

**ANÁLISIS DE REDES COMPLEJAS PARA GENES ORTÓLOGOS EN  
VERTEBRADOS: UN ESTUDIO A NIVEL GENÉTICO DE LA EVOLUCIÓN**

**Complex Network Analysis of Orthologous Genes in Vertebrates: A Genetic Study of  
Evolution**

**Título corto: ANÁLISIS DE REDES COMPLEJAS PARA GENES ORTÓLOGOS**

Juan ORJUELA VÁSQUEZ <sup>1a</sup>

Jaime HURTADO-GIRALDO <sup>2b</sup>

Víctor BUCHELI <sup>3c\*</sup>

<sup>1.</sup> Facultad de Salud, Universidad del Valle, Calle 4B # 36-00, Cali, Colombia,  
[juan.orjuela@correounivalle.edu.co](mailto:juan.orjuela@correounivalle.edu.co).

<sup>2</sup> Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Calle 13 # 100-00, Cali, Colombia, [jaime.hurtado@correounivalle.edu.co](mailto:jaime.hurtado@correounivalle.edu.co).

<sup>3</sup> Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Calle 13 # 100-00, Cali, Colombia, [victor.bucheli@correounivalle.edu.co](mailto:victor.bucheli@correounivalle.edu.co).

<sup>a</sup> ORCID: 0009-0009-3788-3605

<sup>b</sup> ORCID: 0009-0000-0824-389X

<sup>c</sup> ORCID: 0000-0002-0885-8699

\* *For correspondence:* [victor.bucheli@correounivalle.edu.co](mailto:victor.bucheli@correounivalle.edu.co)

**Recibido:** 22 de enero de 2024. **Revisado:** 07 de mayo de 2024. **Aceptado:** 29 de agosto de 2024

**Editora asociada:** Rosa Angélica Plata Rueda

**Citation/ citar este artículo como:** Análisis de redes complejas para genes ortólogos. (2025). Análisis de redes complejas para genes ortólogos en vertebrados: un estudio a nivel genético de la evolución. *Acta Biol Colomb.*, 30(1), XX-XX. <https://doi.org/10.15446/abc.v30n1.111588>

## RESUMEN

La especiación es un concepto fundamental en la biología evolutiva. En este trabajo, se analiza dicho concepto utilizando redes complejas. Se estudió el gen ortólogo apolipoproteína E (APOE) presente en vertebrados para identificar información sobre la clasificación y evolución de las especies. Desde el portal del NCBI, se descargó la

secuencia genética del gen APOE para 239 especies. Para el análisis de los datos, se generó una matriz de identidad a partir de alineamientos, y posteriormente dicha matriz se transformó en matrices de adyacencia para llevar a cabo los análisis de redes complejas. Los resultados muestran relaciones de tipo evolutivo entre las especies y confirman la especiación como un proceso evolutivo a nivel genético. Además, indican que la evolución a nivel de genes ortólogos presenta las relaciones taxonómicas y de derivación entre especies. En el estudio, se identificó que las redes de genes ortólogos son similares a las redes de pequeño mundo y que éstas, a su vez, permiten reconocer automáticamente las especies a nivel de orden taxonómico como grupos inherentes a la evolución. El método propuesto permite obtener evidencia de la teoría de la evolución a nivel genético y es una alternativa para hacer clasificación taxonómica semi-automática.

**Palabras clave:** Alineación de secuencia, Apolipoproteínas, Filogenia, Redes de genes, Redes semánticas.

### **ABSTRACT**

This study addresses the speciation concept within the evolutionary biology context, employing complex network approaches. It primarily focuses on the analysis of the orthologous gene apolipoprotein E (APOE), which is present in vertebrates, to investigate species classification and evolution. Genetic sequences of APOE from 239 species were collected from the NCBI portal. The data were analyzed using an identity matrix generated from alignments, subsequently transformed into adjacency matrices for complex network analysis. The results revealed evolutionary relationships among species and affirmed

specialization as a genetic evolutionary process. Furthermore, it was observed that orthologous gene networks resemble "small-world" networks and possess the ability to automatically identify species at the taxonomic order level, enhancing our understanding of evolutionary groupings. This proposed approach provides evidence for the theory of evolution at the genetic level and offers a semi-automated alternative for taxonomic classification.

**Keywords:** Apolipoproteins, Gene Networks, Phylogeny, Semantic Networks, Sequence Alignment.

## INTRODUCCIÓN

La capacidad para generar información genética, almacenarla, procesarla y compartirla ha incrementado en las últimas décadas (Perez De Souza et al., 2020). Estas fuentes de datos son ricas en patrones e información (Sreeja et al., 2019) con el potencial de contribuir en trabajos de clasificación de las especies. Una aproximación para el análisis de sistemas biológicos es representarlos como una red (Perez De Souza et al., 2020). Una red o grafo permite definir un conjunto de nodos y encontrar las diferentes estructuras de relaciones entre ellos (Wasserman y Faust, 1994) y puede usarse como una generalización simple, estructurada y eficiente de sistemas biológicos complejos (Sreeja et al., 2019). Por ejemplo, se usa el agrupamiento de genes bajo la teoría espectral de grafos para recuperar comunidades de genes, relacionados con el trastorno del espectro autista, mejorando la eficiencia en la identificación respecto a otros métodos tradicionales y contribuyendo a diagnósticos y tratamientos (Sreeja et al., 2019).

Este trabajo se basa en redes complejas. Estas describen la estructura de las interrelaciones o interacciones (Valavanis et al., 2010) entre los nodos de una red. De esta manera, la evolución a nivel genético se entiende no sólo como un conjunto de elementos independientes (Telesford et al., 2011), sino como un modelo descrito por sus relaciones y estructuras matemáticas (Valavanis et al., 2010). Las propiedades y características del mismo son representadas por una red que puede ser caracterizada por medidas topológicas (Valavanis et al., 2010) tales como el coeficiente de agrupamiento (*clustering coefficient*), la longitud promedio de la ruta (*average path length*) o medidas de eficiencia (Telesford et al., 2011), entre otras. Con estas medidas, es posible clasificar la red como aleatoria, libre de escala o de mundo pequeño (*small-worldness*), evidenciando información complementaria que puede resultar de gran interés para el sistema estudiado (Valavanis et al., 2010).

En este trabajo se presenta un método para la clasificación semiautomática de las especies y para ello se estudiarán los genes ortólogos. Estos genes permiten describir las diferencias y similitudes en la composición de los diferentes genomas a partir de un gen ancestral, confirmando eventos de especiación (Gabaldón y Koonin, 2013). Así, se usó el gen ortólogo Apolipoproteína E (APOE) para 239 especies de animales. Este gen desempeña un papel importante en la unión, el transporte de lípidos y es un miembro de la familia de las apolipoproteínas ampliamente distribuidas en vertebrados (Fang et al., 2010; Liu et al., 2019).

Para este estudio, se propuso una red cuyos nodos fueron cada una de las especies y los enlaces representan una característica genética común entre cada par de especies. En este

caso, la característica utilizada fue el porcentaje de identidad genética entre dos especies analizadas, calculado a partir de los alineamientos del gen APOE. Los resultados muestran que fue posible calcular métricas para análisis de redes a partir de las diferentes secuencias del gen. Este análisis reveló características emergentes de los datos genéticos como comunidades (agrupamiento de las especies), especies de importancia en cada comunidad, especies de eslabón entre dos conjuntos de especies. Asimismo, se determinó que el tipo de la red es de mundo pequeño para identidades de entre el 50 al 80 %, valores para los cuales es mejor la clasificación de las especies. Finalmente, las comunidades obtenidas fueron cotejadas con las clasificaciones taxonómicas descritas en trabajos de biología evolutiva, descargables desde National Center for Biotechnology Information (NCBI, 2023) con resultados prometedores.

## **MATERIALES Y MÉTODOS**

A continuación, se presenta el procedimiento metodológico (Fig. 1) para desarrollar un análisis de redes complejas con información genética estudiando el gen ortólogo APOE.

### **BÚSQUEDA DE SECUENCIAS Y ALINEAMIENTO**

Se encontraron secuencias del gen APOE para 239 especies (NCBI, 2022), que se listaron y alojaron en un repositorio (Hurtado, 2023). Posteriormente, se procedió a realizar un alineamiento de las secuencias a partir de un único archivo en formato FASTA, utilizando el algoritmo ClustalW (Thompson et al., 1994) implementado en el software Bioedit (Alzohairy, 2011).

## **MATRIZ DE IDENTIDAD**

A partir del alineamiento se calculó una matriz de identidad de las secuencias (Hurtado, 2023) en el software Bioedit (Alzohairy, 2011) y se generó un archivo separado por tabuladores. Las filas y columnas de la matriz representan cada especie y los coeficientes indican el porcentaje de identidad entre las secuencias genéticas de las dos especies referenciadas. Los valores de los coeficientes de la matriz de identidad están en un rango de cero a uno, es decir, representan un porcentaje de qué tan parecidas son las secuencias genéticas entre las dos especies a las que hace referencia el coeficiente.

## **MATRICES DE ADYACENCIA**

Con la matriz de identidad se calcularon matrices de adyacencia mediante un proceso (Hurtado, 2023) en el que las relaciones representadas como porcentajes de la matriz de identidad son transformadas a una relación binaria en las matrices de adyacencia, que muestra si la relación existe o no. Por ejemplo, para la matriz de adyacencia del 10 %, se pone en cada coeficiente el valor de uno si la identidad entre dos especies es superior al 10 % y un coeficiente de cero en caso contrario. Se calcularon matrices de adyacencia para valores de identidad entre 10 y 100 % con intervalos de 10 %.

## **ANÁLISIS DE REDES Y MÉTRICAS**

Cada matriz de adyacencia obtenida representa una red compleja que puede analizarse usando metodologías de ciencias de redes (León et al., 2018; Leytón Yela et al., 2022). Para ello, se calculan y analizan métricas como centralidad por intermediación, estructura

de comunidades y diversas representaciones gráficas de las redes con algunos parámetros representativos como el coeficiente de agrupamiento y la longitud promedio de la ruta. La lógica empleada para este análisis (Hurtado, 2023) se desarrolló en libretas de Wolfram Mathematica (Wolfram Research Inc., 2022).

## **RANGO CLASIFICACIÓN ESPECIES**

Teniendo en cuenta las métricas calculadas, se delimitó un rango para escoger matrices de adyacencia entre 50 y 80 %. Este rango se definió al considerar las características de las redes obtenidas y sus métricas (Perera et al., 2017), dado que en estos valores de identidad se observan parámetros que corresponden a redes de mundo pequeño.

## **DETERMINACIÓN DE TIPO DE RED**

Sumado a las métricas obtenidas para redes, se generaron grafos aleatorios siguiendo el modelo de Erdős-Rényi, y se obtuvo la métrica de coeficiente de mundo pequeño (*small-worldness*) empleando la ecuación  $sw = \frac{\gamma}{\lambda}$ , donde  $\gamma = \frac{Cg}{Cn}$ ,  $\lambda = \frac{lg}{ln}$ .  $Cg$  y  $Cn$  son los coeficientes de agrupamiento de la red evaluada y la red aleatoria simulada, respectivamente, y  $lg$  y  $ln$  son las longitudes promedio del camino de la red evaluada y la red aleatoria simulada (Humphries y Gurney, 2008). La obtención de la red aleatoria simulada se realiza con el mismo número de nodos y ejes de la red biológica propuesta.

## CLASIFICACIÓN FILOGENÉTICA

A partir de la red representada por la matriz de adyacencia, se obtuvo la clasificación automática de comunidades (Wolfram Research Inc., 2023). Adicionalmente, se consultó para cada nodo-especie la taxonomía en una base de datos (NCBI, 2023). Cada comunidad representa una lista de especies, y al etiquetar la taxonomía consultada, se evidencia la concordancia de la clasificación taxonómica a nivel de orden con las comunidades encontradas automáticamente.

## RESULTADOS

El análisis de las diferentes redes obtenidas a partir de cada matriz dio como resultado múltiples grafos, de los cuales se escogieron los de mayor claridad respecto a relaciones filogenéticas y evolutivas. Los grafos generados permitieron identificar casos con un claro agrupamiento de especies (Fig. 2) estrechamente relacionadas entre sí, al menos desde el conocimiento general de sus relaciones filogenéticas (historia evolutiva).

En la red de especies de 80 % de identidad (Fig. 3), se muestran los nodos de mayor importancia con colores que tienden al amarillo de acuerdo con el valor de la métrica de centralidad por intermediación calculada para cada especie de la red. Esta métrica sugiere especies eslabón entre otras especies o clasificaciones. Adicionalmente, se presentan algunos nodos con imágenes representativas de las especies, mostrando la obtención de comunidades. En la primera comunidad, se observa que todas las especies corresponden al orden Chiroptera (Mamíferos voladores, murciélagos). En este grupo, la especie *Myotis myotis* (216) resalta porque es un eslabón entre los diferentes individuos del grupo, es decir,

tiene una similaridad genética desde la cual se puede derivar hacia las otras especies del mismo orden. Análisis similares se pueden realizar para los otros grupos representados en la red (Fig. 3).

En el grupo orden Primates (Fig. 3), tenemos especies representativas como *Propithecus coquereli* (103) y *Microcebus murinus* (104) con mayores valores de centralidad por intermediación. Otros órdenes de importancia agrupados por la red biológica creada fueron Artiodactyla (Mamíferos ungulados, incluye cetáceos), Carnivora (Mamíferos con caninos muy desarrollados), Cichliformes (Peces cíclidos, óseos), Rodentia (Mamíferos de incisivos afilados), Cyprinodontiformes (Peces de agua dulce), Testudines (Reptiles, tortugas). Estos datos revelan que cada especie en cada comunidad corresponde al mismo grupo a nivel de orden en la taxonomía. Por tanto, se muestra que el análisis de redes permite la clasificación de especies a nivel de orden taxonómico como grupos inherentes a la evolución.

El resultado anterior fue comparado (Fig. 4) con la taxonomía recuperadora de las especies (NCBI, 2023) donde cada color corresponde a las comunidades obtenidas en este trabajo. Por otra parte, se anotó para cada especie el orden taxonómico, como se evidencia en la clasificación de comunidades (Hurtado, 2023). En esta clasificación, se muestra que la comunidad uno es la más grande con 27 especies, entre ellas el *Homo sapiens*. De esta comunidad, se obtuvo que la clasificación de taxonomía era del orden Primates, con solo una especie de un orden diferente, *Tupaia chinensis*, que corresponde al orden Scandentia. Se observa dentro del grupo Primates que es una de las especies externas al grupo principal, nodo 64. Estudios han discutido sobre la clasificación de *T. chinensis* dentro del orden de

Primates (Yao, 2017), encontrando un gran potencial frente a los roedores como animal de experimentación para el modelado de enfermedades humanas y respuestas terapéuticas. Las comunidades dos con 25 especies y tres con 23 especies fueron clasificadas en los órdenes Carnivora y Artiodactyla, sin ninguna especie de orden diferente.

Otro resultado obtenido (Fig. 5) muestra el grafo de comunidades, el *coeficiente de agrupamiento* (CC), la *longitud promedio del camino* (APL) y la métrica *small-worldness* (SW) para diferentes porcentajes de identidad (ID). Se encontró que entre más baja la identidad entre las secuencias, se definen comunidades con mayor número de nodos. Al aumentar la identidad entre cada par de nodos de las especies seleccionadas para la generación de las redes, las comunidades se obtienen con menor cantidad de nodos, lo que indicaría una clasificación más específica. En el cálculo de la métrica SW para definir el tipo de red se evidencia (Fig. 5) que mientras el porcentaje de identidad se mantiene bajo (< 40 % de identidad), el valor de Small-worldness está entre 1 y 2,45. A partir de 50 % de identidad, el valor incrementa de forma pronunciada.

En este sentido, se muestra que las redes obtenidas son similares a una red de mundo pequeño debido principalmente a que SW tiene valores muy superiores a uno. Este comportamiento sugiere que el proceso de evolución sigue un patrón donde las especies no tienen una relación aleatoria con sus predecesores, sino que, en la medida que hay una adaptación, la especialización va siguiendo una estructura similar a la de una red de mundo pequeño. En este tipo de estructura de red, se observan grupos cuyos individuos están conectados con las demás especies relacionadas en pocos saltos, como también se evidencia en los valores de APL (Fig. 5). Estos resultados muestran que las especies en las diferentes

comunidades obtenidas presentan una alta similitud y, pese a ser un agrupamiento automatizado, revelan concordancia (Fig. 4) con la clasificación a nivel taxonómico de orden consultado (NCBI, 2023).

Los resultados obtenidos muestran entonces que las variaciones en el porcentaje de identidad implican cambios en la complejidad de la red, y para efectos de analizar interacción entre las especies, se puede obtener información de importancia en un rango de identidad entre el 50 y 80 %. Para rangos menores o mayores, los análisis cambiarían, por lo que la decisión del porcentaje de identidad en el rango definido es un factor clave en trabajos que se deseen hacer con un enfoque similar al presentado en líneas anteriores.

## DISCUSIÓN



Entre las herramientas computacionales más reconocidas para el análisis de secuencias genómicas se encuentran Bioedit (Alzohairy, 2011), MEGA (Kumar et al., 2018) y Gene Runner (Shapouri Moghaddam et al., 2020) con diversas capacidades para la clasificación y filogenia de las especies. Aun así, el análisis de redes se muestra como una alternativa simple con potencial de revelar información de interés en estudios de estos sistemas biológicos. En este trabajo fue posible identificar relaciones filogenéticas de forma automática. El análisis de redes complejas agrupó las especies por órdenes y familias, consistente con la filogenia de los vertebrados producto de los diferentes estudios a nivel evolutivo y filogenético (NCBI, 2023).

Como la red biológica seleccionada para este estudio es del tipo mundo pequeño, se puede inferir que el gen APOE ha evolucionado entre las especies de forma no aleatoria, sino que

las presiones de la adaptación han logrado que el gen esté presente en grupos de especies fuertemente relacionadas entre sí y, al mismo tiempo, pueda estar presente en grupos de especies de diferente orden a pocos saltos en la red. De acuerdo con los resultados obtenidos, se cree que el gen APOE puede explicar los cambios en la estructura ósea en el proceso evolutivo, en la divergencia de peces cartilagosos a óseos (Liu et al., 2019) observada en los gráficos obtenidos. Así, su función no solo es clave en el metabolismo de lípidos, sino también en el metabolismo óseo, por lo que su alteración puede estar relacionada con enfermedades como la osteoporosis y la arteriosclerosis, así como el Alzheimer en los humanos (Noguchi et al., 2018).

El gen APOE exhibe funciones específicas en diferentes órdenes de la taxonomía de especies, lo que se traduce en comunidades de especies identificadas en este estudio (Xu et al., 1996; Witting et al., 2000). En primates, incluidos los humanos, APOE está estrechamente relacionado con la regulación del metabolismo cerebral, la neuroprotección y la respuesta a lesiones cerebrales (McIntosh et al., 2012). Por otro lado, en roedores como ratones y ratas, APOE desempeña un papel fundamental en el metabolismo lipídico, el transporte de colesterol y la modulación de respuestas inmunológicas y plasticidad neuronal (Myers y McGonigle, 2019). Además, APOE contribuye a la respuesta inmunológica, defendiendo contra patógenos y manteniendo la salud celular en entornos propensos a la contaminación y cambios estacionales, como se observó en peces, en los cuales este gen podría disminuir los niveles de expresión de factores proinflamatorios, vías relacionadas con la respuesta inmunológica y apoptosis, al tiempo que aumenta los niveles de expresión de factores antiinflamatorios (Feng et al., 2023). La variabilidad genética observada entre

las comunidades estudiadas refleja adaptaciones específicas a las necesidades de cada orden de especies.

Respecto al modelo comparativo para redes, se obtuvo para el rango seleccionado de 50 a 80 % que la red se comporta como mundo pequeño, que derivado del modelo de Watts-Strogatz, que es utilizado en el estudio de diferentes campos, incluyendo la ecología, economía, epidemiología y neurociencia (Humphries y Gurney, 2008). Este tipo de estructura muestra cierta robustez y eficiencia en términos de compartir información genética en todo el sistema de especies, lo que podría asegurar su adaptabilidad y persistencia en el proceso evolutivo. Así, la construcción de redes de especies a partir de secuencias de genes ortólogos permite reconocer automáticamente las especies a nivel taxonómico de orden (Fig. 4). Sin embargo, este tipo de aproximaciones también podría realizarse con el análisis de la secuencia de la proteína, para considerar aspectos relacionados con dominios específicos, lo cual es útil para identificar y caracterizar regiones funcionales dentro de la proteína, comprender las interacciones moleculares y predecir posibles funciones biológicas (Wang et al., 2021). Por ejemplo, el gen APOE en humanos es polimórfico, presenta tres alelos principales (E2, E3, E4), lo que se asocia con variaciones fenotípicas en varios rasgos médicos, incluyendo los niveles de colesterol, la salud cardiovascular, el riesgo de enfermedad de Alzheimer y la longevidad (McIntosh et al., 2012; Tudorache et al., 2017).

## CONCLUSIONES

Los esfuerzos recientes en el uso de redes complejas para el análisis de datos, como aquellas enfocadas en secuencias génicas, muestran resultados prometedores para ampliar el conocimiento de cómo los genes u otras entidades moleculares, especies o entidades biológicas están íntimamente relacionados y su potencial en campos como la biología evolutiva, farmacéutica, e ingeniería biomédica. De esta forma, el método propuesto permite una alternativa para hacer clasificación taxonómica semi-automática.

El análisis de redes basado en matrices de similitud genética revela estructuras complejas que agrupan especies según relaciones filogenéticas y evolutivas. Por otra parte, los grafos generados destacaron comunidades claramente definidas donde especies estrechamente relacionadas compartían nodos significativos, indicando una agrupación taxonómica coherente. Adicionalmente, este enfoque demostró ser una herramienta poderosa para el estudio de sistemas biológicos complejos, proporcionando perspectivas sobre la evolución y la adaptación genética.

La simplicidad del modelo radicó en su capacidad para transformar datos genéticos complejos en representaciones vectoriales claras y estructuradas, que permiten computar métricas y obtener visualizaciones mediante el análisis de redes complejas. Estas facilitan la identificación de comunidades taxonómicas y nodos de interés, proporcionando una comprensión intuitiva y accesible de la estructura y la dinámica evolutiva de los sistemas biológicos estudiados.

Para futuras investigaciones, se sugiere incorporar un análisis comparativo de los dominios proteicos. Este enfoque permitirá identificar y contrastar las variaciones estructurales y funcionales de las proteínas entre distintas especies, proporcionando una comprensión más profunda y detallada acerca de su evolución, adaptación y función biológica. Además, con este tipo de análisis se podrían establecer correlaciones con otros factores como las presiones ambientales y las necesidades fisiológicas específicas de cada especie.

### **PARTICIPACIÓN DE AUTORES**

Juan ORJUELA VASQUEZ: Conceptualización, investigación, cálculo de alineamientos, análisis de redes, redacción y edición.

Jaime HURTADO-GIRALDO: Investigación, cálculo de matrices de adyacencia, cálculos de analítica de redes, redacción y edición.

Víctor BUCHELI: Conceptualización, investigación, análisis de redes, redacción, revisión y edición.

### **AGRADECIMIENTOS**

A la Universidad del Valle, Facultad de Ingeniería, Escuela de Ingeniería de Sistemas y Computación por la financiación de asistencia de docencia de Jaime Hurtado Giraldo y Facultad de Salud, Escuela de Ciencias Básicas, Departamentos de Fisiología y Microbiología por la financiación de Juan Orjuela Vasquez y por los recursos representados en tiempo de profesor destinados a la investigación.

## CONFLICTO DE INTERESES

Los autores declaran no tener ningún conflicto de intereses.

## REFERENCIAS

- Alzohairy, A. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, 2, 60-61.
- Fang, G., Bhardwaj, N., Robilotto, R., y Gerstein, M. B. (2010). Getting Started in Gene Orthology and Functional Analysis. *PLOS Computational Biology*, 6(3), e1000703. <https://doi.org/10.1371/journal.pcbi.1000703>
- Feng, J., Huang, Y., Huang, M., Li, X., Amoah, K., Huang, Y., y Jian, J. (2023). Apolipoprotein Eb (On-ApoEb) protects *Oreochromis niloticus* against *Streptococcus agalactiae* infection. *Fish & Shellfish Immunology*, 141, 109069. <https://doi.org/10.1016/j.fsi.2023.109069>
- Gabaldón, T., y Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, 14(5), 360-366. <https://doi.org/10.1038/nrg3456>
- Humphries, M. D., y Gurney, K. (2008). Network ‘Small-World-Ness’: A Quantitative Method for Determining Canonical Network Equivalence. *PLoS ONE*, 3(4), e0002051. <https://doi.org/10.1371/journal.pone.0002051>
- Hurtado, J. (28 de agosto de 2023). *Jandresh/species-network*. <https://github.com/jandresh/species-network>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., y Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology*

- and Evolution*, 35(6), 1547-1549. <https://doi.org/10.1093/molbev/msy096>
- León, D. A., Valdivia, J. A., y Bucheli, V. A. (2018). Modeling of Colombian Seismicity as Small-World Networks. *Seismological Research Letters*, 89(5), 1807-1816. <https://doi.org/10.1785/0220180076>
- Leytón Yela, G. V., Valencia, A. J. y Bucheli Guerrero, V. A. (2022). Relaciones sociales y comunidades en una plataforma virtual para cursos de programación de computadores. *Ingeniería*, 27(2), e17797. <https://doi.org/10.14483/23448393.17797>
- Liu, J.-Q., Li, W.-X., Zheng, J.-J., Tian, Q.-N., Huang, J.-F., y Dai, S.-X. (2019). Gain and loss events in the evolution of the apolipoprotein family in vertebrata. *BMC Evolutionary Biology*, 19(1), 209. <https://doi.org/10.1186/s12862-019-1519-8>
- McIntosh, A. M., Bennett, C., Dickson, D., Anestis, S. F., Watts, D. P., Webster, T. H., Fontenot, M. B., y Bradley, B. J. (2012). The Apolipoprotein E (APOE) Gene Appears Functionally Monomorphic in Chimpanzees (*Pan troglodytes*). *PLoS ONE*, 7(10), e47760. <https://doi.org/10.1371/journal.pone.0047760>
- Myers, A., y McGonigle, P. (2019). Overview of Transgenic Mouse Models for Alzheimer's Disease. *Current Protocols in Neuroscience*, 89(1), e81. <https://doi.org/10.1002/cpns.81>
- NCBI. (2 de noviembre de 2022). *APOE orthologs*. NCBI. <https://www.ncbi.nlm.nih.gov/gene/348/ortholog/>
- NCBI. (23 de junio de 2023). *Taxonomy—NCBI*. <https://www.ncbi.nlm.nih.gov/taxonomy>
- Noguchi, T., Ebina, K., Hirao, M., Otsuru, S., Guess, A. J., Kawase, R., Ohama, T., Yamashita, S., Etani, Y., Okamura, G., y Yoshikawa, H. (2018). Apolipoprotein E plays crucial roles in maintaining bone mass by promoting osteoblast differentiation

- via ERK1/2 pathway and by suppressing osteoclast differentiation via c-Fos, NFATc1, and NF- $\kappa$ B pathway. *Biochemical and Biophysical Research Communications*, 503(2), 644-650. <https://doi.org/10.1016/j.bbrc.2018.06.055>
- Perera, S., Bell, M. G. H., y Bliemer, M. C. J. (2017). Network science approach to modelling the topology and robustness of supply chain networks: A review and perspective. *Applied Network Science*, 2(1), 33. <https://doi.org/10.1007/s41109-017-0053-0>
- Perez De Souza, L., Alseekh, S., Brotman, Y., y Fernie, A. R. (2020). Network-based strategies in metabolomics data analysis and interpretation: From molecular networking to biological interpretation. *Expert Review of Proteomics*, 17(4), 243-255. <https://doi.org/10.1080/14789450.2020.1766975>
- Shapouri Moghaddam, A., Mansouri, S., Neshani, A., Firoozeh, F., Matinpur, A., Khaledi, A., y Ghazalibina, M. (2020). Construction, Cloning, and Expression of CagA Recombinant Protein of Helicobacter pylori. *Avicenna Journal of Medical Biotechnology*, 12(2), 135-138.
- Sreeja, A., Krishnakumar, U., y Vinayan, K. P. (2019). Functional Categorization of Disease Genes Based on Spectral Graph Theory and Integrated Biological Knowledge. *Interdisciplinary Sciences: Computational Life Sciences*, 11(3), 460-474. <https://doi.org/10.1007/s12539-017-0279-7>
- Telesford, Q. K., Simpson, S. L., Burdette, J. H., Hayasaka, S., y Laurienti, P. J. (2011). The Brain as a Complex System: Using Network Science as a Tool for Understanding the Brain. *Brain Connectivity*, 1(4), 295-308. <https://doi.org/10.1089/brain.2011.0055>

- Thompson, J. D., Higgins, D. G., y 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. <https://doi.org/10.1093/nar/22.22.4673>
- Tudorache, I. F., Trusca, V. G., y Gafencu, A. V. (2017). Apolipoprotein E - A Multifunctional Protein with Implications in Various Pathologies as a Result of Its Structural Features. *Computational and Structural Biotechnology Journal*, 15, 359-365. <https://doi.org/10.1016/j.csbj.2017.05.003>
- Valavanis, I., Spyrou, G., y Nikita, K. (2010). A similarity network approach for the analysis and comparison of protein sequence/structure sets. *Journal of Biomedical Informatics*, 43(2), 257-267. <https://doi.org/10.1016/j.jbi.2010.01.005>
- Wang, Y., Zhang, H., Zhong, H., y Xue, Z. (2021). Protein domain identification methods and online resources. *Computational and Structural Biotechnology Journal*, 19, 1145-1153. <https://doi.org/10.1016/j.csbj.2021.01.041>
- Wasserman, S., y Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815478>
- Witting, P. K., Pettersson, K., Letters, J., y Stocker, R. (2000). Anti-atherogenic effect of coenzyme Q10 in apolipoprotein E gene knockout mice. Dedicated to Lars Ernster for his pioneering contributions to research into coenzyme Q and his genuine passion and enthusiasm we were privileged to experience. *Free Radical Biology and Medicine*, 29(3-4), 295-305. [https://doi.org/10.1016/S0891-5849\(00\)00311-7](https://doi.org/10.1016/S0891-5849(00)00311-7)
- Wolfram Research Inc. (1 de noviembre de 2022). *Wolfram Mathematica: Computación técnica moderna*. <https://www.wolfram.com/mathematica/>

Wolfram Research Inc. (26 de junio de 2023). *FindGraphCommunities*.

<https://reference.wolfram.com/language/ref/FindGraphCommunities.html>

Xu, P.-T., Schmechel, D., Rothrock-Christian, T., Burkhart, D. S., Qiu, H.-L., Popko, B.,

Sullivan, P., Maeda, N., Saunders, A. M., Roses, A. D., y Gilbert, J. R. (1996).

Human Apolipoprotein E2, E3, and E4 Isoform-Specific Transgenic Mice: Human-like Pattern of Glial *and* Neuronal Immunoreactivity in Central Nervous System Not Observed in Wild-Type Mice. *Neurobiology of Disease*, 3(3), 229-245.

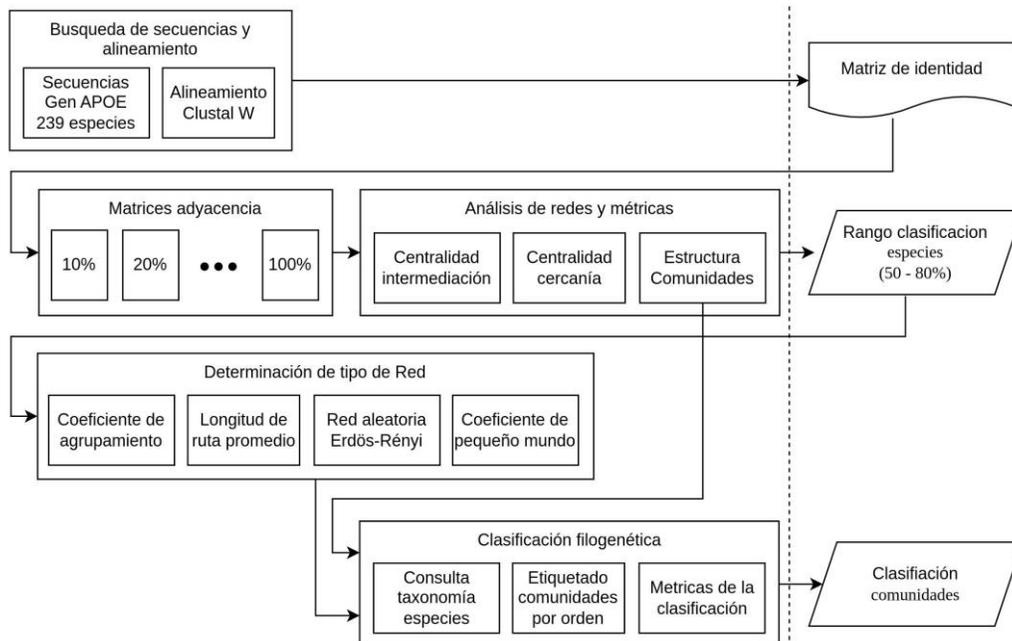
<https://doi.org/10.1006/nbdi.1996.0023>

Yao, Yong-Gang. (2017). Creating Animal Models, Why Not Use the Chinese Tree Shrew

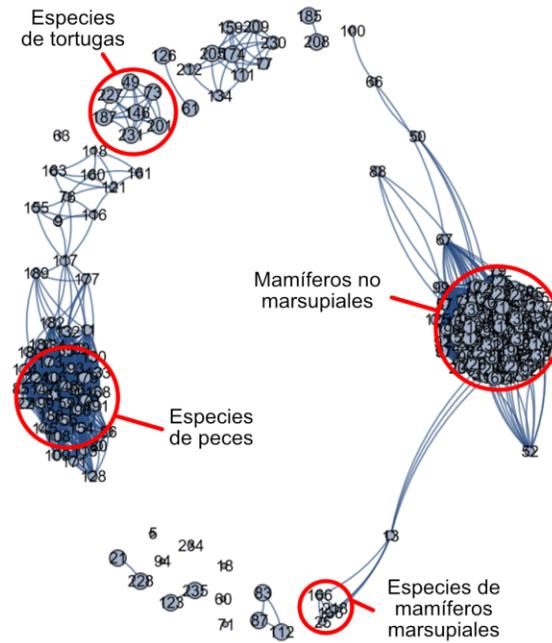
(*Tupaia belangeri chinensis*)?. *Zoological Research*, 38(3), 118-126.

<https://doi.org/10.24272/j.issn.2095-8137.2017.032>

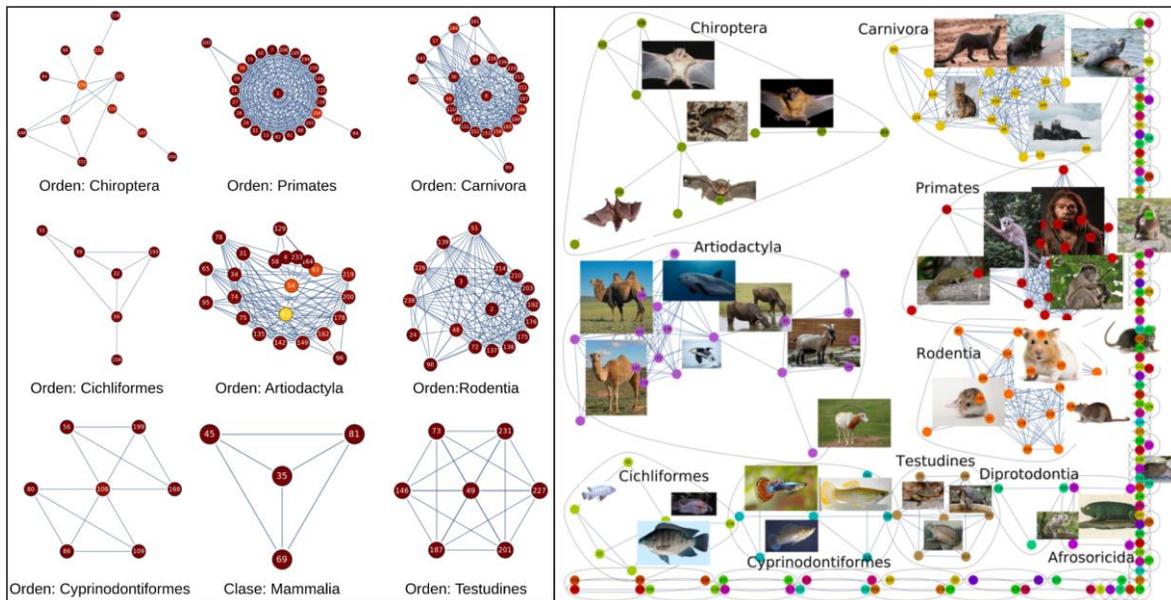
## TABLAS Y FIGURAS



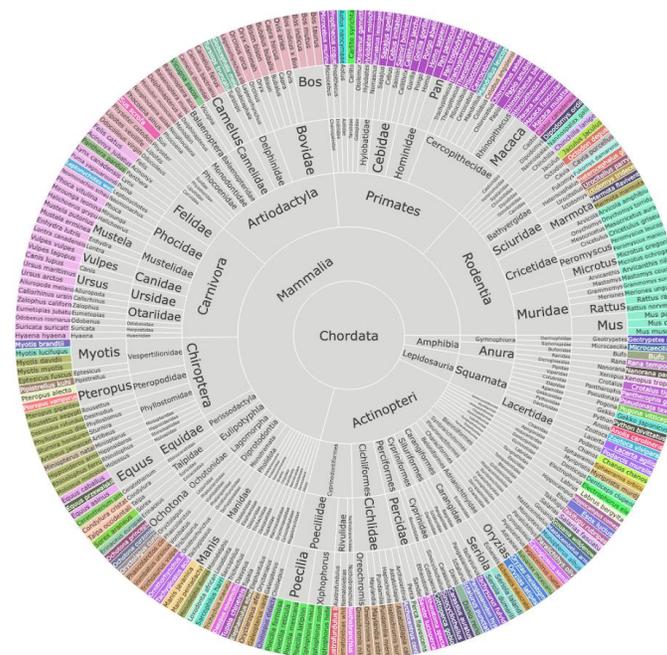
**Figura 1.** Metodología de construcción de una red biológica a partir de las secuencias genéticas del gen ortólogo APOE.



**Figura 2.** Red APOE para el porcentaje 50 % de identidad.

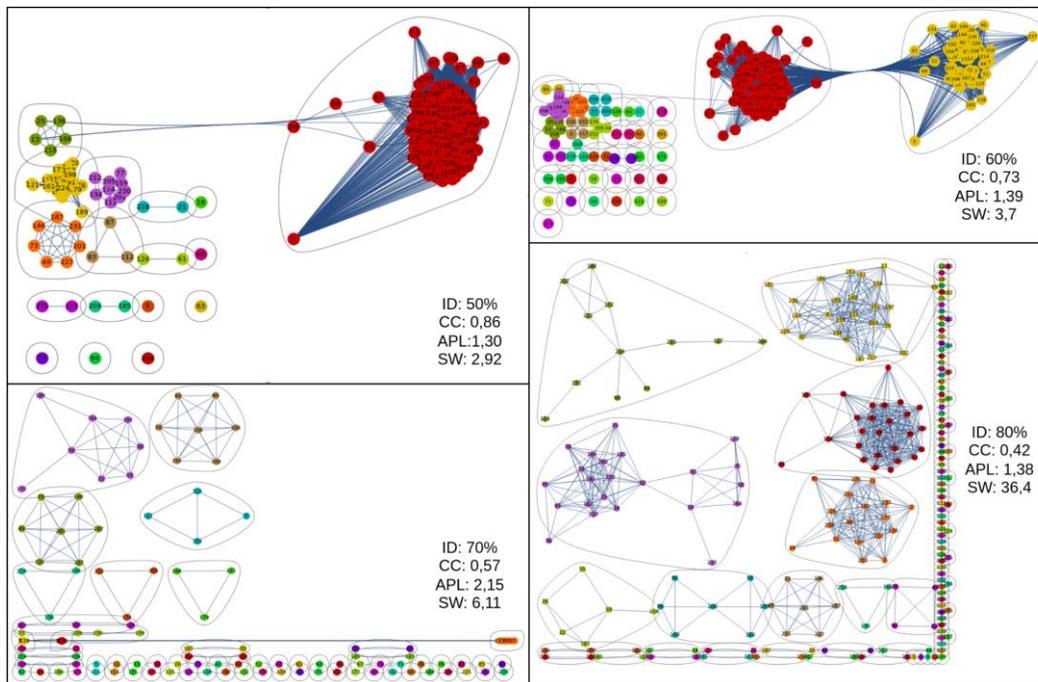


**Figura 3.** APOE 80 % centralidad por intermediación con anotación del orden taxonómico encontrado e ilustraciones con imágenes de algunas especies.



**Figura 4.** Taxonomía de especies por colores de acuerdo a las comunidades obtenidas.

Accept



**Figura 5.** Comunidades y métricas para APOE por porcentajes de identidad.

Accepted