

Proteínas multifuncionales: desplazamiento de gen no ortólogo como posible mecanismo para obtener una función *moonlighting*.

Rocío López Martínez

Máster en bioinformática y bioestadística
Área 2

Director: Dr. Luis Franco Serrano

PRA: Dr. Carles Ventura Royo

24 de junio de 2020



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Proteínas multifuncionales: desplazamiento de gen no ortólogo como posible mecanismo para obtener una función <i>moonlighting</i>.
Nombre del autor:	Rocío López Martínez
Nombre del consultor/a:	Luis Franco Serrano
Nombre del PRA:	Carles Ventura Royo
Fecha de entrega (mm/aaaa):	06/2020
Titulación:	Máster en bioinformática y bioestadística
Área del Trabajo Final:	Área 2, subárea de bioinformática y proteómica en virulencia y enfermedades humanas
Idioma del trabajo:	Español
Palabras clave	Proteínas multifuncionales, proteínas NOGD, evolución

Resumen del Trabajo

Las proteínas *moonlighting* son aquellas que realizan más de una función utilizando la misma cadena polipeptídica. Identificar estas proteínas es necesario tanto para su anotación funcional como por su relevancia en múltiples aspectos bioquímicos y fisiopatológicos, estando implicadas en virulencia de microorganismos y patologías humanas. Es un fenómeno de gran complejidad, pues comprender cómo surgen y coexisten diferentes funciones sin interferir entre ellas supone un reto. Aún quedan muchas cuestiones por resolver como cuál es su origen evolutivo o cuáles son los mecanismos por los que se adquieren estas funciones adicionales.

En este trabajo se ha estudiado la relación entre las proteínas multifuncionales y el *Non Orthologous Gene Displacement* (NOGD), un fenómeno por el que un componente esperado de una vía enzimática está ausente y su función es realizada por otra proteína. Se han encontrado 124 proteínas *moonlighting* que realizan funciones de proteínas NOGD e incluso hay varios ejemplos de proteínas NOGD que son multifuncionales.

Se analizó la relación filogenética entre las proteínas multifuncionales y NOGD que realizan la misma función y se ha visto que las dos proteínas NOGD tienen líneas evolutivas diferentes y que una de ellas es o se relaciona con una proteína multifuncional. Este hecho sugiere que el NOGD puede ser un mecanismo por el cual una proteína podría adquirir una nueva función para sustituir al componente faltante de la vía y se convertiría, por tanto, en una

proteína multifuncional si conserva su función original.

Abstract:

Moonlighting proteins performs more than one function using the same polypeptide chain. Identifying these proteins is necessary for their functional annotation and for their relevance in multiple biochemical and pathophysiological aspects, being implicated in the virulence of microorganisms and human diseases. It is a highly complex phenomenon, since understanding how different functions arise and coexist without interfering with each other is a challenge. Many questions remain to be resolved, such as their evolutionary origin or the mechanisms by which these additional functions are acquired.

In this work, the relationship between multifunctional proteins and Non Orthologous Gene Displacement (NOGD) has been studied. NOGD is a phenomenon in which an expected component of an enzyme pathway is absent and its function is performed by another protein. We have found 124 moonlighting proteins with functions of NOGD proteins and even some examples in which NOGD proteins are multifunctional themselves.

We have analyzed the phylogenetic relationship between multifunctional proteins and NOGD performing the same function. We obtained that NOGD proteins have different evolutionary lines and almost one of them is a multifunctional protein or is related to that. This fact suggests that NOGD may be a mechanism by which a protein could acquire a new function to replace the missing component of the pathway and would therefore become a multifunctional protein if it retains its original function.

Índice

1. Introducción	1
1.1. Contexto y justificación del Trabajo	1
1.2. Objetivos del Trabajo.....	3
1.3. Enfoque y método seguido.....	3
1.4. Planificación del Trabajo	4
1.5. Breve resumen de productos obtenidos	5
1.6. Breve descripción de los otros capítulos de la memoria.....	5
2. Métodos	6
2.1. Bases de datos y herramientas bioinformáticas empleadas	6
2.2. Actualización del listado de proteínas multifuncionales y NOGD que realizan la misma función.	8
2.3. Estudio evolutivo entre las proteínas que realizan la misma función.....	9
3. Resultados	11
3.1. Actualización del listado de proteínas multifuncionales y NOGD que realizan la misma función.	11
3.2. Estudio evolutivo entre las proteínas que realizan la misma función.....	14
3.2.1. Selección de proteínas para el análisis evolutivo.	14
3.2.2. Alineamiento global.	15
3.2.3. PSI-BLAST.	15
3.2.4. Árboles filogenéticos y alineamientos múltiples.....	18
4. Discusión: Fenómeno NOGD como mecanismo de obtención de una función <i>moonlighting</i>	28
5. Conclusiones	30
6. Glosario	32
7. Bibliografía	33
8. Anexos	35

Lista de figuras

- Figura 1.** Diagrama de Gantt de la planificación del trabajo.
- Figura 2.** Filos seleccionados para la construcción de árboles filogenéticos.
- Figura 3.** Árbol filogenético de proteínas con función 6,7-dihidropteridina reductasa.
- Figura 4.** Alineamiento de proteínas con función 6,7-dihidropteridina reductasa.
- Figura 5.** Árbol filogenético de proteínas con función óxido nítrico sintasa.
- Figura 6.** Alineamiento de proteínas con función óxido nítrico sintasa.
- Figura 7.** Árbol filogenético de proteínas con función corismato mutasa.
- Figura 8.** Alineamiento de proteínas con función corismato mutasa.
- Figura 9.** Árbol filogenético de proteínas con función fosfolipasa A2.
- Figura 10.** Árbol filogenético de proteínas con función lisofosfolipasa.
- Figura 11.** Árbol filogenético de proteínas con función α -N-acetilgalactosaminidasa.
- Figura 12.** Árbol filogenético de proteínas con función endo- β -N-acetilglucosaminidasa.
- Figura 13.** Árbol filogenético de proteínas con función ARNt metiltransferasa.

Lista de tablas

- Tabla 1.** Planificación del trabajo.
- Tabla 2.** Ejemplos de proteínas *moonlighting* y NOGD que realizan la misma función.
- Tabla 3.** Proteínas seleccionadas para el análisis evolutivo.
- Tabla 4.** Matrices de identidad obtenidas en Clustal.

1. Introducción

1.1. Contexto y justificación del Trabajo

Las proteínas *moonlighting* o multifuncionales, como su nombre indica, son aquellas que realizan más de una función utilizando la misma cadena polipeptídica. Se llama canónica a la primera función que se descubrió y *moonlighting* a las demás, sin que esta nomenclatura implique mayor importancia para una de las funciones. Estas múltiples funciones se suelen descubrir por serendipia. Una localización inesperada de una proteína conocida, diferente expresión de la misma según el factor que la active, diferentes estados de oligomerización, modificaciones postraduccionales o interacciones con factores inesperados, por ejemplo, pueden sugerir la multifuncionalidad de una proteína [1].

También mediante el uso de la bioinformática se pueden predecir nuevas funciones de una proteína [2], pero estas no se aceptan hasta que se demuestran experimentalmente. Por ello, tampoco se puede suponer la multifuncionalidad de proteínas de organismos evolutivamente cercanos si no se demuestra en cada uno de ellos.

La multifuncionalidad es un fenómeno de gran complejidad, desde su definición hasta su propia existencia, pues se dan diversas combinaciones que dificultan establecer un límite a la hora de considerar o no que una proteína es multifuncional [3]. Diferentes autores discrepan en ciertos aspectos a la hora de definir o identificar estas proteínas. Por ejemplo, para Constance J. Jeffery, quién acuñó el término de proteína multifuncional, no se considerarían aquellas proteínas producto de una fusión génica, mientras otros autores sí lo hacen, incluido el grupo del Dr. Franco-Serrano. Por otro lado, comprender cómo surgen y coexisten diferentes funciones sin interferir entre ellas supone un reto [4].

La primera proteína multifuncional descrita fue la aldehído deshidrogenasa en 1987 [5]. Tras ella, se han descrito numerosas proteínas *moonlighting*, habiendo recogidas 694 en la última base de datos creada por el grupo del Dr. Franco-Serrano [6]. Estas proteínas se presentan de manera ubicua en cualquier organismo y abarcan todo tipo de funciones, siendo las más frecuentes enzimas metabólicas y factores de transcripción. Una misma proteína puede tener la misma función *moonlighting* en diferentes organismos, incluso genéticamente lejanos, o, al contrario, realizar funciones *moonlighting* diferentes en cada uno de ellos.

Se trata de un mecanismo clave para la célula al dotarla de más complejidad sin necesidad de producir más proteínas, diferente del *splicing* alternativo, en el que se obtienen diferentes proteínas a partir de una misma molécula de ARN. Las proteínas multifuncionales realizan diferentes funciones con una misma cadena polipeptídica. Identificar las funciones *moonlighting* de una proteína es importante tanto para su anotación funcional como por su relevancia en múltiples aspectos bioquímicos y fisiopatológicos. Conocer la multifuncionalidad de una proteína es necesario a la hora de interpretar ensayos funcionales,

estudios *Knock-outs* o redes de interactómica, por ejemplo. También para predecir posibles efectos secundarios de los fármacos de los que son diana, de hecho, el 48% de proteínas de la base de datos MultitaskProtDBII son dianas conocidas de fármacos. Así mismo, se han relacionado un 25% de proteínas de esta base de datos con mecanismos de virulencia en microorganismos patógenos. Además, un 78% de las proteínas *moonlighting* humanas están implicadas en patologías [7-9].

Aún quedan muchas cuestiones por resolver acerca de estas proteínas, entre ellas, cuál es su origen evolutivo, si la multifuncionalidad se conserva filogenéticamente o cuáles son los mecanismos por los que se adquieren estas funciones adicionales. Por su relevancia bioquímica y fisiopatológica, es necesario entender los mecanismos de adquisición de nuevas funciones, así como investigar su origen evolutivo para identificar la multifuncionalidad en más proteínas. Esto permitirá poder diseñar proteínas con estas características o fármacos más eficaces contra las mismas.

Se ha propuesto, por ejemplo, que una proteína puede adquirir una nueva función con sólo unas pocas mutaciones clave que afecten a regiones no necesarias para su función canónica. También a través de modificaciones postraduccionales, cambios conformacionales o incluso mutaciones en regiones reguladoras que favorezcan la adquisición de esta nueva función, todo ello promovido por su interacción con nuevos factores [10].

El grupo del Dr. Franco-Serrano hipotetiza que el desplazamiento de gen no ortólogo (NOGD, por sus siglas en inglés) podría ser un mecanismo por el cual una proteína podría adquirir una nueva función. Las proteínas NOGD surgen como una forma alternativa de una vía enzimática en la que un componente esperado está ausente pero su función es realizada por otra proteína. Estas difieren en su estructura, principalmente se trata de proteínas diferentes de distintos organismos que realizan una misma función. Este fenómeno sugiere un mecanismo por el cual una proteína podría adquirir una nueva función para sustituir al componente faltante de la vía y se convertiría, por tanto, en una proteína multifuncional si conserva su función original.

Este fenómeno fue descrito por Koonin et al., quienes crearon la primera base de datos de proteínas NOGD en 1998, que fue actualizada en 2010 [11]. En esta base se encuentran 224 funciones enzimáticas que son realizadas por al menos dos proteínas con estructuras diferentes.

En estudios previos con la primera base de proteínas *moonlighting* creada por el grupo del Dr. Franco-Serrano, se encontró que 26 de estas proteínas estaban también en la base de proteínas NOGD. Se hipotetizó entonces la posible relación entre estos dos fenómenos. En el presente trabajo, nos disponemos a estudiar esta relación con el fin de comprobar si el desplazamiento de gen no ortólogo puede ser un mecanismo por el cuál se adquiera una función *moonlighting*.

1.2. Objetivos del Trabajo

El objetivo principal de este trabajo es estudiar la posible relación entre la multifuncionalidad y el fenómeno NOGD. Para ello, se definen los siguientes objetivos generales:

1. Actualizar el listado de proteínas que son multifuncionales y NOGD con la nueva base MultitaskProtDBII.
2. Realizar un estudio filogenético entre todas las proteínas que realizan la misma función.
3. Evaluar si el fenómeno NOGD podría ser un mecanismo de obtención de una función *moonlighting*.

Para cumplir estos objetivos, se estructurará el trabajo en los siguientes *objetivos específicos*:

1. Actualizar el listado de proteínas que son multifuncionales y NOGD:
 - 1.1. Acceder a las bases de datos de proteínas *moonlighting* y NOGD y analizar su estructura y los datos que contienen.
 - 1.2. Buscar todas aquellas proteínas comunes a ambas bases de datos.
 - 1.3. Crear una tabla con todas estas proteínas e información relevante acerca de las mismas.
2. Realizar un estudio filogenético entre las proteínas que realizan la misma función:
 - 2.1. Realizar una selección entre todas las proteínas del listado anterior y obtener sus secuencias de aminoácidos.
 - 2.2. Realizar una comparación entre las secuencias de las proteínas que realizan la misma función para ver el grado de homología.
 - 2.3. Crear árboles evolutivos para ver el origen de estas proteínas y su relación.
3. Evaluar si el fenómeno NOGD podría ser un mecanismo de obtención de una función *moonlighting*.
 - 3.1. Analizar los árboles evolutivos creados.
 - 3.2. Establecer la relación entre la(s) proteína(s) *moonlighting* y la(s) proteína(s) NOGD.

1.3. Enfoque y método seguido

Partiendo de las bases de datos de proteínas *moonlighting* y NOGD, se recogieron las que estaban en ambas y se realizó un estudio de homología con el objetivo de determinar su origen y la relación filogenética entre ellas. Entre todas las proteínas NOGD que realizan la misma función, si se encontrara que al menos una de las formas estructurales está relacionada evolutivamente con la estructura de la proteína *moonlighting* y la(s) otra(s) no, quedaría demostrado que esas proteínas tienen una línea evolutiva diferente. Esto

sugeriría que la proteína multifuncional podría haber surgido por un mecanismo NOGD, al adquirir una función extra que supliera la falta de la proteína original. Este hecho quedaría demostrado si se encontrara el origen de estas proteínas, comprobando si efectivamente proceden de una proteína que no realizaba esa función antes.

A grandes rasgos, la metodología seguida fué la siguiente:

Para actualizar el listado de proteínas que son multifuncionales y NOGD, se realizó una búsqueda comparativa entre las bases de datos de proteínas *moonlighting* y NOGD para encontrar las que estén en ambas. En caso de incongruencias en la nomenclatura o dificultades para identificar proteínas que realicen la misma función, se recurrió a UniProt para establecer la funcionalidad de las proteínas en cuestión. Se creó una tabla que recoge todas estas proteínas e información relevante acerca del origen y funcionalidad de las mismas.

Una vez creado el listado de todas las proteínas comunes a ambas bases de datos, se seleccionaron unos cuantos ejemplos representativos para realizar un estudio de homología entre todas las proteínas que realizan la misma función. Para ello se obtuvieron sus secuencias de aminoácidos de UniProt y se realizaron estudios con diferentes herramientas de alineamiento y construcción de árboles filogenéticos para establecer relaciones evolutivas entre las mismas.

Con los árboles evolutivos obtenidos se evaluará si el fenómeno NOGD podría ser un mecanismo de obtención de una función *moonlighting*. Si se encontrara que cada forma estructural de la proteína NOGD tiene una línea evolutiva diferente y, una de ellas, procede de otra proteína que no realizaba esa función, se puede asumir que esa proteína de la que procede adquirió esta función para suplir la falta de la proteína que la realizaba originalmente. Si además esta proteína mantiene su función original se convertiría en una proteína multifuncional, aunque esto no es necesario porque puede que la perdiera al evolucionar. En cualquier caso, este hallazgo sugeriría un mecanismo por el cual una proteína adquiriría una nueva función para suplir la falta de otra y, por tanto, relacionaría la multifuncionalidad y el NOGD.

1.4. Planificación del Trabajo

Las primeras semanas se dedicaron a aspectos de planificación, como la definición del trabajo y sus objetivos, el establecimiento de la metodología a emplear, búsqueda bibliográfica y la planificación de las tareas en base a los plazos de entrega de las PEC.

La parte central del trabajo la constituyó el trabajo bioinformático, que se dividió en tres partes de acuerdo con los objetivos principales: actualización del listado de proteínas, análisis evolutivo y análisis de resultados.

Además, en mitad de este período, marcado por la entrega de la PEC2, se aprovechó para elaborar un primer boceto de la memoria con todos los apartados realizados hasta el momento.

Los principales hitos que marcaron este trabajo fueron:

Cumplimiento del objetivo 1: actualización del listado de proteínas multifuncionales que realizan funciones NOGD.

Cumplimiento del objetivo 2: creación de árboles filogenéticos con proteínas NOGD y multifuncionales que realizan una misma función.

Tabla 1. Planificación del trabajo.

	Fecha de inicio	Fecha de fin
Definición del trabajo	19/02/20	24/02/20
Objetivos y búsqueda bibliográfica	25/02/20	9/03/20
PEC0	2/03/20	2/03/20
Planificación	10/03/20	16/03/20
PEC1	16/03/20	16/03/20
Listado de proteínas	17/03/20	13/04/20
Inicio memoria	14/04/20	20/04/20
Análisis de homología	22/04/20	10/06/20
PEC2	22/04/20	22/04/20
Análisis de resultados	18/05/20	12/06/20
PEC3	1/06/20	1/06/20
Elaboración memoria final	8/06/20	22/06/20
PEC4	24/06/20	24/06/20
Elaboración presentación	22/06/20	26/06/20

Diagrama de Gantt

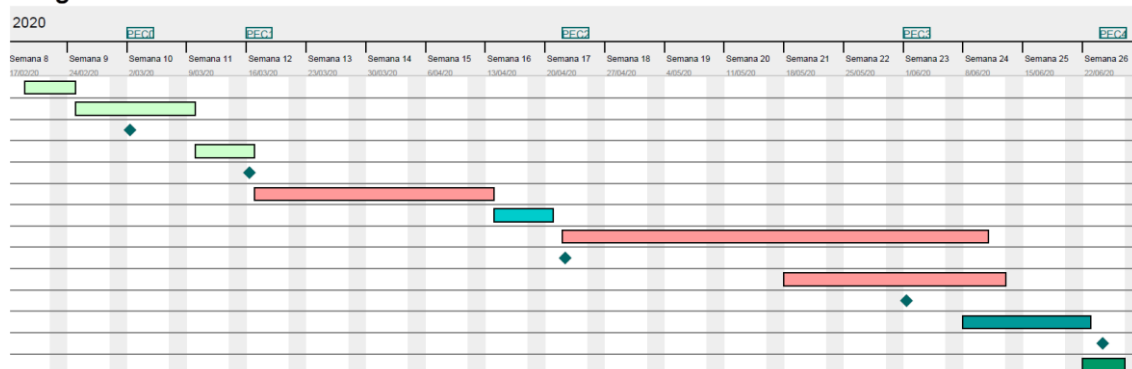


Figura 1. Diagrama de Gantt. Se representan en verde claro los aspectos de planificación, en salmón la realización del trabajo bioinformático, en azul la elaboración de la memoria y en verde la de la presentación final. Se marcan las fechas de entrega de cada PEC con un rombo azul.

1.5. Breve resumen de productos obtenidos

Anexo I. Listado de proteínas multifuncionales y NOGD que realizan la misma función.

Anexo II. Árboles filogenéticos y alineamientos realizados con varios ejemplos de proteínas multifuncionales y NOGD que realizan la misma función.

1.6. Breve descripción de los otros capítulos de la memoria

Apartado 2: Métodos

Se enumerarán las bases de datos y herramientas bioinformáticas empleadas, así como las dificultades que surgieron y como se afrontaron. Se describirá la

estrategia que se siguió para obtener el listado actualizado de proteínas multifuncionales y NOGD y en el análisis evolutivo de las mismas.

Apartado 3: Resultados

Se mostrarán los resultados obtenidos en la actualización del listado de proteínas multifuncionales y NOGD y en el estudio evolutivo entre proteínas que realizan la misma función.

Apartado 4. Discusión.

A partir de los resultados obtenidos, se analizará si el fenómeno NOGD podría ser un mecanismo de obtención de una función *moonlighting* y se relacionará con otras hipótesis acerca del origen y evolución de las proteínas multifuncionales.

Apartado 5: Conclusiones

Se enumerarán las conclusiones obtenidas en este trabajo y las vías de investigación que puedan quedar pendientes, así como una reflexión sobre el cumplimiento de la planificación y los objetivos iniciales.

2. Métodos

2.1. Bases de datos y herramientas bioinformáticas empleadas

MultitaskProtDBII

<http://wallace.uab.es/multitaskII/>

Base de datos creada por el grupo del Dr. Franco-Serrano, actualizada en 2018 [6]. Recopila las proteínas multifuncionales encontradas en la literatura e información relevante acerca de las mismas. Contiene 694 proteínas y sus funciones canónica y *moonlighting* con sus correspondientes códigos GO, así como los códigos UniProt y PDB de la proteína, el organismo en el que es multifuncional y referencias bibliográficas e información sobre enfermedades humanas en las que están involucradas o fármacos de los que son diana.

Analogous enzymes

https://www.ncbi.nlm.nih.gov/Complete_Genomes/AnalEnzymes.html

Base de datos creada por Koonin et al., actualizada en 2010 [11]. Contiene 224 entradas correspondientes a funciones enzimáticas realizadas por al menos dos proteínas estructuralmente diferentes que tienen el mismo nombre, correspondiente a la función que realizan (proteínas NOGD). Se muestra el nombre de la enzima, sus códigos EC, UniProt, PDB y NCBI, clasificación estructural, familia y superfamilia, referencia bibliográfica y el número de proteínas estructuralmente diferentes que realizan esa misma función.

UniProt

<https://www.uniprot.org/>

Base de datos con información funcional y estructural de las proteínas, sus secuencias de aminoácidos, proteínas análogas, taxonomía y bibliografía relevante [12]. Se describe principalmente la función canónica pero en algunos casos también las *moonlighting*. Se utilizó para comprobar la función realizada por las proteínas a analizar y obtener su secuencia de aminoácidos, así como para buscar esa proteína en otras especies.

Clustal Omega

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

Herramienta que realiza alineamientos globales entre al menos 3 secuencias mediante el uso de árboles guía [13]. Estos pueden ser aportados por el usuario o, por defecto, creados a partir de las secuencias introducidas. Es un proceso iterativo en el que se alinean dos secuencias y éstas se alinean con una tercera que se incorporará a ellas o generará un nuevo cluster. Este árbol se usará como guía para el alineamiento múltiple.

Aunque Clustal permite representar los alineamientos y el cladograma con la distancia entre las secuencias, no es la herramienta adecuada para ello en este trabajo, ya que queremos estudiar proteínas muy diferentes tanto en tamaño como en su secuencia y se obtienen mejores resultados con alineamientos locales que globales. Además, los cladogramas obtenidos no son árboles filogenéticos inferidos a partir de las secuencias si no simplemente el árbol guía creado para el alineamiento.

Por tanto, en este trabajo Clustal se utilizó para realizar el alineamiento de las secuencias de aminoácidos de las proteínas NOGD y *moonlighting* que realizan una misma función para conocer el grado de homología global entre estas secuencias.

Protein BLAST

<https://blast.ncbi.nlm.nih.gov/>

Permite realizar alineamientos locales de secuencias de aminoácidos. Generalmente se usa para buscar secuencias similares a una introducida en diferentes bases de datos, pero dispone de múltiples herramientas con otros fines, como el alineamiento de dos o más secuencias entre sí [14].

Se usó para buscar otras proteínas relacionadas con la introducida con el objetivo de establecer su origen y línea evolutiva. Para ello se utilizó el algoritmo PSI-BLAST [15], que permite hacer un análisis iterativo en el que con secuencias seleccionadas de los resultados del BLAST se crean matrices que se usan como base para el siguiente BLAST permitiendo realizar homología remota.

Como se detallará en el apartado 2.3., se utilizó para hacer homología remota e intentar encontrar el origen de la proteína introducida, así como para obtener proteínas relacionadas evolutivamente con las que construir los árboles filogenéticos. Se probaron varias configuraciones según el objetivo de cada búsqueda, modificando principalmente el número máximo de secuencias a alinear. En el resto de aspectos se mantuvo la configuración por defecto, con

un valor E máximo de 10, un tamaño mínimo a alinear de 3 aminoácidos y una matriz de puntuación BLOSUM 62, con penalización por gaps de 11 por su creación y 1 por la extensión.

COBALT

https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi

Herramienta de alineamiento múltiple local que utiliza bases de datos de dominios conservados y motivos para realizar el alineamiento [16]. Se puede lanzar desde los resultados de un BLAST o bien introducir directamente las secuencias que se quieren alinear. Permite construir árboles filogenéticos con las secuencias alineadas mediante los algoritmos *Fast Minimum Evolution* [17] y *Neighbor Joining* [18].

Sin embargo, aunque permite alinear secuencias muy alejadas no se puede construir el árbol filogenético con ellas, la máxima diferencia permitida entre las secuencias del árbol es una fracción de *mismatched* de 0.9. Como el objetivo era representar un árbol filogenético amplio que abarcara muchos filos distantes definitivamente no se utilizó esta herramienta para este fin, únicamente se usó para representar los alineamientos.

NGPhylogeny

<https://ngphylogeny.fr/>

Diseñado para realizar árboles filogenéticos, integra numerosas herramientas de búsqueda de secuencias, alineamiento múltiple e inferencia filogenética [19]. Permite hacer todo el análisis en un sólo paso mediante *workflows* preestablecidos, modificaciones de éstos o creados de cero por el usuario.

Para *datasets* de tamaño pequeño-mediano, como es el caso, se recomienda realizar los alineamientos con el método MAFFT [20], la curación de los mismos mediante BMGE [21] y la inferencia del árbol filogenético con PhyML [22], basado en maximum-likelihood. Este es el análisis integrado en el *workflow* PhyML/OneClick y fue el utilizado.

2.2. Actualización del listado de proteínas multifuncionales y NOGD que realizan la misma función.

Se accedió a las bases de datos de proteínas *moonlighting* y NOGD y se analizó la información contenida en las mismas para establecer cómo afrontar la búsqueda de proteínas presentes en ambas.

Los campos comunes a ambas bases son el nombre de la proteína y los códigos UniProt, PDB y EC. El único de estos campos que define la funcionalidad de las proteínas sin importar su estructura es el código EC, pues éste es exclusivamente funcional y, por tanto, común a proteínas de diferentes organismos que realicen la misma función enzimática. Sin embargo, otros identificadores como los de UniProt o PDB son exclusivos de cada forma estructural y no coincidirán entre proteínas con diferentes estructuras aunque realicen la misma función, por lo que no sirven para nuestro propósito.

Por tanto, la primera estrategia de búsqueda que se usó fue a través de los códigos EC. Sin embargo, surgen dos problemas con este método de búsqueda: no todas las proteínas de la base MultitaskProtDBII los contienen y, de ser así, estos van asociados al nombre de la proteína haciendo referencia a su función canónica pero no al resto de sus funciones. Por lo que se perdería mucha información si nos basamos sólo en los códigos EC. Por ello, fue necesaria otra estrategia de búsqueda. Lamentablemente, no hay otro elemento funcional común a las dos bases de datos que permita hacer una búsqueda automática, ni siquiera los nombres de las proteínas coinciden en todos los casos, porque estos generalmente están asociados a su función canónica pero no a las *moonlighting*, por lo que se procedió a realizar la búsqueda manualmente.

Se hizo a través de los nombres de las proteínas NOGD, que indican esta función enzimática realizada por proteínas de diferentes organismos, y sus “variantes lógicas” para alcanzar el mayor número posible de resultados. Por ejemplo, para la función “Mannosyl-glycoprotein endo- β -N-acetylglucosaminidase” se buscaron los términos “Mannosyl-glycoprotein endo- β -N-acetylglucosaminidase”, “endo-beta-N-acetylglucosaminidase” y “acetylglucosaminidase” en la base MultitaskProtDBII.

De esta forma, se abarcaron todas las proteínas multifuncionales que realizaban estas funciones de manera tanto canónica como *moonlighting* aunque su nombre no fuera el mismo que el de las proteínas NOGD correspondientes.

Se recurrió a UniProt para corroborar que todas las proteínas realizaban la misma función y solventar así dudas debidas a la diferente nomenclatura.

2.3. Estudio evolutivo entre las proteínas que realizan la misma función.

Dadas las diversas relaciones encontradas entre las proteínas multifuncionales y NOGD, se seleccionaron unas cuantas proteínas representativas para realizar estudios evolutivos, tal como se indicará en el apartado 3.2.1.

El primer paso en el análisis fue comprobar que las proteínas *moonlighting* y NOGD encontradas en las bases de datos correspondientes realizaban la misma función. Para ello se recurrió a UniProt, de dónde también se obtuvieron sus secuencias de aminoácidos. Se procedió a realizar el alineamiento en Clustal Omega de estas secuencias para ver el grado de homología global entre las proteínas.

La estrategia de análisis inicial consistía en realizar un estudio de homología remota con PSI-BLAST para establecer la línea evolutiva de cada proteína e intentar encontrar si la proteína multifuncional podía proceder de otra relacionada con alguna de sus otras funciones. Para ello se realizaba PSI-BLAST con la secuencia de la proteína y se buscaba entre los resultados otras proteínas diferentes con un alto porcentaje de homología, mayor del 60%, que se seleccionaban para siguientes BLAST. La intención era ir obteniendo cada vez proteínas más similares a éstas y no a la introducida para ir estableciendo la línea evolutiva y el origen de la misma. Se probó en varias proteínas sin obtener los resultados esperados, incluso aumentando el número máximo de secuencias a encontrar y disminuyendo el % de similitud de las proteínas

seleccionadas para siguientes BLAST hasta un 40%. Estos resultados se detallarán en el apartado 3.2.3.

Se estableció entonces otra estrategia de análisis que consistió en construir árboles filogenéticos con las proteínas *moonlighting* y NOGD a analizar, así como con sus correspondientes en otras especies para ver la relación evolutiva entre todas ellas. Estas otras especies se seleccionaron estratégicamente para abarcar distintos puntos evolutivos. En la Figura 2 se representa el abanico de filos que se abarcó cuando fue posible, pues no en todos los casos se encontraron estas proteínas en todas las especies deseadas.

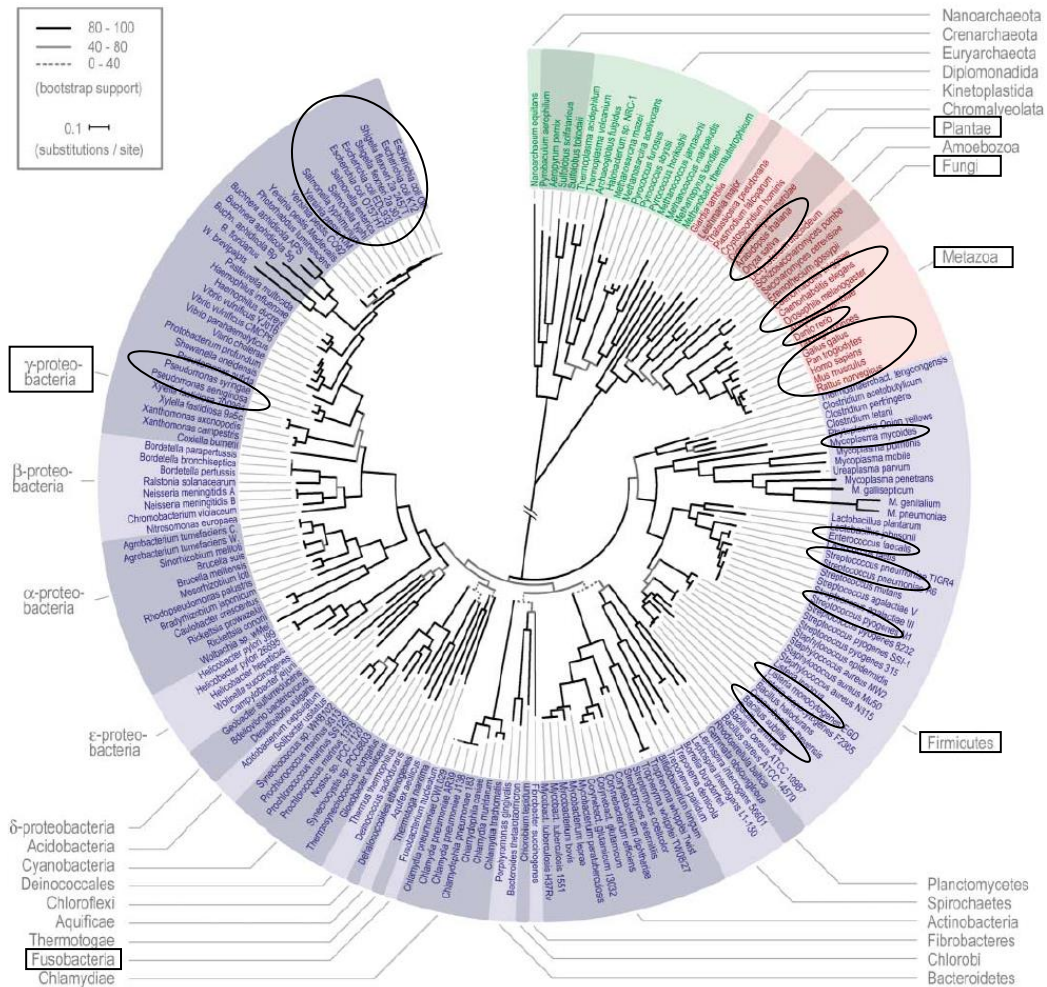


Figura 2. Filos abarcados en la construcción de árboles filogenéticos. Modificado de Ciccarelli. et al. [23]. Se remarcan los filos y principales especies que se incluyeron en la construcción de árboles filogenéticos cuando fue posible, abarcando así especies ampliamente distribuidas a lo largo de la evolución. Principalmente se incluyeron las especies *Escherichia coli*, *Shigella flexneri*, *Salmonella enterica*, *Elizabethkingia meningoseptica*, *Streptococcus pneumoniae*, *Bacillus subtilis*, *Listeria monocytogenes*, *Enterococcus faecalis*, *Arabidopsis thaliana*, *Oryza sativa*, *Danio rerio*, *Gallus gallus*, *Pan troglodytes*, *Rattus norvegicus*, *Mus musculus* y *Homo sapiens*.

Se obtuvieron las secuencias proteicas de las especies de interés a partir de los resultados de PSI-BLAST de las proteínas *moonlighting* y NOGD y de otras más distantes que se buscaron en UniProt. Con todas estas secuencias se realizaron alineamientos en COBALT y se construyeron árboles filogenéticos con NGPhylogeny, como se ha indicado en el apartado 2.1.

3. Resultados

3.1. Actualización del listado de proteínas multifuncionales y NOGD que realizan la misma función.

Con la primera búsqueda a través de los códigos EC, se encontraron 14 códigos comunes a ambas bases de datos, que se correspondían con 16 proteínas multifuncionales y 34 proteínas NOGD.

Con la segunda estrategia, se amplió este resultado hasta 46 funciones enzimáticas de proteínas NOGD que eran realizadas por proteínas *moonlighting*. Realizan estas funciones un total de 124 proteínas multifuncionales y 122 proteínas NOGD, de las cuales 8 proteínas son a la vez multifuncionales y NOGD. El número de proteínas que realizan una misma función difiere según cuál sea la función en cuestión. Desde los casos más sencillos en los que hay dos proteínas NOGD siendo una de ellas multifuncional, hasta los más complejos en los que se encuentran entre 8 y 11 proteínas diferentes que realizan la misma función.

Proporcionalmente, se obtuvo que un 20,5% de las funciones enzimáticas de la base Analogous Enzymes están realizadas por proteínas multifuncionales, lo que supone una relación funcional entre el 24% de proteínas NOGD y el 18% de proteínas multifuncionales de sus bases de datos respectivas. Estos porcentajes son lo suficientemente altos para sugerir la relación entre ambos fenómenos.

Así mismo, se encontraron todo tipo de relaciones. Por ejemplo, se vió que proteínas multifuncionales totalmente distintas podían realizar una misma función coincidente con la de una proteína NOGD, como es el caso de la fosfoproteína fosfatasa, cuya función también es realizada por las proteínas multifuncionales *pyruvate*, *orthophosphatedikinase regulatory protein (PPDK-RP)*, *stage II sporulation protein E*, *serine/threonine-protein phosphatase PP1-gamma catalytic subunit* y *developmental protein eyes absent*. También se observó que la misma proteína multifuncional podía realizar varias funciones de proteínas NOGD, como por ejemplo, la peroxirredoxina-6 humana, con funciones peroxirredoxina, fosfolipasa A2 y lisofosfolipasa.

Un hallazgo curioso es que hay casos en los que hay dos proteínas NOGD de la misma especie. Concretamente, ocurre en 12 de las 47 funciones encontradas. Por la propia definición del fenómeno NOGD, estas proteínas surgen cuando un componente esperado en una vía enzimática está ausente y su función es realizada por otra proteína. Por tanto, que se encuentren dos proteínas NOGD en el mismo organismo me resultó contradictorio con esta definición y me pareció necesario comentarlo aunque no sea el objetivo de este trabajo revisar las proteínas NOGD si no su relación con las proteínas multifuncionales.

En la tabla 2 se muestra un fragmento representativo de los resultados obtenidos, el listado completo se puede consultar en el anexo I. Se muestra la función que relaciona estas proteínas, su nombre tal como aparece en las bases de datos correspondientes, el organismo en el que se ha visto que

realizan dicha función, su código UniProt y si son *moonlighting*, NOGD o ambas.

Tabla 2. Ejemplos de proteínas *moonlighting* y NOGD que realizan la misma función.

Función	Proteína	Organismo	UniProt	Moon/NOGD
Alpha-N-acetylgalactosaminidase EC. 3.2.1.49	Alpha-N-acetylgalactosaminidase	<i>Homo sapiens</i>	P17050	Moonlighting + NOGD
	Alpha-N-acetylgalactosaminidase	<i>Elizabethkingia meningoseptica</i>	A4Q8F7	NOGD
Acetaldehyde dehydrogenase EC. 1.2.1.10	Alcohol dehydrogenase (EhADH2)	<i>Entamoeba histolytica</i>	Q24803	Moonlighting + NOGD
	Alcohol acetaldehyde dehydrogenase	<i>Listeria monocytogenes</i>	Q6Q3I2	Moonlighting
	Acetaldehyde dehydrogenase (acetylating)	<i>Escherichia coli</i>	P77580	NOGD
	Acetaldehyde dehydrogenase (acetylating)	<i>Pseudomonas putida</i>	Q49KG0	NOGD
Glutamate dehydrogenase EC. 1.4.1.2	Glutamate dehydrogenase	<i>Bacillus subtilis</i>	P39633	Moonlighting + NOGD
	Glutamate dehydrogenase 1, mitochondrial	<i>Gallus gallus</i>	P00368	Moonlighting
	Isocitrate--homocitrate dehydrogenase	<i>Pyrococcus horikoshii</i>	O59394	Moonlighting
	Glutamate dehydrogenase	<i>Achlya klebsiana</i>	P41755	NOGD
Glucose-6-phosphate isomerase EC. 5.3.1.9	Phosphoglucose isomerase	<i>Homo sapiens</i>	P06744	Moonlighting + NOGD
	Phosphoglucose isomerase	<i>Geobacillus stearothermophilus</i>	P13376	Moonlighting
	Phosphoglucose isomerase	<i>Xanthomonas oryzae</i>	B2SJM1	Moonlighting
	Glucose-6-phosphate isomerase	<i>Echinococcus multilocularis</i>	A7LKP5	Moonlighting
	Phosphoglucose isomerase	<i>Lactobacillus crispatus</i>	D5H2G7	Moonlighting
	Glucose-6-phosphate isomerase	<i>Mus musculus</i>	P06745	Moonlighting
	Glucose-6-phosphate isomerase	<i>Mustela nigripes</i>	M3Y807	Moonlighting
	Bifunctional phosphoglucose/phosphomannose isomerase	<i>Pyrobaculum aerophilum</i>	Q8ZWV0	Moonlighting
	Bifunctional phosphoglucose/phosphomannose isomerase	<i>Aeropyrum pernix</i>	Q9YE01	Moonlighting
	Bifunctional phosphoglucose/phosphomannose isomerase	<i>Thermoplasma acidophilum</i>	Q9HIC2	Moonlighting
Glucose-6-phosphate isomerase	<i>Thermococcus litoralis</i>	P84140	NOGD	
tRNA (guanine-N1-)-methyltransferase EC. 2.1.1.31	Mitochondrial RNase P	<i>Homo sapiens</i>	Q7L0Y3	Moonlighting
	tRNA (guanine-N1-)-methyltransferase	<i>Homo sapiens</i>	Q32P41	NOGD
	tRNA (guanine-N1-)-methyltransferase	<i>Escherichia coli</i>	P0A873	NOGD
3',5'-cyclic-AMP phosphodiesterase EC. 3.1.4.53	3',5'-cyclic AMP phosphodiesterase CpdA	<i>Mycobacterium tuberculosis</i>	P9WP65	Moonlighting
	3',5'-cyclic-AMP phosphodiesterase	<i>Dictyostelium discoideum</i>	Q23917	NOGD
	3',5'-cyclic-AMP phosphodiesterase	<i>Dictyostelium discoideum</i>	P12019	NOGD
Fumarate hydratase EC:4.2.1.2	Fumarate hydratase	<i>Homo sapiens</i>	P07954	Moonlighting
	Fumarase	<i>Paracoccidioides brasiliensis</i>	C1GPZ3	Moonlighting
	Fumarate hydratase	<i>Escherichia coli</i>	P05042	NOGD
	Fumarate hydratase	<i>Escherichia coli</i>	P0AC33	NOGD
Amidase EC. 3.5.1.4	Amidase	<i>Listeria monocytogenes</i>	O33983	Moonlighting
	Putative endo-beta-N-acetylglucosaminidase	<i>Streptococcus pneumoniae</i>	P59206	Moonlighting

	Amidase	<i>Pseudomonas chlororaphis</i>	P27765	NOGD
	Amidase	<i>Pseudomonas aeruginosa</i>	P11436	NOGD
Testosterone 17- β -dehydrogenase EC. 1.1.1.63 Testosterone 17- β -dehydrogenase (NADP+) EC. 1.1.1.64	3-hydroxyacyl-CoA dehydrogenase type-2	<i>Homo sapiens</i>	Q99714	Moonlighting
	Testosterone 17- β -dehydrogenase	<i>Homo sapiens</i>	O14756	NOGD
	Testosterone 17- β -dehydrogenase	<i>Homo sapiens</i>	P37058	NOGD
	Testosterone 17- β -dehydrogenase	<i>Homo sapiens</i>	P42330	NOGD
Peptidylprolyl isomerase EC. 5.2.1.8	Peptidyl Prolyl cis,trans-Isomerase (PPI)	<i>Helicobacter pylori</i>	D0K1B1	Moonlighting
	Peptidyl-prolyl cis-trans isomerase A	<i>Homo sapiens</i>	P62937	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Chlamydia trachomatis</i>	P26623	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Histoplasma capsulatum</i>	A6R3K7	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Neisseria gonorrhoeae</i>	P0A0W3	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Rickettsia prowazekii</i>	Q9ZCX6	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Streptococcus pneumoniae</i>	J1NX65	Moonlighting
	Peptidylprolyl isomerase	<i>Escherichia coli</i>	P0AFL3	NOGD
	Peptidylprolyl isomerase	<i>Escherichia coli</i>	P45523	NOGD
Peroxidase EC. 1.11.1.7	Cyclooxygenase-1	<i>Homo sapiens</i>	P23219	Moonlighting
	Prostaglandin G/H synthase 1	<i>Ovis aries</i>	P23219	Moonlighting
	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial	<i>Homo sapiens</i>	P36969	Moonlighting
	Peroxiredoxin	<i>Aeropyrum pernix</i>	Q9Y9L0	NOGD
	Peroxiredoxin	<i>Mycobacterium tuberculosis</i>	P9WQB5	NOGD
	Peroxiredoxin	<i>Escherichia coli</i>	P0C0L2	NOGD
	Peroxidase	<i>Mus musculus</i>	O08709	NOGD
	Peroxidase	<i>Homo sapiens</i>	P05164	NOGD
Peroxidase	<i>Escherichia coli</i>	P31545	NOGD	
Peroxiredoxin EC. 1.11.1.15	Peroxiredoxin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Tsa1p (Peroxiredoxin TSA1)	<i>Candida albicans</i>	Q9Y7F0	Moonlighting
	cPrxl (Peroxiredoxin TSA1)	<i>Saccharomyces cerevisiae</i>	P34760	Moonlighting
	Peroxiredoxin TSA2	<i>Saccharomyces cerevisiae</i>	Q04120	Moonlighting
	Peroxiredoxin	<i>Aeropyrum pernix</i>	Q9Y9L0	NOGD
	Peroxiredoxin	<i>Mycobacterium tuberculosis</i>	P9WQB5	NOGD
	Peroxiredoxin	<i>Escherichia coli</i>	P0C0L2	NOGD
	Peroxidase	<i>Mus musculus</i>	O08709	NOGD
Phospholipase A2 EC. 3.1.1.4	Peroxiredoxin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Cytosolic phospholipase A2	<i>Homo sapiens</i>	P47712	NOGD
	Group IIE secretory phospholipase A2	<i>Homo sapiens</i>	Q9NZK7	NOGD
	Phospholipase A2	<i>Escherichia coli</i>	P0A921	NOGD
Lysophospholipase EC. 3.1.1.5	Peroxiredoxin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Lysophospholipase	<i>Homo sapiens</i>	P47712	NOGD
	Lysophospholipase	<i>Rattus norvegicus</i>	O88202	NOGD
	Lysophospholipase	<i>Escherichia coli</i>	P07000	NOGD

3.2. Estudio evolutivo entre las proteínas que realizan la misma función.

3.2.1. Selección de proteínas para el análisis evolutivo.

Dada la cantidad de proteínas encontradas y las diversas relaciones observadas entre ellas, se seleccionaron unas cuantas proteínas que resultaron más llamativas para realizar los estudios evolutivos.

Una relación particularmente interesante son aquellos casos en los que hay dos proteínas NOGD y una de ellas es multifuncional. Este simple hallazgo ya sugiere la estrecha relación entre los dos fenómenos. Por ello se analizaron varios casos de este tipo.

En otros casos, las proteínas NOGD y multifuncional(es) no coinciden pero se presentan en especies cercanas o incluso son diferentes proteínas de la misma especie. Esto podría indicar una relación evolutiva entre ellas, por lo que también se seleccionaron ejemplos de este tipo.

Otras relaciones son más complejas por haber numerosas proteínas de muy diversos organismos realizando la misma función e incluso proteínas multifuncionales que realizan varias funciones de proteínas NOGD, por lo que es más complicado realizar un estudio evolutivo con ellas y sólo se seleccionó un ejemplo de este tipo.

Tabla 3. Proteínas seleccionadas para el análisis evolutivo.

Función	Proteína	Organismo	UniProt	Moon/NOGD
6,7-dihydropteridine reductase EC. 1.5.1.34	Dihydropteridine reductase	<i>Homo sapiens</i>	P09417	Moonlighting + NOGD
	6,7-dihydropteridine reductase	<i>Escherichia coli</i>	P38489	NOGD
Alpha-N-acetylgalactosaminidase EC. 3.2.1.49	Alpha-N-acetylgalactosaminidase	<i>Homo sapiens</i>	P17050	Moonlighting + NOGD
	Alpha-N-acetylgalactosaminidase	<i>Elizabethkingia meningoseptica</i>	A4Q8F7	NOGD
Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase EC. 3.2.1.96	Putative endo-beta-N-acetylglucosaminidase	<i>Streptococcus pneumoniae</i>	P59206	Moonlighting + NOGD
	Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	<i>Elizabethkingia meningoseptica</i>	P36912	NOGD
tRNA (guanine-N1-)-methyltransferase EC. 2.1.1.31	Mitochondrial RNase P	<i>Homo sapiens</i>	Q7L0Y3	Moonlighting
	tRNA (guanine-N1-)-methyltransferase	<i>Homo sapiens</i>	Q32P41	NOGD
	tRNA (guanine-N1-)-methyltransferase	<i>Escherichia coli</i>	P0A873	NOGD
Nitric-oxide synthase EC. 1.14.13.39	Endothelial Nitric Oxide Synthase	<i>Homo sapiens</i>	P29474	Moonlighting
	Nitric-oxide synthase	<i>Rattus norvegicus</i>	P29476	NOGD
	Nitric-oxide synthase	<i>Arabidopsis thaliana</i>	Q66GP9	NOGD
Chorismate mutase EC. 5.4.99.5	Chorismate mutase	<i>Xanthomonas oryzae</i>	Q5H4Y2	Moonlighting
	Chorismate mutase	<i>Enterobacter agglomerans</i>	P42517	NOGD
Phospholipase A2 EC. 3.1.1.4	Peroxisome oxidoreductin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Cytosolic phospholipase A2	<i>Homo sapiens</i>	P47712	NOGD
	Group IIE secretory phospholipase A2	<i>Homo sapiens</i>	Q9NZK7	NOGD
	Phospholipase A2	<i>Escherichia coli</i>	P0A921	NOGD
Lysophospholipase EC. 3.1.1.5	Peroxisome oxidoreductin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Lysophospholipase	<i>Homo sapiens</i>	P47712	NOGD
	Lysophospholipase	<i>Rattus norvegicus</i>	O88202	NOGD
	Lysophospholipase	<i>Escherichia coli</i>	P07000	NOGD

3.2.2. Alineamiento global.

En la tabla 4 se muestran los resultados obtenidos en el alineamiento global en Clustal de las secuencias de aminoácidos de las proteínas multifuncionales y NOGD seleccionadas que realizan la misma función.

Se observa que los porcentajes de identidad son muy bajos, lo que indica la gran diferencia entre estas proteínas que realizan la misma función.

Tabla 4. Matrices de identidad obtenidas en Clustal.

Proteína	Organismo	Matriz de identidad (Clustal)		
DHPR (peptide C) dihydropteridine reductase	<i>Homo sapiens</i>			
6.7-dihydropteridine reductase	<i>Escherichia coli</i>	24.52		
Alpha-N-acetylgalactosaminidase	<i>Homo sapiens</i>			
Alpha-N-acetylgalactosaminidase	<i>Elizabethkingia meningoseptica</i>	17.43		
Putative endo-beta-N-acetylglucosaminidase	<i>Streptococcus pneumoniae</i>			
Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	<i>Elizabethkingia meningoseptica</i>	19.68		
Mitochondrial RNase P	<i>Homo sapiens</i>			
tRNA (guanine-N1-)-methyltransferase	<i>Homo sapiens</i>	19.42		
tRNA (guanine-N1-)-methyltransferase	<i>Escherichia coli</i>	13.43	22.03	
Endothelial Nitric Oxide Synthase	<i>Homo sapiens</i>			
Nitric-oxide synthase	<i>Rattus norvegicus</i>	57.56		
Nitric-oxide synthase	<i>Arabidopsis thaliana</i>	22.33	23.72	
Chorismate mutase	<i>Xanthomonas oryzae</i>			
Chorismate mutase	<i>Enterobacter agglomerans</i>	19.30		
Peroxiredoxin-6	<i>Homo sapiens</i>			
Cytosolic phospholipase A2	<i>Homo sapiens</i>	16.93		
Group IIE secretory phospholipase A2	<i>Homo sapiens</i>	16.13	12.69	
Phospholipase A2	<i>Escherichia coli</i>	9.78	16.94	14.29
Peroxiredoxin-6	<i>Homo sapiens</i>			
Lysophospholipase	<i>Homo sapiens</i>	17.65		
Lysophospholipase	<i>Rattus norvegicus</i>	14.41	23.47	
Lysophospholipase	<i>Escherichia coli</i>	11.54	16.89	20.14

3.2.3. PSI-BLAST.

Como se ha explicado en el apartado 2.3., la estrategia de análisis inicial consistía en realizar un estudio de homología remota con PSI-BLAST para establecer la línea evolutiva de cada proteína e intentar encontrar si la proteína multifuncional podía proceder de otra relacionada con alguna de sus otras funciones. Para ello se realizaba PSI-BLAST con la secuencia de la proteína y se buscaba entre los resultados otras proteínas diferentes con un alto porcentaje de homología que se seleccionaban para siguientes BLAST, con la intención de ir obteniendo cada vez proteínas similares a éstas y no a la introducida para ir estableciendo la línea evolutiva y el origen de la misma. A continuación se detallan los resultados obtenidos con cada proteína analizada, agrupadas según la función que realizan.

6,7-dihidropteridina reductasa

Se encontró que realizaban esta función la proteína dihidropteridina reductasa (DHPR) de *Homo sapiens*, que es a la vez NOGD y *moonlighting*, y la proteína NOGD dihidropteridina reductasa de *Escherichia coli*.

En el PSI-BLAST de la dihidropteridina reductasa humana se obtienen DHPR y DHPR quinoide (QDPR) de distintas especies eucariotas con alto porcentaje de similitud. Las únicas proteínas bacterianas que aparecen son *SDR oxidorreductase/short chain dehydrogenase* con un 50% de identidad. Otras proteínas que se obtienen y que selecciono para la segunda iteración son: *Claudin-16* [Platysternon megacephalum], con un 89% de similitud, y *PTCD3-like protein* [Trichinella] y *exonuclease 3'-5' domain contain protein* [Cassostrea virginica], con alrededor del 50% de similitud. En la segunda iteración, se obtienen nuevas DHPR/QDPR pero no otras proteínas para continuar con la homología remota.

En el PSI-BLAST de la dihidropteridina reductasa de *E.coli* se encuentran nitrorreductasas de diferentes proteobacterias (enterobacterias). Otras proteínas encontradas son *Glutamate-cysteina ligasa* [Salmonella], con un 82.47% de identidad, y *GGPP synthase chloroplastic* [Tanacetum cinerariifolium] con un 52%. En la siguiente iteración, no hay nuevas proteínas con las que realizar más iteraciones para establecer el origen de esta proteína.

α -N-acetilgalactosaminidasa

Realizan esta función la proteína α -N-acetilgalactosaminidasa (NAGA) de *Homo sapiens*, que es a la vez NOGD y *moonlighting*, y la proteína NOGD α -N-acetilgalactosaminidasa de *Elizabethkingia meningoseptica*.

En el PSI-BLAST de la α -N-acetilgalactosaminidasa humana se obtienen diferentes alpha-N-acetylglactosaminidase eucariotas. Se selecciona para la segunda búsqueda la única proteína diferente encontrada: *Transmembrane protein 45B* [Platysternon megacephalum], con 74.84% de similitud. Todas las nuevas proteínas que se añaden en la segunda iteración son NAGA por lo que no se puede continuar con la homología remota.

En el PSI-BLAST de la α -N-acetilgalactosaminidasa de *Elizabethkingia meningoseptica* sólo aparece la familia de oxidoreductasas Gfo/ldh/MocA sin ninguna proteína diferente para hacer una segunda iteración.

Endo- β -N-acetilglucosaminidasa

Tienen esta función la endo- β -N-acetilglucosaminidasa putativa de *Streptococcus pneumoniae*, que es a la vez NOGD y *moonlighting*, y la proteína NOGD manosil-glicoproteína endo- β -N-acetilglucosaminidasa de *Elizabethkingia meningoseptica*.

En el PSI-BLAST de la endo- β -N-acetilglucosaminidasa putativa de *Streptococcus pneumoniae* se obtienen *putative endo-beta-N-Ac-glucosaminidase*, *mannosyl-glycoprotein endo-beta-N-Ac-glucosaminidase* y *glucosaminidase domain containing protein* de *S.pneumoniae* con más del 90% de identidad, así como otras proteínas que se seleccionan para la segunda

iteración: 1´4-beta-N-acetyl muramidase, glucosaminidase domain protein, N-acetylmuramoyl-L-alanine amidase family protein y cell wall binding repeat family protein de *Streptococcus pneumoniae*, todas con identidad mayor del 90%. En las siguientes iteraciones aparecen más acetilglucosaminidasas de *S.pneumoniae* pero ninguna de las otras proteínas para continuar.

En el PSI-BLAST de la manosil-glicoproteína endo-β-N-acetilglucosaminidasa de *Elizabethkingia meningoseptica* sólo se obtienen endo-β-N-acetilglucosaminidasas de distintas especies de *Elizabethkingia*, *Chryseobacterium* o *Pedobacter* por encima del 60% Id, por lo que no se continúa con la homología remota.

ARNt (guanine-N1-)-metiltransferasa

Tienen esta función las proteínas NOGD tRNA (guanine-N1-)-metiltransferasa de *Homo sapiens* y *Escherichia coli* y la proteína multifuncional RNasa P mitocondrial de *Homo sapiens*.

En el PSI-BLAST de la ARNt(guanine-N1-)-metiltransferasa de *E.coli* sólo se obtuvieron resultados correspondientes a esta proteína en diferentes enterobacterias con más del 85% de similitud.

En el de ARNt(guanine-N1-)-metiltransferasa (TRM5) de *Homo sapiens* se obtuvieron diferentes isoformas de esta proteína en especies eucariotas.

Con la RNasa P mitocondrial (TM10C) de *Homo sapiens* tampoco se obtuvieron otras proteínas que no fueran “homólogos 10C de tRNA (guanine-N1-)-metiltransferasa” de distintas especies eucariotas.

Por tanto, con ninguna de ellas se pudieron seleccionar proteínas para hacer homología remota.

Óxido nítrico sintasa (NOS)

Hay varias isoformas de esta proteína: inducibles, constitutivas (endotelial o cerebral) y bacterianas. Se encontró la proteína multifuncional óxido nítrico sintasa endotelial de *Homo sapiens* y las proteínas NOGD óxido nítrico sintasa cerebral de *Rattus norvegicus* y óxido nítrico sintasa mitocondrial/cloroplástica de *Arabidopsis thaliana*.

En el PSI-BLAST de la óxido nítrico sintasa endotelial de *Homo sapiens* se obtienen diferentes isoformas de NOS endoteliales de diferentes especies, incluidas las de *Rattus norvegicus* con un 95% de homología. Con menos del 60% de homología aparecen NOS cerebrales. No aparecen otras proteínas con las que hacer una segunda iteración.

En el PSI-BLAST de la óxido nítrico sintasa cerebral de *Rattus norvegicus* se obtienen diferentes isoformas de NOS cerebrales de diferentes especies, incluidas las de *Homo sapiens* con un 95% de homología, pero no aparecen NOS endoteliales. No aparecen otras proteínas con las que hacer una segunda iteración.

En el PSI-BLAST de la óxido nítrico sintasa mitocondrial/cloroplástica de *Arabidopsis thaliana* se obtienen óxido nítrico sintasa, proteína mitocondrial asociada a óxido nítrico, GTPasa y NTP hidrolasa de plantas y bacterias. Estas

últimas se seleccionan para siguientes iteraciones pero no se obtienen nuevas proteínas para seguir con la homología remota.

Corismato mutasa

Realizan esta función la corismato mutasa multifuncional de *Xanthomonas oryzae* y las NOGD de *Bacillus subtilis* y *Enterobacter agglomerans*. En sus respectivos PSI-BLAST no se obtienen otras proteínas que no sean corismato mutasas para hacer homología remota.

Fosfolipasa A2

Se han encontrado las NOGD fosfolipasa A1 de *E.coli* (con funciones fosfolipasa A1 y A2) y dos formas de fosfolipasa A2 humana (secretada y citosólica). Esta función también es realizada por la proteína multifuncional peroxirredoxina-6 humana. En sus respectivos PSI-BLAST se obtuvo lo siguiente:

Fosfolipasa A2 citosólica de *Homo Sapiens*: la única proteína diferente que se obtiene para siguientes iteraciones es *XK-related protein 8* con un 84% de homología, pero en los siguientes resultados sólo se obtienen nuevas fosfolipasas A2.

Fosfolipasa A2 secretada de *Homo Sapiens*: todos los resultados observados fueron fosfolipasas secretadas del grupo IIE. Con un porcentaje de identidad en torno al 40% aparecen fosfolipasas dependientes de calcio y fosfolipasas secretadas del grupo IIF, pero ninguna diferente para seleccionar para sucesivos BLAST.

Fosfolipasa A1 de *E. coli*: se obtienen fosfolipasas de *E.coli* y otras enterobacterias, así como la ATP dependiente–DNA helicasa de *E.coli* que se selecciona para la segunda iteración donde sólo aparecen más fosfolipasas de *E.coli* y, por tanto, no se puede continuar con la homología remota.

Peroxirredoxina-6 de *Homo sapiens*: se obtienen diferentes peroxirredoxinas-6 y otras proteínas que se seleccionan para siguientes iteraciones sin obtener resultados satisfactorios en la tercera iteración.

Ante estos resultados, se decidió cambiar la estrategia de análisis y no utilizar PSI-BLAST para hacer homología remota si no para obtener proteínas con las que construir árboles filogenéticos como se ha detallado en el apartado 2.3.

3.2.4. Árboles filogenéticos y alineamientos múltiples.

6,7-dihidropteridina reductasa

Se encontró que realizaban esta función la proteína dihidropteridina reductasa de *Homo sapiens*, que es a la vez NOGD y *moonlighting*, y la proteína NOGD dihidropteridina reductasa de *Escherichia coli*. Al analizar estas proteínas mediante PSI-BLAST, se encontraron nitrorreductasas bacterianas y dihidropteridina reductasa de especies superiores. En UniProt no se encuentra esta nitrorreductasa en *Homo sapiens* ni la dihidropteridina reductasa en

Escherichia coli, lo que corrobora los resultados de PSI-BLAST y sugiere un “hueco” evolutivo entre estas proteínas.

En la representación filogenética (figura 3) se observan estas proteínas en dos líneas separadas, correspondientes a las nitrorreductasas bacterianas y a las dihidropteridina reductasas respectivamente.

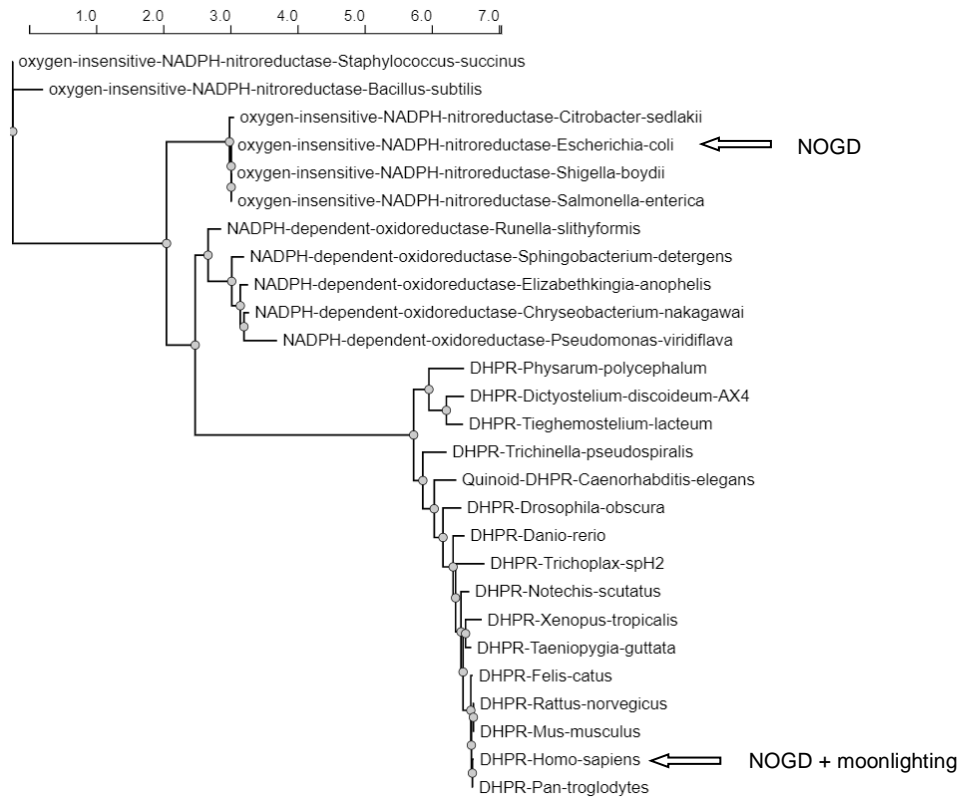


Figura 3. Árbol filogenético de proteínas con función 6,7-dihidropteridina reductasa. Se señalan las proteínas NOGD y moonlighting correspondientes.

En la figura 4 se muestra el alineamiento de estas proteínas en COBALT. También se observan aquí dos grupos bien diferenciados, constituidos por proteínas similares entre sí, y con regiones conservadas a lo largo de ambos grupos (en azul), probablemente responsables de realizar la función que las relaciona.

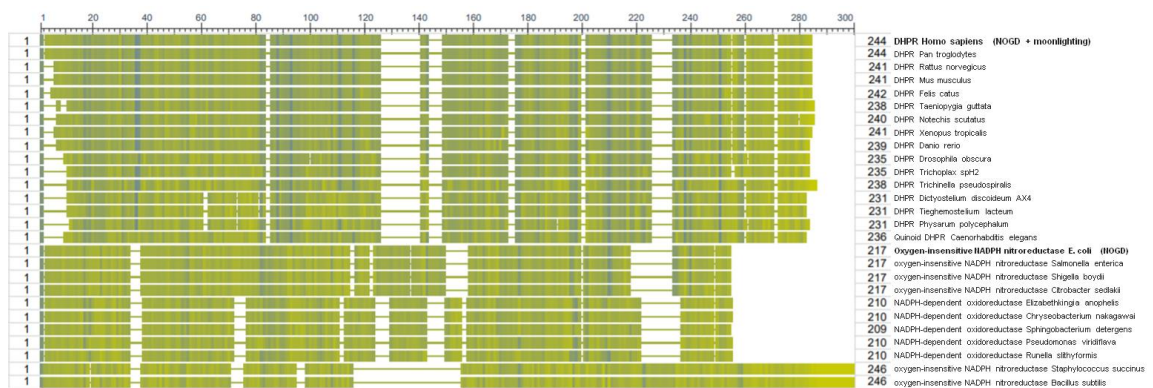


Figura 4. Alineamiento de proteínas con función 6,7-dihidropteridina reductasa. Realizado con COBALT y coloreado con BLOSUM45. Se representa una vista global del alineamiento entre

las proteínas a través del coloreado, sin mostrar las secuencias por simplicidad. Se colorean en verde los aminoácidos no compartidos y en azul los conservados.

Óxido nítrico sintasa

Se encontraron la proteína multifuncional óxido nítrico sintasa endotelial de *Homo sapiens* y las proteínas NOGD óxido nítrico sintasa cerebral de *Rattus norvegicus* y óxido nítrico sintasa mitocondrial/cloroplástica de *Arabidopsis thaliana*. Se añadieron óxido nítrico sintasas de especies relacionadas con éstas y de especies bacterianas.

Se obtuvo un árbol filogenético (figura 5) en el que se observa la relación entre las óxido nítrico sintasas endotelial (*moonlighting* en humanos), cerebral y bacterianas. Las óxido nítrico sintasa de plantas constituyen una línea evolutiva diferente. Por tanto, se observa una relación evolutiva entre una de las formas estructurales de las proteínas NOGD y la proteína multifuncional.

En el alineamiento (figura 6) se obtiene que las óxido nítrico sintasa endotelial y cerebral son muy similares. También se observa que conservan muchos aminoácidos con respecto a las bacterianas, a pesar de ser éstas mucho más cortas, lo que no ocurre con las de plantas, que son muy diferentes.

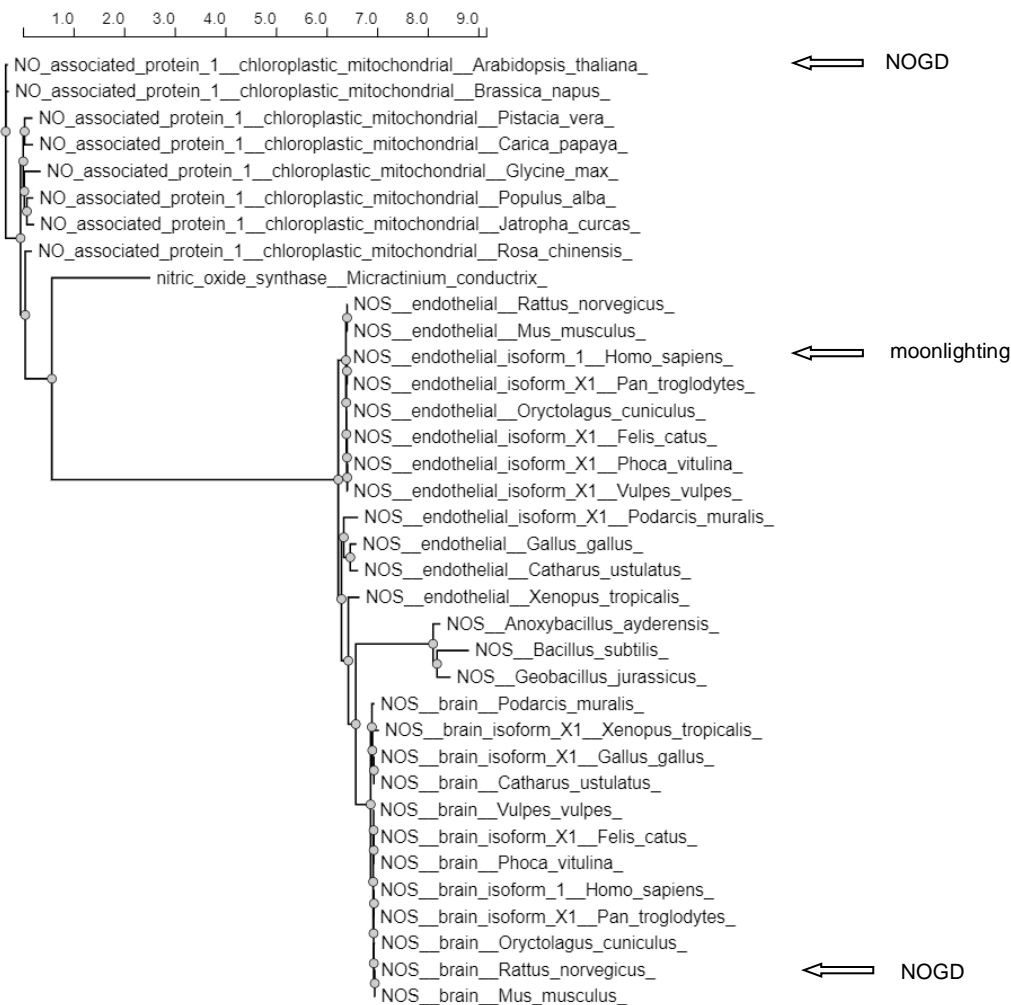


Figura 6. Árbol filogenético de proteínas con función óxido nítrico sintasa. Se señalan las proteínas NOGD y *moonlighting* correspondientes.



Figura 5. Alineamiento de proteínas con función óxido nítrico sintasa. Realizado con COBALT y coloreado con BLOSUM45. Se representa una vista global del alineamiento entre las proteínas a través del coloreado, sin mostrar las secuencias, por simplicidad. Se colorean en verde los aminoácidos no compartidos y en azul los conservados.

Corismato mutasa

Realizan esta función la corismato mutasa multifuncional de *Xanthomonas oryzae* y las NOGD de *Bacillus subtilis* y *Enterobacter agglomerans* (*Pantoea agglomerans*).

Esta enzima existe en bacterias, plantas, algas y algunos hongos pero no en protozoos ni animales, pues participa en la ruta del ácido shikímico para la síntesis de aminoácidos aromáticos, que no se da en animales. Se conocen dos proteínas participantes en esta ruta: una bifuncional, que cataliza dos reacciones enzimáticas (el paso de corismato a prefenato, que es la actividad corismato mutasa, y la conversión de prefenato a fenilpiruvato) y que está presente por ejemplo en *E.coli*, y otra monofuncional, que sólo cataliza la conversión de corismato a prefenato como es, por ejemplo, la de *Bacillus subtilis*.

Además de las proteínas NOGD y multifuncional encontradas, se incluyó también en el estudio esta proteína bifuncional de *E.coli* y otras cercanas a ellas. En el PSI-BLAST de estas proteínas se observa que muchas especies de enterobacterias tienen ambas enzimas, la bifuncional y la monofuncional.

En el árbol filogenético (figura 7) se observa una relación evolutiva entre las corismato mutasas bifuncionales, la corismato mutasa multifuncional y la corismato mutasa de *Pantoea agglomerans*, mientras que la de *Bacillus subtilis* está relacionada con la de otras bacterias.

En la figura 8 se muestra el alineamiento entre estas proteínas. Se observa que la corismato mutasa multifuncional de *Xanthomonas oryzae* y la de *Pantoea agglomerans* son muy similares y, a su vez, son también muy similares a la primera mitad de las enzimas bifuncionales, probablemente esta sea la zona correspondiente a la actividad corismato mutasa. El resto de corismato mutasas de bacterias y plantas forman dos grupos diferenciados pero también conservan algunos aminoácidos comunes con las primeras.

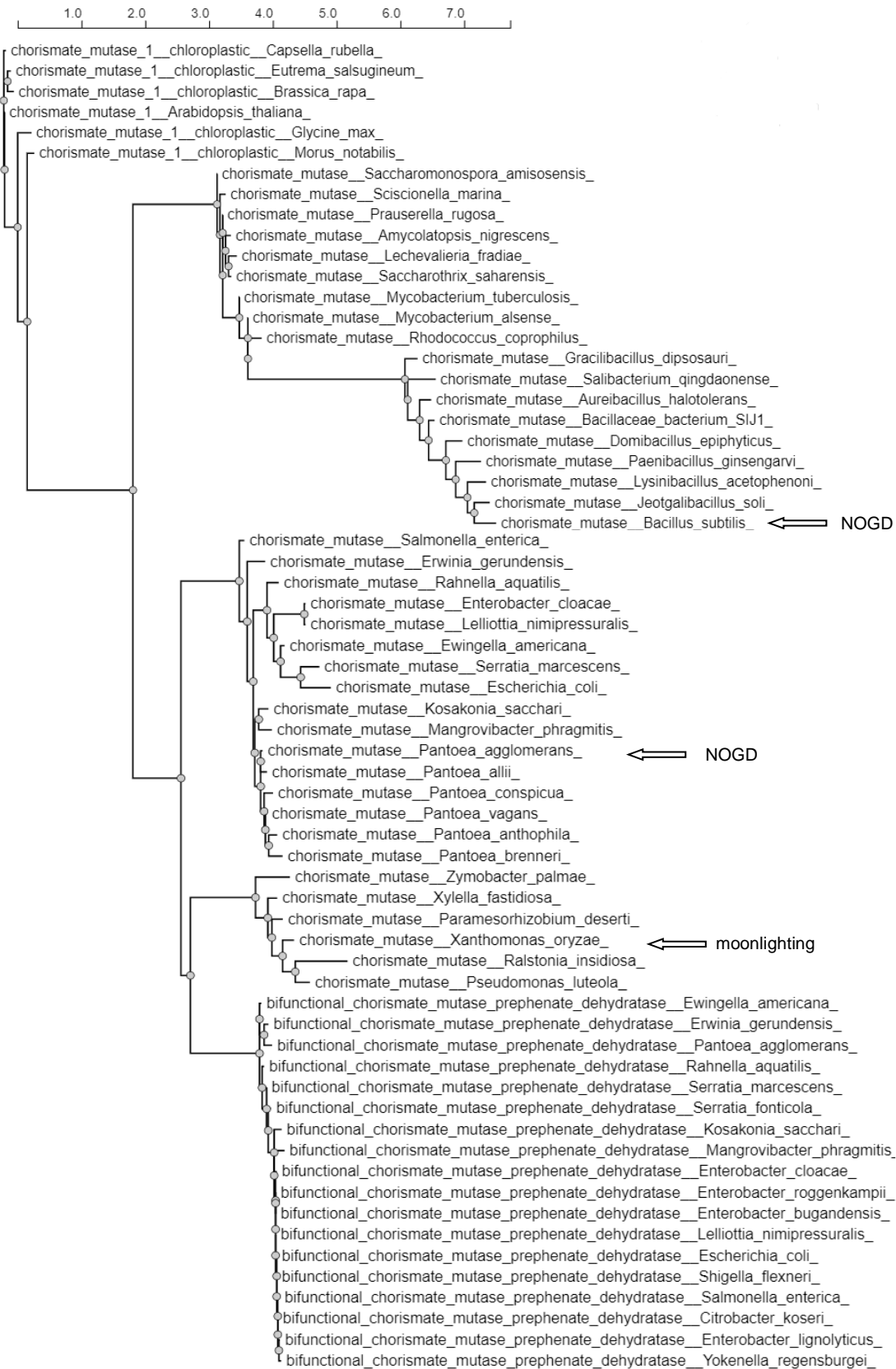


Figura 7. Árbol filogenético de proteínas con función corismato mutasa. Se señalan las proteínas NOGD y *moonlighting* correspondientes.

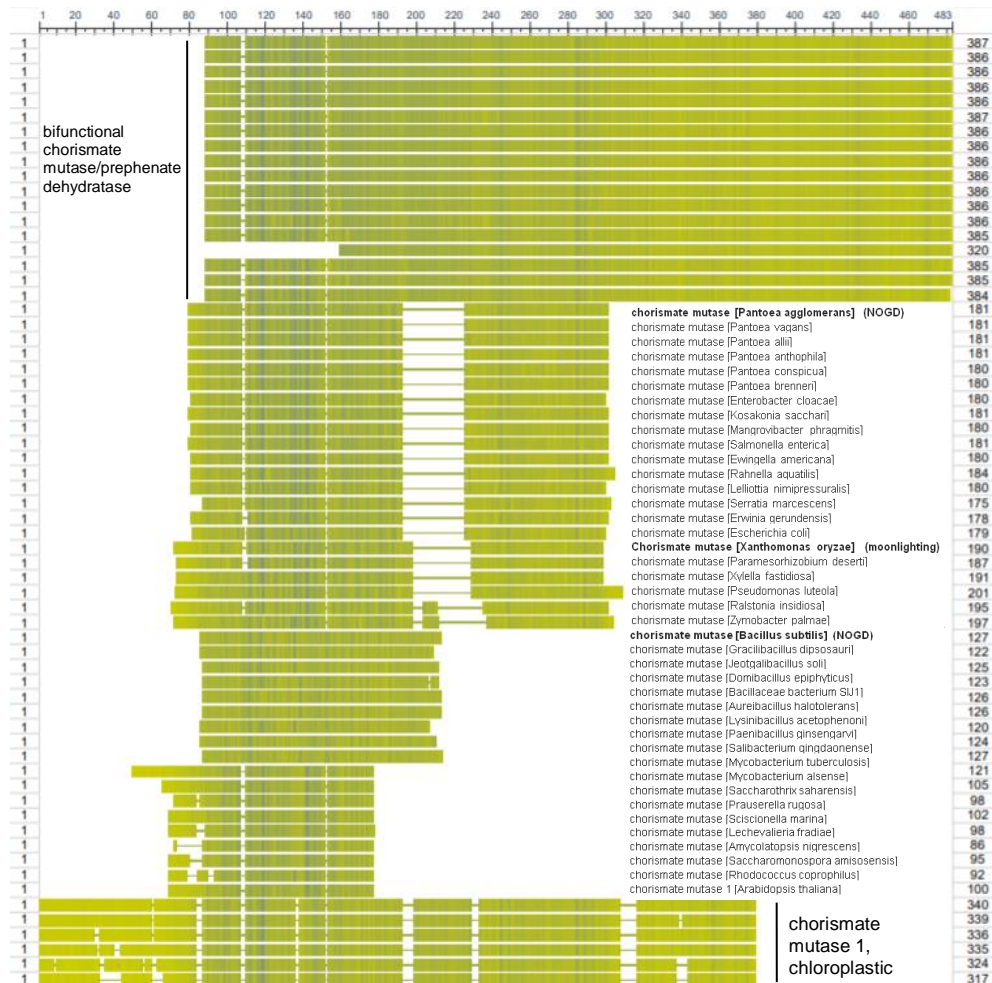


Figura 8. Alineamiento de proteínas con función corismato mutasa. Realizado con COBALT y coloreado con BLOSUM45. Se representa una vista global del alineamiento entre las proteínas a través del coloreado, sin mostrar las secuencias, por simplicidad. Se colorean en verde los aminoácidos no compartidos y en azul los conservados.

Fosfolipasa A2

Este es un ejemplo de relación más compleja. Se han encontrado tres proteínas NOGD que realizan esta función, la fosfolipasa A1 de *E.coli* (con funciones fosfolipasa A1 y A2) y dos formas de fosfolipasa A2 humana (secretada y citosólica). Esta función también es realizada por la proteína multifuncional peroxirredoxina-6 humana. A su vez, la peroxirredoxina-6 tiene otras funciones en común con otras proteínas NOGD: la función lisofosfolipasa, que también realiza la fosfolipasa A2 humana citosólica y se analizará a continuación, y la función peroxirredoxina, que no se ha analizado por simplicidad al ser muchas las proteínas que la realizan.

En la figura 9 se muestra el árbol obtenido, en el que se observa una relación evolutiva entre la *moonlighting* peroxirredoxina-6 de *Homo sapiens* y las fosfolipasas A2 citosólicas, mientras que las otras proteínas NOGD son más lejanas a estas. El alineamiento entre estas proteínas se puede consultar en el anexo II.

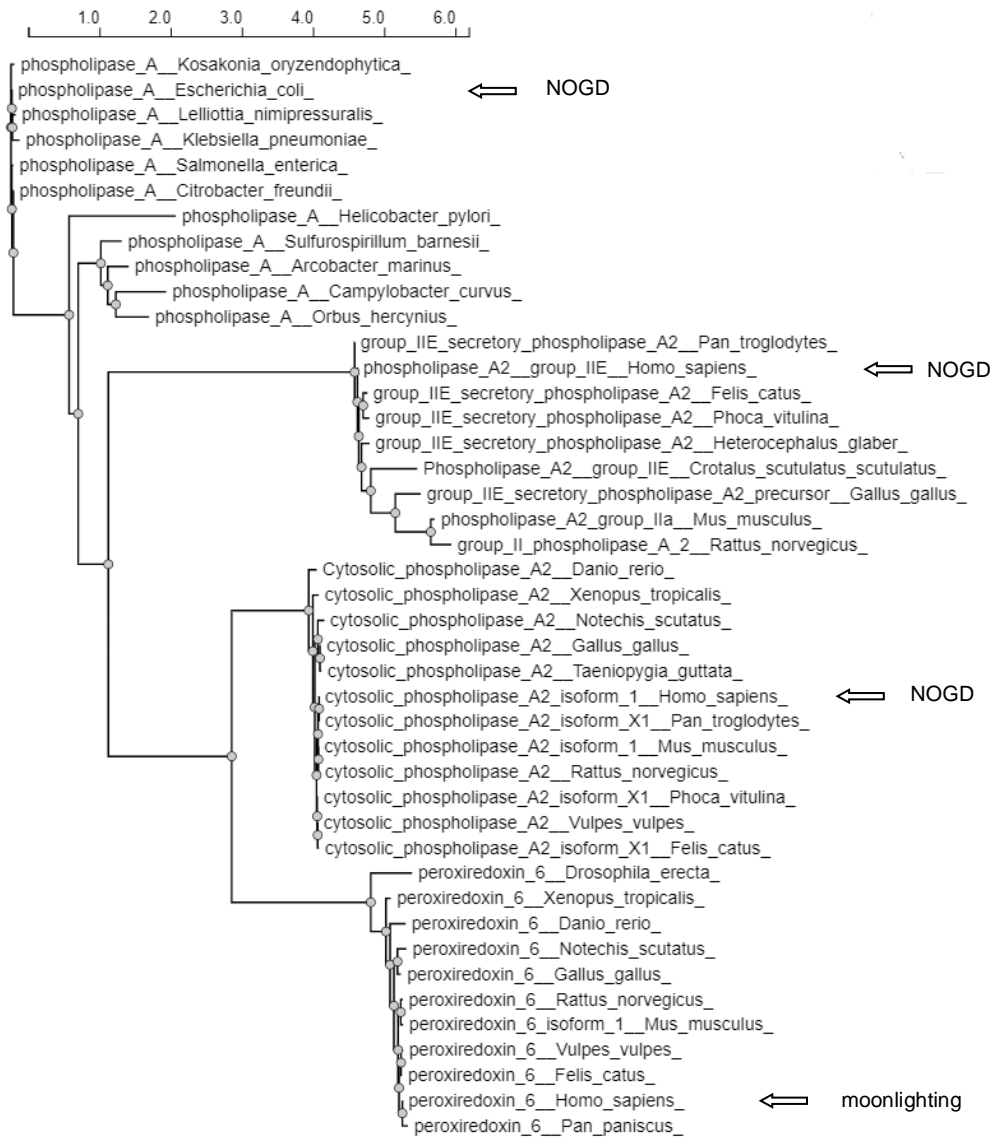


Figura 9. Árbol filogenético de proteínas con función fosfolipasa A2. Se señalan las proteínas NOGD y *moonlighting* correspondientes.

Lisofosfolipasa

Además de la peroxirredoxina-6 y la fosfolipasa A2 humana citosólica, realizan esta función las proteínas NOGD lisofosfolipasa de *Rattus norvegicus* y *Escherichia coli*. En el árbol filogenético correspondiente (figura 10), se observa de nuevo la relación entre la peroxirredoxina-6 y la fosfolipasa A2 y también una mayor relación entre éstas y la lisofosfolipasa NOGD de *Escherichia coli*, mientras el resto de lisofosfolipasas animales relacionadas con la de *Rattus norvegicus* se separan de ellas. El alineamiento entre estas proteínas se puede consultar en el anexo II.

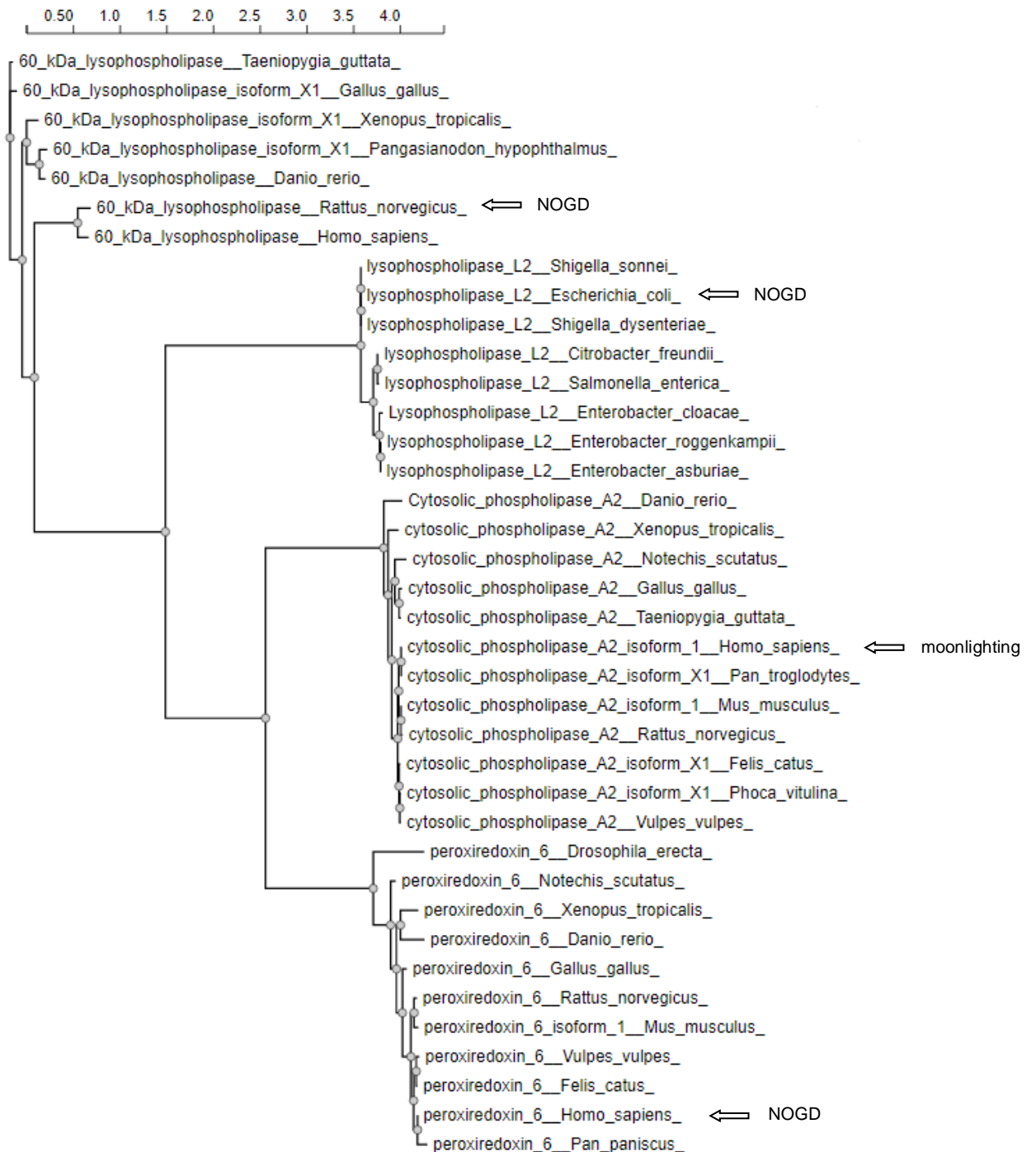


Figura 10. Árbol filogenético de proteínas con función lisofosfolipasa. Se señalan las proteínas NOGD y *moonlighting* correspondientes.

α-N-acetilgalactosaminidasa

Realizan esta función la proteína α-N-acetilgalactosaminidasa (NAGA) de *Homo Sapiens*, que es a la vez NOGD y *moonlighting*, y la proteína NOGD α-N-acetilgalactosaminidasa de *Elizabethkingia meningoseptica*. En el árbol filogenético (figura 11) se observa que pertenecen a dos grupos de proteínas con líneas evolutivas diferentes: la familia Gfo/LDH/MocA oxidorreductasa y las α-N-acetilgalactosaminidasa. El alineamiento entre estas proteínas se puede consultar en el anexo II.

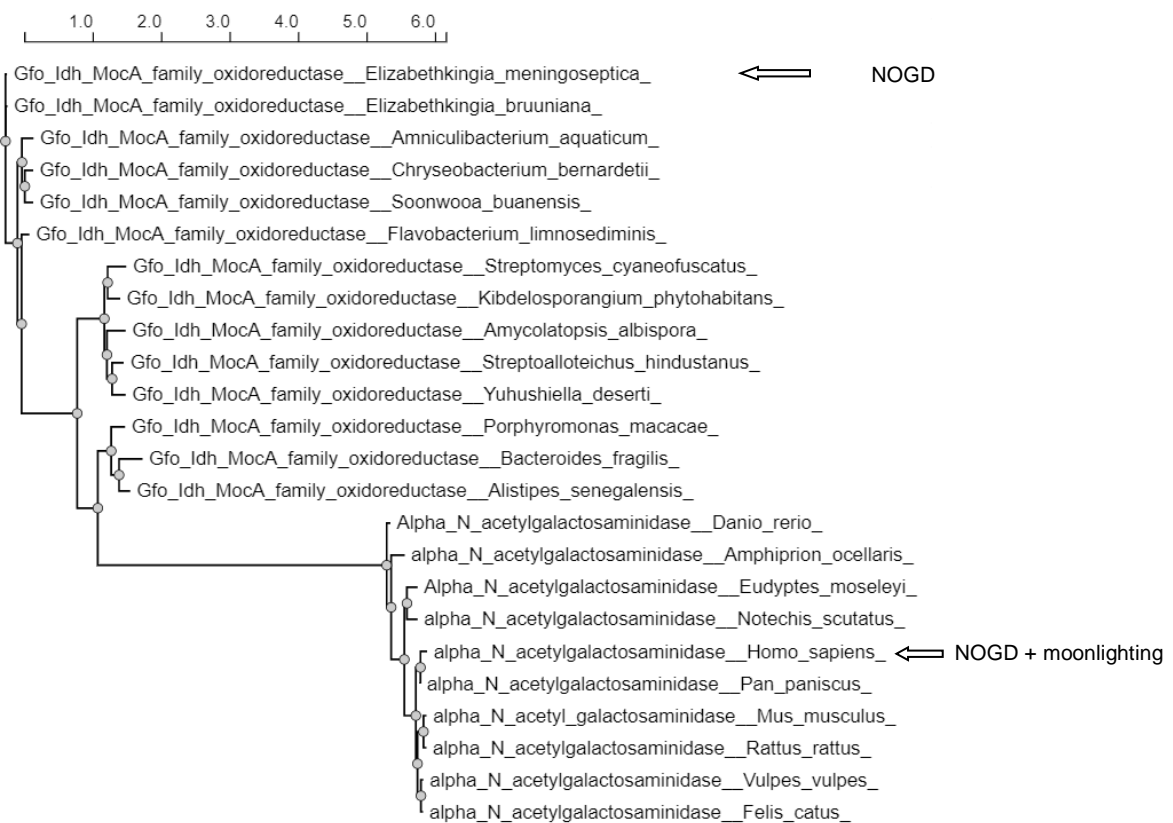


Figura 11. Árbol filogenético de proteínas con función α -N-acetilgalactosaminidasa. Se señalan las proteínas NOGD y *moonlighting* correspondientes.

Endo- β -N-acetilglucosaminidasa

Tienen esta función la endo- β -N-acetilglucosaminidasa putativa de *Streptococcus pneumoniae*, que es a la vez NOGD y *moonlighting*, y la proteína NOGD manosil-glicoproteína endo- β -N-acetilglucosaminidasa de *Elizabethkingia meningoseptica*. En la figura 12 se muestra el árbol filogenético construido en el que no se observan dos líneas evolutivas diferenciadas como en los ejemplos anteriores. Sin embargo, es destacable el hecho de que la proteína multifuncional (endo- β -N-acetilglucosaminidasa de *Streptococcus pneumoniae*) se agrupe con las de otras dos especies de *Streptococcus* pero la de *Streptococcus pyogenes* esté separada de estas. De hecho, la endo- β -N-acetilglucosaminidasa de *Streptococcus pyogenes* se obtuvo del PSI-BLAST de la de *Elizabethkingia meningoseptica* y no del de *Streptococcus pneumoniae*.

El alineamiento entre estas proteínas se puede consultar en el anexo II.

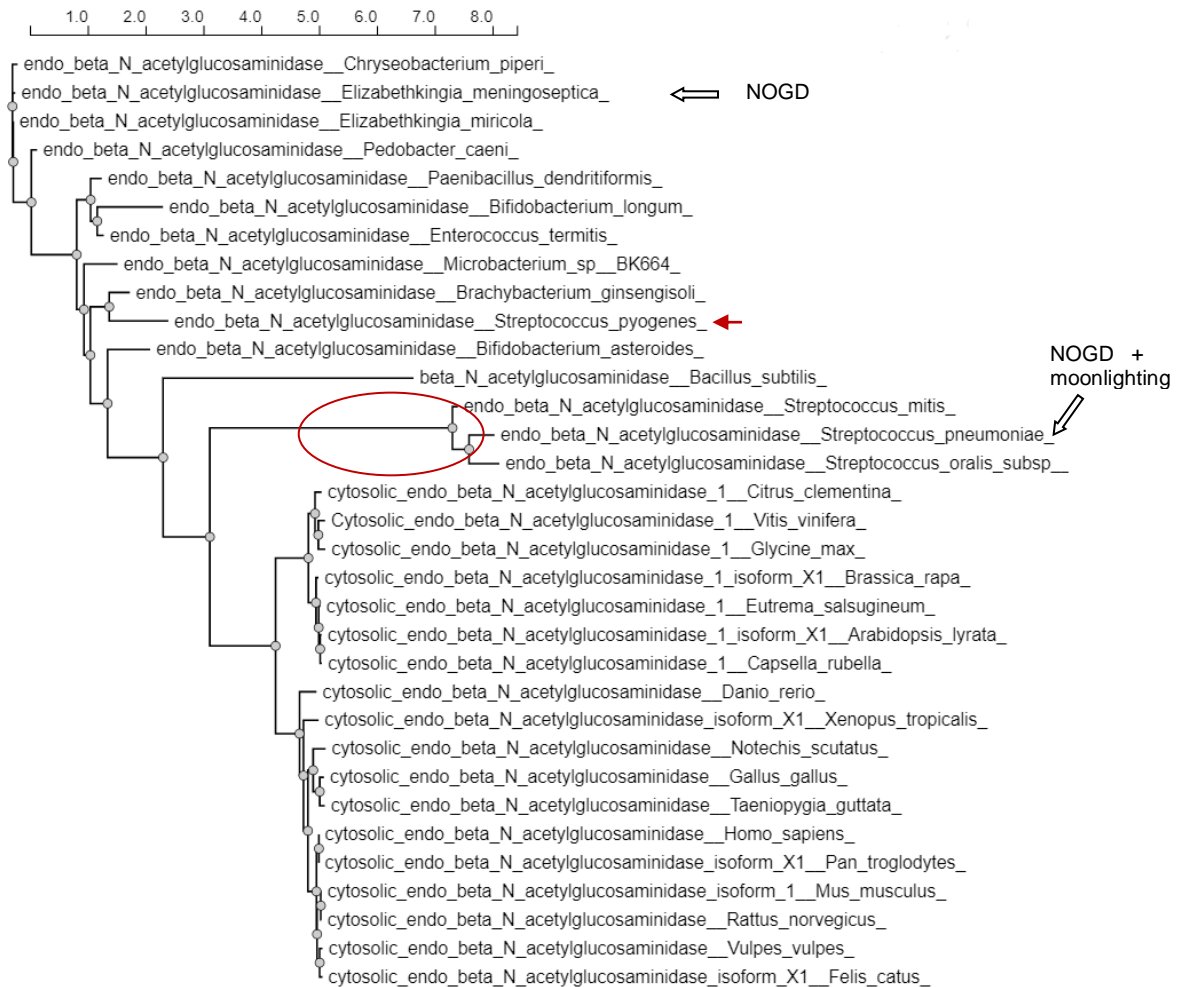


Figura 12. Árbol filogenético de proteínas con función endo-β-N-acetilglucosaminidasa. Se señalan las proteínas NOGD y *moonlighting* correspondientes, así como la de *Streptococcus pyogenes* (flecha roja) por su llamativa separación de las del resto de especies de *Streptococcus* (círculo rojo).

ARNt (guanine-N1-)-metiltransferasa

Tienen esta función las proteínas NOGD tRNA (guanine-N1-)-metiltransferasa de *Homo Sapiens* y *Escherichia coli* y la proteína multifuncional RNasa P mitocondrial de *Homo Sapiens*. En el árbol filogenético construido (figura 13), se observa, como era de esperar, mayor relación entre la proteína multifuncional RNasa P mitocondrial/ARNt-metiltransferasa homólogo C y la NOGD tRNA (guanine-N1-)-metiltransferasa de *Homo Sapiens* que con la de *Escherichia coli*. Sin embargo, están más relacionadas entre sí que con las metiltransferasas de algas y hongos. El alineamiento entre estas proteínas se puede consultar en el anexo II.

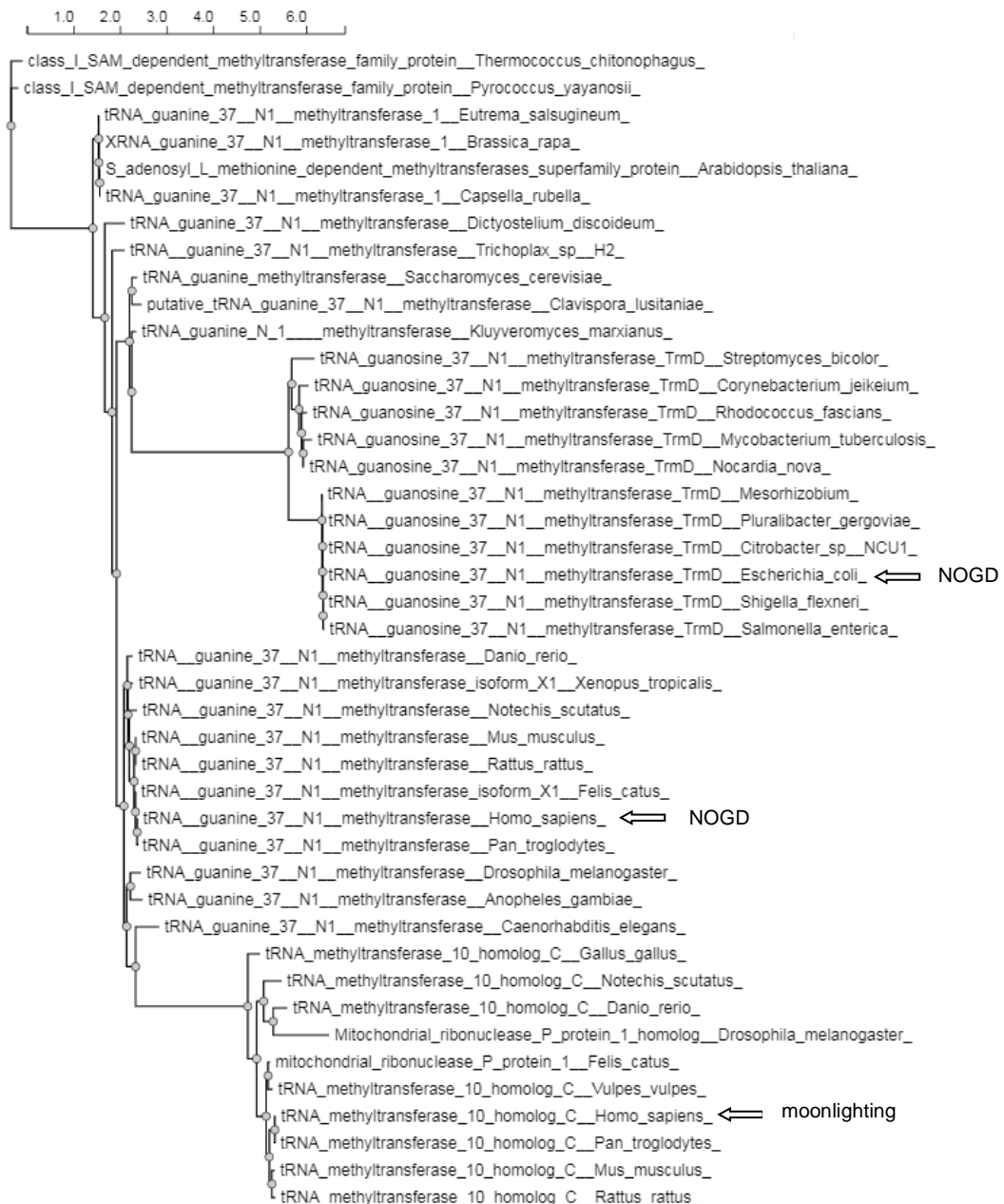


Figura 13. Árbol filogenético de proteínas con función ARNt metiltransferasa. Se señalan las proteínas NOGD y *moonlighting* correspondientes.

4. Discusión: Fenómeno NOGD como mecanismo de obtención de una función *moonlighting*

Como se comentó en la introducción, el fenómeno NOGD se da cuando un componente esperado en una vía enzimática está ausente pero su función es realizada por otra proteína estructuralmente diferente. Este fenómeno sugiere un mecanismo por el cual una proteína podría adquirir una nueva función para sustituir al componente faltante de la vía y se convertiría, por tanto, en una proteína multifuncional si conserva su función original. El objetivo principal del trabajo era comprobar esta hipótesis, motivada por el hecho de haberse

encontrado en trabajos previos varias proteínas multifuncionales que realizaban funciones de proteínas NOGD.

Al hacer una búsqueda comparativa entre la base de datos de proteínas multifuncionales y la de proteínas NOGD, se encontró que el 20% de las funciones recogidas en esta última estaban realizadas por proteínas multifuncionales, lo que refuerza los resultados observados en trabajos previos. Esto supone una relación funcional entre 124 proteínas *moonlighting* y 122 proteínas NOGD de las bases de datos consultadas, pero probablemente haya muchos más casos de proteínas no recogidas en las mismas o que aún queden por descubrir.

Para demostrar esta hipótesis es necesario realizar un estudio evolutivo de estas proteínas y encontrar su origen. Habría que encontrar en qué momento de la evolución surgió la sustitución de una proteína NOGD por otra y si ésta procedía de otra proteína que no realizaba esa función, habiendo perdido su función original para adquirir esta nueva o pudiendo mantener ambas y convertirse en multifuncional.

Sin embargo, este tipo de estudios son muy complicados y se conoce muy poco acerca de la evolución de las proteínas *moonlighting*, los mecanismos por los que surgen o si la multifuncionalidad se conserva filogenéticamente. Además, existen muchas proteínas multifuncionales y se cree que aún quedan muchas más por descubrir, por lo que estudiar estos aspectos se hace más complicado ya que cada proteína se habrá podido originar por mecanismos diferentes según las necesidades funcionales del organismo en el que surgió. De hecho, se ha observado que proteínas multifuncionales de diferentes organismos pueden tener diferentes funciones *moonlighting* en cada uno de ellos. Como concluye Copley, Shelley (2014) “*La evolución de cada proteína multifuncional es una historia que tiene lugar en el contexto de un conjunto de condiciones genéticas, ambientales y poblacionales escondidas a lo largo de millones de años de evolución*” [10].

De hecho, los pocos estudios publicados acerca de la evolución de estas proteínas suelen centrarse en una sola proteína o un grupo muy reducido. Por ejemplo, se ha hipotetizado el surgimiento de funciones *moonlighting* mediante duplicación génica en las chaperoninas [24], hexokinasa de *Saccharomyces cerevisiae* y galactokinasa de *Kluyveromyces lactis* [25].

Se ha propuesto que para que una proteína pueda ganar una función extra basta con unas pocas mutaciones en regiones clave para realizar esta función, que a su vez deben ocurrir en regiones no necesarias para su función original para poder mantener ambas. Se ha estudiado esta hipótesis en la proteína SMN de metazoos [26]. En los alineamientos múltiples realizados en este trabajo con las secuencias de proteínas que realizan la misma función se observó que realmente son muy pocos los aminoácidos conservados entre todas ellas, por lo que estos serían los aminoácidos clave que hay que adquirir para realizar esa función.

También se propone que estos cambios estarían promovidos por la interacción con nuevos factores, como otras proteínas, DNA o RNA. Si el fenómeno NOGD fuera un mecanismo por el cual adquirir una función adicional, la proteína multifuncional surgiría al interaccionar con el sustrato de la enzima ausente dando una reacción favorable que se conservaría evolutivamente.

En este trabajo, se ha estudiado la relación filogenética entre las proteínas *moonlighting* y NOGD que realizan una misma función y con otras relacionadas con cada una de ellas, abarcando un amplio abanico de filos. Los árboles filogenéticos construidos muestran que al menos una de las estructuras de las proteínas NOGD tiene relación evolutiva con la proteína *moonlighting*. Este hecho apoya la hipótesis que se quería comprobar, a pesar de no haber podido encontrar el origen de estas proteínas mediante homología remota con PSI-BLAST ni determinar en qué momento exacto surgió la adquisición de esta nueva función en las proteínas *moonlighting*.

Se ha comprobado, por tanto, una relación evolutiva entre los dos fenómenos que apoya la hipótesis inicial por la que el fenómeno NOGD podría ser un mecanismo por el cual una proteína podría adquirir una función adicional y convertirse en una proteína multifuncional.

5. Conclusiones

1.- Hay una evidente relación funcional entre proteínas *moonlighting* y NOGD, con numerosos ejemplos de las mismas que realizan la misma función y diferentes tipos de relación entre ellas. La relación más evidente entre los dos fenómenos se da en aquellas proteínas que son NOGD y multifuncionales a la vez. Se han encontrado 124 proteínas multifuncionales que realizan funciones de proteínas NOGD, pero probablemente haya muchos más casos no recogidos en las bases de datos consultadas o que aún queden por descubrir.

2.- No se ha podido encontrar el origen de las proteínas analizadas mediante homología remota con PSI-BLAST, lo que era necesario para demostrar la hipótesis a estudiar. Si se hubiera encontrado que al menos una de las proteínas NOGD que realizan la misma función procede de otra diferente, se demostraría con ello que este puede ser un mecanismo por el que surjan proteínas multifuncionales si al adquirir una nueva función por un mecanismo NOGD conserva su función anterior.

3.- Los árboles filogenéticos construidos muestran que una de las estructuras de las proteínas NOGD tiene relación evolutiva con la proteína *moonlighting*. Si bien no se ha podido demostrar la hipótesis de estudio hallando el origen de estas proteínas, esta relación evolutiva la apoya. Quedan abiertas futuras líneas de comparación entre estas proteínas para determinar la posible conservación filogenética de la multifuncionalidad.

4.- La relación entre los dos fenómenos parece estar clara y no se debería abandonar la hipótesis inicial. Quedan muchos más ejemplos de proteínas a las que no se les ha hecho el estudio evolutivo dada su cantidad y mayor complejidad, seguro que se obtendrán resultados muy interesantes al analizar estas proteínas con relaciones más complejas.

5.- Dado que no se puede asumir la multifuncionalidad de una proteína si no se demuestra experimentalmente, quedarían abiertas otras líneas de investigación

para intentar comprobar experimentalmente si las proteínas más cercanas evolutivamente a una proteína multifuncional también lo son.

6.- Como lección a nivel técnico he aprendido que al crear bases de datos o algoritmos de análisis es muy importante tener en cuenta el uso que se les vaya a poder dar para intentar facilitarlos en medida de lo posible. Por ejemplo, encontré que PSI-BLAST no estaba optimizado para el tipo de análisis que quería realizar pues en lugar de incluir una opción para deseleccionar todos los resultados a la vez para luego poder seleccionar los que interesen para la siguiente iteración, aparecen todos seleccionados por defecto y hay que ir deseleccionándolos uno a uno, lo cual es mucho más laborioso.

7.- Por último, como reflexión sobre el trabajo realizado y el seguimiento de la planificación inicial, creo que se pudieron solventar bien los inconvenientes tanto temporales como técnicos con una reprogramación adecuada y la búsqueda de herramientas alternativas, si bien se podrían haber analizado más proteínas con una mejor planificación y gestión del tiempo por mi parte.

6. Glosario

BLAST = Basic Local Alignment Search Tool

DHPR = dihydropteridine reductase

EC = Enzyme Commission

GO = Gen Ontology

NAGA = α -N-acetylgalactosaminidase

NOGD = Non-Orthologous gen displacement

NOS = nitric oxide synthase

PDB = Protein Data Bank

PSI-Blast = Position Specific Iterated BLAST

QDPR = quinoid dihydropteridine reductase

SDR = short chain dehydrogenase

7. Bibliografía

1. Jeffery, Constance J. "Moonlighting Proteins." *Trends in Biochemical Sciences*, vol. 24, no. 1, 1999, pp. 8–11.
2. Khan, Ishita K., and Daisuke Kihara. "Computational Characterization of Moonlighting Proteins." *Biochemical Society Transactions*, vol. 42, no. 6, Dec. 2014, pp. 1780–85.
3. Jeffery, Constance J. "An Introduction to Protein Moonlighting." *Biochemical Society Transactions*, vol. 42, no. 6, Dec. 2014, pp. 1679–83.
4. Jeffery, Constance J. "Moonlighting Proteins – Nature's Swiss Army Knives." *Science Progress*, vol. 100, no. 4, 2017, pp. 363–73.
5. Wistow, G., and J. Piatigorsky. "Recruitment of Enzymes as Lens Structural Proteins." *Science*, vol. 236, no. 4808, June 1987, pp. 1554–56.
6. Franco-Serrano, Luís, et al. "MultitaskProtDB-II: An Update of a Database of Multitasking/Moonlighting Proteins." *Nucleic Acids Research*, vol. 46, no. D1, Jan. 2018, pp. D645–48.
7. Franco-Serrano, Luis, Juan Cedano, et al. "A Hypothesis Explaining Why so Many Pathogen Virulence Proteins Are Moonlighting Proteins." *Pathogens and Disease*, vol. 76, no. 5, July 2018.
8. Franco-Serrano, Luis, Mario Huerta, et al. "Multifunctional Proteins: Involvement in Human Diseases and Targets of Current Drugs." *The Protein Journal*, vol. 37, no. 5, 2018, pp. 444–53.
9. Gupta, Mn, et al. "Medical Implications of Protein Moonlighting." *Indian Journal of Medical Research*, vol. 149, no. 3, 2019, p. 322.
10. Copley, Shelley D. "An Evolutionary Perspective on Protein Moonlighting." *Biochemical Society Transactions*, vol. 42, no. 6, Dec. 2014, pp. 1684–91.
11. Omelchenko, Marina V., et al. "Non-Homologous Isofunctional Enzymes: A Systematic Analysis of Alternative Solutions in Enzyme Evolution." *Biology Direct*, vol. 5, no. 1, 2010, p. 31.
12. "UniProt: A Worldwide Hub of Protein Knowledge." *Nucleic Acids Research*, vol. 47, no. D1, Jan. 2019, pp. D506–15.
13. Madeira, Fábio, et al. "The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019." *Nucleic Acids Research*, vol. 47, no. W1, July 2019, pp. W636–41.
14. Ye, Jian, et al. "BLAST: Improvements for Better Sequence Analysis." *Nucleic Acids Research*, vol. 34, no. Web Server issue, July 2006, pp. W6-9, doi:10.1093/nar/gkl164.
15. Altschul, S. F., et al. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research*, vol. 25, no. 17, Sept. 1997, pp. 3389–402.
16. Papadopoulos, Jason S., and Richa Agarwala. "COBALT: Constraint-Based Alignment Tool for Multiple Protein Sequences." *Bioinformatics (Oxford, England)*, vol. 23, no. 9, May 2007, pp. 1073–79.
17. Desper, Richard, and Olivier Gascuel. "Theoretical Foundation of the Balanced Minimum Evolution Method of Phylogenetic Inference and Its Relationship to Weighted Least-Squares Tree Fitting." *Molecular Biology and Evolution*, vol. 21, no. 3, Mar. 2004, pp. 587–98.
18. Saitou, N., and M. Nei. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution*, vol. 4, no. 4, July 1987, pp. 406–25.
19. Lemoine, Frédéric, et al. "NGPhylogeny.Fr: New Generation Phylogenetic Services for Non-Specialists." *Nucleic Acids Research*, vol. 47, no. W1, July 2019, pp. W260–65.
20. Katoh, Kazutaka, and Daron M. Standley. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution*, vol. 30, no. 4, Apr. 2013, pp. 772–80.

21. Criscuolo, Alexis, and Simonetta Gribaldo. "BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments." *BMC Evolutionary Biology*, vol. 10, no. 1, July 2010, p. 210.
22. Guindon, Stéphane, et al. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology*, vol. 59, no. 3, May 2010, pp. 307–21.
23. Ciccarelli, Francesca D., et al. "Toward Automatic Reconstruction of a Highly Resolved Tree of Life." *Science (New York, N.Y.)*, vol. 311, no. 5765, Mar. 2006, pp. 1283–87.
24. Fares, Mario A. "The Evolution of Protein Moonlighting: Adaptive Traps and Promiscuity in the Chaperonins." *Biochemical Society Transactions*, vol. 42, no. 6, Dec. 2014, pp. 1709–14, doi:10.1042/BST20140225.
25. Gancedo, Carlos, et al. "Evolution of Moonlighting Proteins: Insight from Yeasts." *Biochemical Society Transactions*, vol. 42, no. 6, Dec. 2014, pp. 1715–19, doi:10.1042/BST20140199.
26. Stump, Aram D., et al. "Molecular Evolution of the Moonlighting Protein SMN in Metazoans." *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, vol. 8, no. 3, 2013, pp. 220–30.

8. Anexos

Anexo I. Listado de proteínas multifuncionales y NOGD que realizan la misma función.

Función	Proteína	Organismo	UniProt	Moon/NOGD
6,7-dihydropteridine reductase EC. 1.5.1.34	DHPR (peptide C) dihydropteridine reductase	<i>Homo sapiens</i>	P09417	Moonlighting + NOGD
	6,7-dihydropteridine reductase	<i>Escherichia coli</i>	P38489	NOGD
Alpha-N-acetylgalactosaminidase EC. 3.2.1.49	Alpha-N-acetylgalactosaminidase	<i>Homo sapiens</i>	P17050	Moonlighting + NOGD
	Alpha-N-acetylgalactosaminidase	<i>Elizabethkingia meningoseptica</i>	A4Q8F7	NOGD
Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase EC. 3.2.1.96	Putative endo-beta-N-acetylglucosaminidase	<i>Streptococcus pneumoniae</i>	P59206	Moonlighting + NOGD
	Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	<i>Elizabethkingia meningoseptica</i>	P36912	NOGD
DUTP diphosphatase EC. 3.6.1.23	Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	<i>Homo sapiens</i>	P33316	Moonlighting + NOGD
	dCTP deaminase, dUMP-forming	<i>Methanocaldococcus jannaschii</i>	Q57872	Moonlighting
	DUTP diphosphatase	<i>Trypanosoma cruzi</i>	O15923	NOGD
Acetaldehyde dehydrogenase EC. 1.2.1.10	Alcohol dehydrogenase (EhADH2)	<i>Entamoeba histolytica</i>	Q24803	Moonlighting + NOGD
	Alcohol acetaldehyde dehydrogenase	<i>Listeria monocytogenes</i>	Q6Q3I2	Moonlighting
	Acetaldehyde dehydrogenase (acetylating)	<i>Escherichia coli</i>	P77580	NOGD
	Acetaldehyde dehydrogenase (acetylating)	<i>Pseudomonas putida</i>	Q49KG0	NOGD
Glutamate dehydrogenase EC. 1.4.1.2	Glutamate dehydrogenase	<i>Bacillus subtilis</i>	P39633	Moonlighting + NOGD
	Glutamate dehydrogenase 1, mitochondrial	<i>Gallus gallus</i>	P00368	Moonlighting
	Isocitrate--homocitrate dehydrogenase	<i>Pyrococcus horikoshii</i>	O59394	Moonlighting
	Glutamate dehydrogenase	<i>Achlya klebsiana</i>	P41755	NOGD
Aminoacyl-tRNA hydrolase EC. 3.1.1.29	Pth2/Bit1 peptidyl-tRNA hydrolase 2/Bcl-2 inhibitor of transcription 1 /Aminoacyl-tRNA hydrolase	<i>Homo sapiens</i>	Q9Y3E5	Moonlighting + NOGD
	DNA primase	<i>Pyrococcus abyssi</i>	Q9V1F2	Moonlighting
	Aminoacyl-tRNA hydrolase	<i>Escherichia coli</i>	P0A7D1	NOGD
	Aminoacyl-tRNA hydrolase	<i>Saccharolobus solfataricus</i>	Q980V1	NOGD
Glucose-6-phosphate isomerase EC. 5.3.1.9	Phosphoglucose isomerase	<i>Homo sapiens</i>	P06744	Moonlighting + NOGD
	Phosphoglucose isomerase	<i>Geobacillus stearothermophilus</i>	P13376	Moonlighting
	Phosphoglucose isomerase	<i>Xanthomonas oryzae</i>	B2SJM1	Moonlighting
	Glucose-6-phosphate isomerase	<i>Echinococcus multilocularis</i>	A7LKP5	Moonlighting
	Phosphoglucose isomerase	<i>Lactobacillus crispatus</i>	D5H2G7	Moonlighting
	Glucose-6-phosphate isomerase	<i>Mus musculus</i>	P06745	Moonlighting
	Glucose-6-phosphate isomerase	<i>Mustela nigripes</i>	M3Y807	Moonlighting
	Bifunctional phosphoglucose/phosphomannose isomerase	<i>Pyrobaculum aerophilum</i>	Q8ZWV0	Moonlighting
	Bifunctional phosphoglucose/phosphomannose isomerase	<i>Aeropyrum pernix</i>	Q9YE01	Moonlighting

	Bifunctional phosphoglucose/phosphomannose isomerase Glucose-6-phosphate isomerase	<i>Thermoplasma acidophilum</i> <i>Thermococcus litoralis</i>	Q9HIC2 P84140	Moonlighting NOGD
tRNA (guanine-N1-)-methyltransferase EC. 2.1.1.31	Mitochondrial RNase P tRNA (guanine-N1-)-methyltransferase tRNA (guanine-N1-)-methyltransferase	<i>Homo sapiens</i> <i>Homo sapiens</i> <i>Escherichia coli</i>	Q7L0Y3 Q32P41 P0A873	Moonlighting NOGD NOGD
Nitric-oxide synthase EC. 1.14.13.39	Endothelial Nitric Oxide Synthase Nitric-oxide synthase Nitric-oxide synthase	<i>Homo sapiens</i> <i>Rattus norvegicus</i> <i>Arabidopsis thaliana</i>	P29474 P29476 Q66GP9	Moonlighting NOGD NOGD
Chorismate mutase EC. 5.4.99.5	Chorismate mutase Chorismate mutase Chorismate mutase	<i>Xanthomonas oryzae</i> <i>Enterobacter agglomerans</i> <i>Bacillus subtilis</i>	Q5H4Y2 P42517 P19080	Moonlighting NOGD NOGD
Oleoyl-[acyl-carrier-protein] hydrolase EC. 3.1.2.14	Fatty acid synthase Oleoyl-[acyl-carrier-protein] hydrolase Oleoyl-[acyl-carrier-protein] hydrolase	<i>Homo sapiens</i> <i>Rattus norvegicus</i> <i>Saccharomyces cerevisiae</i>	P49327 P08635 P07149	Moonlighting NOGD NOGD
Protein-glutamine gamma-glutamyltransferase EC. 2.3.2.13	Protein-glutamine gamma-glutamyltransferase 2 Protein-glutamine gamma-glutamyltransferase Protein-glutamine gamma-glutamyltransferase Protein-glutamine gamma-glutamyltransferase	<i>Homo sapiens</i> <i>Homo sapiens</i> <i>Streptomyces cmobaraensis</i> <i>Bacillus subtilis</i>	P21980 P00488 P81453 P40746	Moonlighting NOGD NOGD NOGD
3',5'-cyclic-AMP phosphodiesterase EC. 3.1.4.53	3',5'-cyclic AMP phosphodiesterase CpdA 3',5'-cyclic-AMP phosphodiesterase 3',5'-cyclic-AMP phosphodiesterase	<i>Mycobacterium tuberculosis</i> <i>Dictyostelium discoideum</i> <i>Dictyostelium discoideum</i>	P9WP65 Q23917 P12019	Moonlighting NOGD NOGD
Fumarate hydratase EC:4.2.1.2	Fumarate hydratase Fumarase Fumarate hydratase Fumarate hydratase	<i>Homo sapiens</i> <i>Paracoccidioides brasiliensis</i> <i>Escherichia coli</i> <i>Escherichia coli</i>	P07954 C1GPZ3 P05042 P0AC33	Moonlighting Moonlighting NOGD NOGD
6-phosphofructokinase EC. 2.7.1.11	Phosphofructokinase Phosphofructokinase ATP-dependent 6-phosphofructokinase ATP-dependent 6-phosphofructokinase 6-phosphofructokinase 6-phosphofructokinase	<i>Pichia pastoris</i> <i>Bacillus subtilis</i> <i>Streptococcus oralis</i> <i>Lactococcus lactis</i> <i>Escherichia coli</i> <i>Escherichia coli</i>	Q92448 O34529 F2QDT8 G6FCR9 P0A796 P06999	Moonlighting Moonlighting Moonlighting Moonlighting NOGD NOGD
Amidase EC. 3.5.1.4	Amidase Putative endo-beta-N-acetylglucosaminidase Amidase Amidase	<i>Listeria monocytogenes</i> <i>Streptococcus pneumoniae</i> <i>Pseudomonas chlororaphis</i> <i>Pseudomonas aeruginosa</i>	O33983 P59206 P27765 P11436	Moonlighting Moonlighting NOGD NOGD
Testosterone 17- β -dehydrogenase EC. 1.1.1.63 Testosterone 17- β -dehydrogenase (NADP+) EC. 1.1.1.64	3-hydroxyacyl-CoA dehydrogenase type-2 Testosterone 17- β -dehydrogenase Testosterone 17- β -dehydrogenase Testosterone 17- β -dehydrogenase	<i>Homo sapiens</i> <i>Homo sapiens</i> <i>Homo sapiens</i> <i>Homo sapiens</i>	Q99714 O14756 P37058 P42330	Moonlighting NOGD NOGD NOGD

Peptidylprolyl isomerase EC. 5.2.1.8	Peptidyl Prolyl cis,trans-Isomerase (PPI)	<i>Helicobacter pylori</i>	D0K1B1	Moonlighting
	Peptidyl-prolyl cis-trans isomerase A	<i>Homo sapiens</i>	P62937	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Chlamydia trachomatis</i>	P26623	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Histoplasma capsulatum</i>	A6R3K7	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Neisseria gonorrhoeae</i>	P0A0W3	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Rickettsia prowazekii</i>	Q9ZCX6	Moonlighting
	PPI (Peptidyl Prolyl cis,trans-Isomerase (PPI))	<i>Streptococcus pneumoniae</i>	J1NX65	Moonlighting
	Peptidylprolyl isomerase	<i>Escherichia coli</i>	P0AFL3	NOGD
	Peptidylprolyl isomerase	<i>Escherichia coli</i>	P45523	NOGD
Protein-tyrosine-phosphatase EC. 3.1.3.48	Developmental protein eyes absent	<i>Drosophila melanogaster</i>	Q05201	Moonlighting
	Protein-tyrosine-phosphatase	<i>Streptomyces coelicolor</i>	P53433	NOGD
	Protein-tyrosine-phosphatase	<i>Homo sapiens</i>	P23468	NOGD
	Protein-tyrosine-phosphatase	<i>Homo sapiens</i>	P30307	NOGD
	Protein-tyrosine-phosphatase	<i>Bacillus subtilis</i>	P96717	NOGD
	Protein-tyrosine-phosphatase	<i>Mus musculus</i>	P97480	NOGD
L-lactate dehydrogenase EC.1.1.1.27 D-lactate dehydrogenase EC. 1.1.1.28	Upsilon-crystallin	<i>Ornithorhynchus anatinus</i>	Q7YQK6	Moonlighting
	Lactate dehydrogenase	<i>Homo sapiens</i>	P00338	Moonlighting
	Lactate dehydrogenase	<i>Rattus norvegicus</i>	P04642	Moonlighting
	L-lactate dehydrogenase B chain	<i>Anas platyrhynchos</i>	P13743	Moonlighting
	L-lactate dehydrogenase	<i>Homo sapiens</i>	P00338	NOGD
	L-lactate dehydrogenase	<i>Cupriavidus necator</i>	Q07251	NOGD
	D-lactate dehydrogenase	<i>Escherichia coli</i>	P52643	NOGD
D-lactate dehydrogenase	<i>Escherichia coli</i>	P06149	NOGD	
Superoxide dismutase EC. 1.11.1.5	Superoxide dismutase [Cu-Zn]	<i>Mycobacterium tuberculosis</i>	P9WGE9	Moonlighting
	Superoxide dismutase	<i>Mycobacterium avium subsp. avium</i>	B1A031	Moonlighting
	NAD(P)H dehydrogenase [quinone] 1	<i>Homo sapiens</i>	P15559	Moonlighting
	Superoxide dismutase	<i>Escherichia coli</i>	P06149	NOGD
	Superoxide dismutase	<i>Escherichia coli</i>	P53635	NOGD
	Superoxide dismutase	<i>Streptomyces seoulensis</i>	P80734	NOGD
Glutathione transferase EC. 2.5.1.18	Glutathione S-transferase M3 (GSTM)	<i>Homo sapiens</i>	Q6FGJ9	Moonlighting
	Ure2 (Transcriptional regulator URE2)	<i>Saccharomyces cerevisiae</i>	P23202	Moonlighting
	Eukaryotic translation elongation factor 1 epsilon-1	<i>Mus musculus</i>	Q9D1M4	Moonlighting
	Glutathione transferase	<i>Homo sapiens</i>	P10620	NOGD
	Glutathione transferase	<i>Homo sapiens</i>	Q9Y2Q3	NOGD
	Glutathione transferase	<i>Escherichia coli</i>	P0A9D2	NOGD
	Glutathione transferase	<i>Serratia marcescens</i>	Q56415	NOGD
Phosphoprotein phosphatase EC. 3.1.3.16	Pyruvate, orthophosphatedikinase regulatory protein (PPDK-RP)	<i>Arabidopsis thaliana</i>	O49562	Moonlighting
	Developmental protein eyes absent	<i>Drosophila melanogaster</i>	Q05201	Moonlighting
	Stage II sporulation protein E	<i>Bacillus subtilis</i>	P37475	Moonlighting
	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	<i>Homo sapiens</i>	P36873	Moonlighting

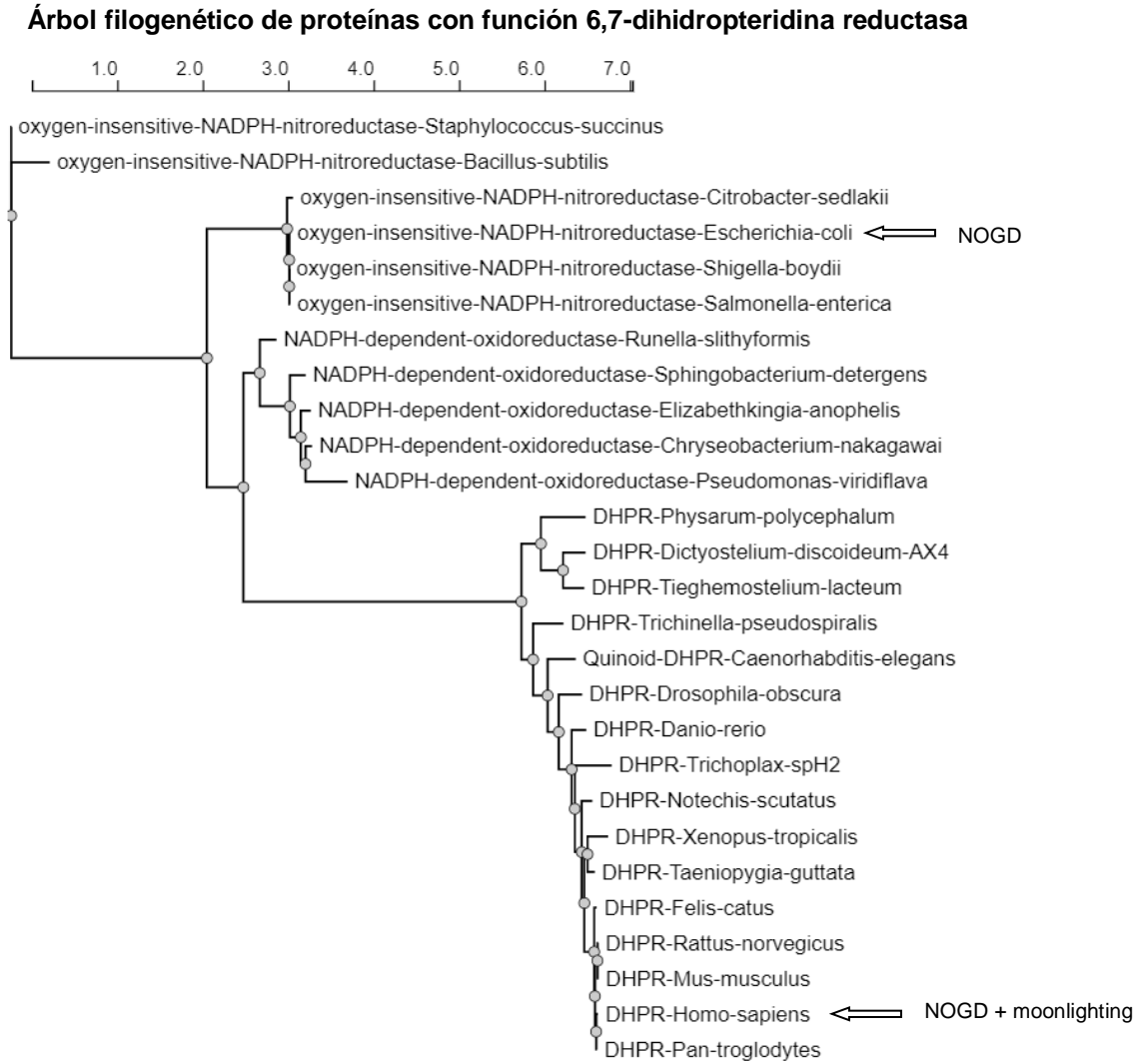
	Phosphoprotein phosphatase	<i>Homo sapiens</i>	Q9GZU7	NOGD
	Phosphoprotein phosphatase	<i>Homo sapiens</i>	Q8WTR2	NOGD
	Phosphoprotein phosphatase	<i>Escherichia coli</i>	P55798	NOGD
	Phosphoprotein phosphatase	<i>Bacillus subtilis</i>	O34779	NOGD
DNA-(apurinic or apyrimidinic site) lyase EC. 4.2.99.18	S3	<i>Homo sapiens</i>	P23396	Moonlighting
	S3	<i>Drosophila melanogaster</i>	Q06559	Moonlighting
	60S acidic ribosomal protein P0	<i>Drosophila melanogaster</i>	P19889	Moonlighting
	DNA-(apurinic or apyrimidinic site) lyase	<i>Homo sapiens</i>	P27695	NOGD
	DNA-(apurinic or apyrimidinic site) lyase	<i>Escherichia coli</i>	P20625	NOGD
	DNA-(apurinic or apyrimidinic site) lyase	<i>Escherichia coli</i>	P05523	NOGD
Thiamine-phosphate synthase EC. 2.5.1.3	Thiamine biosynthetic bifunctional enzyme TH1	<i>Arabidopsis thaliana</i>	Q5M731	Moonlighting
	Thiamine-phosphate synthase	<i>Escherichia coli</i>	P30137	NOGD
	Thiamine-phosphate synthase	<i>Methanothermobacter thermoautotrophicus</i>	O26949	NOGD
Malate dehydrogenase EC. 1.1.1.37	Malate dehydrogenase, mitochondrial	<i>Homo sapiens</i>	P40926	Moonlighting
	Malate dehydrogenase	<i>Haloarcula marismortui</i>	Q07841	NOGD
	Malate dehydrogenase	<i>Methanothermus fervidus</i>	P16142	NOGD
Alpha-glucosidase EC. 3.2.1.20	Trex (Alpha-1,4-Transferase)	<i>Sulfolobus solfataricus</i>	Q7LX99	Moonlighting
	Alpha-glucosidase	<i>Candida albicans</i>	Q02751	NOGD
	Alpha-glucosidase	<i>Thermotoga maritima</i>	O33830	NOGD
Cysteine synthase EC. 2.5.1.47	Cysteine synthase	<i>Bacillus subtilis</i>	P37887	Moonlighting
	Cysteine synthase	<i>Caenorhabditis elegans</i>	Q93244	Moonlighting
	Cysteine synthase	<i>Escherichia coli (strain K12)</i>	P0ABK5	Moonlighting
	Cysteine synthase	<i>Arabidopsis thaliana</i>	P47998	NOGD
	Cysteine synthase	<i>Saccharomyces cerevisiae</i>	P06106	NOGD
IMP cyclohydrolase EC. 3.5.4.10	Phosphoribosylaminoimidazolecarboxamideformyltransferase-IMP cyclohydrolase (AICARFT-IMPCHase)	<i>Arabidopsis thaliana</i>	Q8RWT5	Moonlighting
	IMP cyclohydrolase	<i>Homo sapiens</i>	P31939	NOGD
	IMP cyclohydrolase	<i>Methanotherx thermoacetophila</i>	A0B9N7	NOGD
3-dehydroquinate dehydratase EC. 4.2.1.10	3-Dehydroquinate dehydratase-shikimate-NADP-oxidoreductase (DHQase-SORase)	<i>Arabidopsis thaliana</i>	Q9SQT8	Moonlighting
	3-dehydroquinate dehydratase	<i>Actinobacillus pleuropneumoniae</i>	P43877	NOGD
	3-dehydroquinate dehydratase	<i>Escherichia coli</i>	P05194	NOGD
Lysine-tRNA ligase EC. 6.1.1.6	Lysine-tRNA ligase	<i>Homo sapiens</i>	Q15046	Moonlighting
	Lysine-tRNA ligase	<i>Escherichia coli</i>	P0A8N3	NOGD
	Lysine-tRNA ligase	<i>Aeropyrum pernix</i>	Q9YFT9	NOGD
Dihydrofolate reductase EC. 1.5.1.3	Dihydrofolate-reductasethymidylate-synthase (DHFR-TS)	<i>Arabidopsis thaliana</i>	Q05762	Moonlighting
	Bifunctional dihydrofolate reductase-thymidylate synthase	<i>Homo sapiens</i>	Q86XF0	Moonlighting
	Bifunctional dihydrofolate reductase-thymidylate synthase	<i>Plasmodium falciparum</i>	P13922	Moonlighting
	Dihydrofolate reductase	<i>Bacillus subtilis</i>	P11045	NOGD
	Dihydrofolate reductase	<i>Escherichia coli</i>	P0AFS3	NOGD

Fructokinase EC. 2.7.1.4	Hexokinase	<i>Escherichia coli</i>	Q47223	Moonlighting
	Hexokinase	<i>Arabidopsis thaliana</i>	Q9T071	Moonlighting
	Hexokinase	<i>Saccharomyces cerevisiae</i>	P04806	Moonlighting
	Hexokinase	<i>Leishmania donovani</i>	Q4VRV1	Moonlighting
	Hexokinase	<i>Plasmodium falciparum</i>	Q02155	Moonlighting
	Hexokinase	<i>Homo sapiens</i>	P19367	Moonlighting
	Hexokinase-2	<i>Homo sapiens</i>	P52789	Moonlighting
	Hexokinase (HXK1)	<i>Arabidopsis, various species</i>	Q42525	Moonlighting
	Fructokinase	<i>Escherichia coli</i>	P40713	NOGD
	Fructokinase	<i>Streptococcus mutans</i>	Q07211	NOGD
lysozyme EC. 3.2.1.17	LytC (Lysozyme)	<i>Streptococcus pneumoniae</i>	B2IR92	Moonlighting
	Lactose synthetase (Alpha-lactalbumin)	<i>Bos Taurus</i>	P00711	Moonlighting
	AtlA (autolysin)	<i>Streptococcus mutans</i>	A7VMH8	Moonlighting
	lysozyme	<i>Clostridium acetobutylicum</i>	P34020	NOGD
	lysozyme	<i>Escherichia phage P1</i>	Q37875	NOGD
	lysozyme	<i>Bacillus subtilis</i>	P10773	NOGD
NADH dehydrogenase EC. 1.6.99.3	Zeta-crystallin	<i>Cavia porcellus</i>	P11415	Moonlighting
	Sarcosine oxidase, alpha subunit	<i>Thermococcus kodakarensis</i>	Q5JFD7	Moonlighting
	NADH dehydrogenase	<i>Homo sapiens</i>	P28331	NOGD
	NADH dehydrogenase	<i>Escherichia coli</i>	P00393	NOGD
Non-specific protein-tyrosine kinase EC. 2.7.10.2	HSP90-alpha (Heat shock protein HSP 90-alpha)	<i>Homo sapiens</i>	P07900	Moonlighting
	FGF1 (Fibroblast growth factor 1)	<i>Homo sapiens</i>	P05230	Moonlighting
	FGF2 (fibroblast growth factor 2)	<i>Homo sapiens</i>	P09038	Moonlighting
	Epidermal growth factor receptor	<i>Homo sapiens</i>	P00533	Moonlighting
	Dual specificity protein kinase CLK1	<i>Homo sapiens</i>	P49759	Moonlighting
	Non-specific protein-tyrosine kinase	<i>Homo sapiens</i>	Q9UIG0	NOGD
	Non-specific protein-tyrosine kinase	<i>Streptococcus pneumoniae</i>	Q54520	NOGD
	Fructose- bisphosphatase EC. 3.1.3.11	Bifunctional NADP phosphatase/NAD kinase	<i>Methanococcus jannaschii</i>	Q58327
Fructose-1,6-bisphosphatase/inositol-1-monophosphatase		<i>Archaeoglobus fulgidus</i>	O30298	Moonlighting
Fructose-1,6-bisphosphatase		<i>Sulfolobus tokodaii</i>	F9VMT6	Moonlighting
Fructose-bisphosphatase		<i>Cupriavidus necator</i>	P19912	NOGD
Fructose-bisphosphatase		<i>Bacillus subtilis</i>	Q45597	NOGD
Fructose-bisphosphatase		<i>Pyrococcus furiosus</i>	Q8U359	NOGD
Fructose-bisphosphatase		<i>Saccharomyces cerevisiae</i>	P36136	NOGD
Enoyl-CoA hydratase EC. 4.2.1.17	Fatty acid multifunctional protein (MFP)	<i>Rice (Oryza sativa Japonica Group)</i>	Q8W1L6	Moonlighting
	Fatty acid multifunctional protein (MFP)	<i>Arabidopsis thaliana</i>	Q9ZPI5	Moonlighting
	Trifunctional enzyme subunit beta, mitochondrial	<i>Homo sapiens</i>	P55084	Moonlighting
	Enoyl-CoA hydratase	<i>Escherichia coli</i>	P21177	NOGD
	Enoyl-CoA hydratase	<i>Mycobacterium tuberculosis</i>	P9WNP3	NOGD
Phosphoglycerate mutase EC. 5.4.2.1	Phosphoglycerate mutase	<i>Streptococcus sp. oral</i>	E0PZW1	Moonlighting
	Phosphoglycerate mutase	<i>Bifidobacterium longum</i>	Q8G5N9	Moonlighting

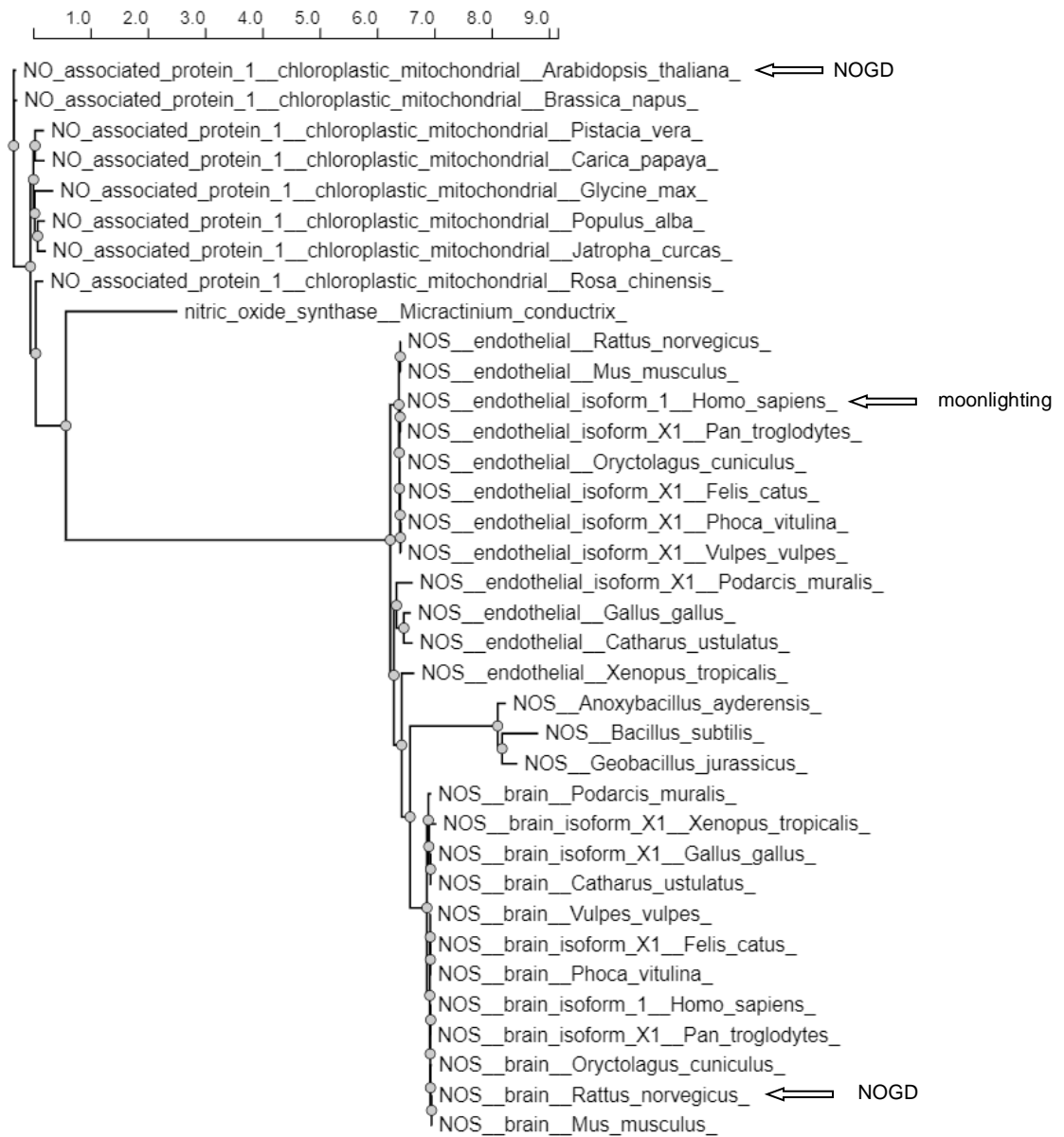
	Phosphoglycerate mutase	<i>Candida albicans</i>	P82612	Moonlighting
	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	<i>Streptococcus oralis</i>	F2QC85	Moonlighting
	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	<i>Streptococcus anginosus</i>	U2Z6L8	Moonlighting
	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	<i>Leishmania spp.</i>	Q4Q090	Moonlighting
	Phosphoglycerate mutase	<i>Bifidobacterium lactis</i>	B8DW74	Moonlighting
	Phosphoglycerate mutase	<i>Bifidobacterium bifidum</i>	E3ER63	Moonlighting
	Phosphoglycerate mutase	<i>Escherichia coli</i>	P31217	NOGD
	Phosphoglycerate mutase	<i>Pseudomonas syringae</i>	P52832	NOGD
N-acetylmuramoyl-L-alanine amidase EC. 3.5.1.28	N-acetylmuramoyl-L-alanine amidase sle1	<i>Staphylococcus aureus</i>	Q2G0U9	Moonlighting
	Amidase	<i>Listeria monocytogenes</i>	O33983	Moonlighting
	Surface protein pspA	<i>Streptococcus pneumoniae</i>	O34097	Moonlighting
	Surface protein PspC	<i>Streptococcus pneumoniae</i>	Q9FDQ1	Moonlighting
	Autolysin, LytA	<i>Streptococcus pneumoniae</i>	P06653	Moonlighting
	N-acetylmuramoyl-L-alanine amidase	<i>Escherichia coli</i>	P13016	NOGD
	N-acetylmuramoyl-L-alanine amidase	<i>Escherichia coli</i>	P26365	NOGD
Alcohol dehydrogenase EC. 1.1.1.1	hRoDH-E2 (Dehydrogenase/reductase SDR family member 9)	<i>Homo sapiens</i>	Q9BPW9	Moonlighting
	Alcohol dehydrogenase 1	<i>Candida albicans</i>	P43067	Moonlighting
	Alcohol dehydrogenase (EhADH2)	<i>Entamoeba histolytica</i>	Q24803	Moonlighting
	Alcohol acetaldehyde dehydrogenase	<i>Listeria monocytogenes</i>	Q6Q3I2	Moonlighting
	Alcohol dehydrogenase	<i>Saccharomyces cerevisiae</i>	P10127	NOGD
	Alcohol dehydrogenase	<i>Drosophila melanogaster</i>	P00334	NOGD
	Alcohol dehydrogenase	<i>Schizosaccharomyces pombe</i>	Q09669	NOGD
	Alcohol dehydrogenase	<i>Entamoeba histolytica</i>	Q24803	NOGD
Peroxidase EC. 1.11.1.7	Cyclooxygenase-1	<i>Homo sapiens</i>	P23219	Moonlighting
	Prostaglandin G/H synthase 1	<i>Ovis aries</i>	P23219	Moonlighting
	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial	<i>Homo sapiens</i>	P36969	Moonlighting
	Peroxiredoxin	<i>Aeropyrum pernix</i>	Q9Y9L0	NOGD
	Peroxiredoxin	<i>Mycobacterium tuberculosis</i>	P9WQB5	NOGD
	Peroxiredoxin	<i>Escherichia coli</i>	P0C0L2	NOGD
	Peroxidase	<i>Mus musculus</i>	O08709	NOGD
	Peroxidase	<i>Homo sapiens</i>	P05164	NOGD
	Peroxidase	<i>Escherichia coli</i>	P31545	NOGD
Peroxiredoxin EC. 1.11.1.15	Peroxiredoxin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Tsa1p (Peroxiredoxin TSA1)	<i>Candida albicans</i>	Q9Y7F0	Moonlighting
	cPrxl (Peroxiredoxin TSA1)	<i>Saccharomyces cerevisiae</i>	P34760	Moonlighting
	Peroxiredoxin TSA2	<i>Saccharomyces cerevisiae</i>	Q04120	Moonlighting
	Peroxiredoxin	<i>Aeropyrum pernix</i>	Q9Y9L0	NOGD
	Peroxiredoxin	<i>Mycobacterium tuberculosis</i>	P9WQB5	NOGD
	Peroxiredoxin	<i>Escherichia coli</i>	P0C0L2	NOGD

	Peroxidase	<i>Mus musculus</i>	O08709	NOGD
Phospholipase A2 EC. 3.1.1.4	Peroxiredoxin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Cytosolic phospholipase A2	<i>Homo sapiens</i>	P47712	NOGD
	Group IIE secretory phospholipase A2	<i>Homo sapiens</i>	Q9NZK7	NOGD
	Phospholipase A2	<i>Escherichia coli</i>	P0A921	NOGD
Lysophospholip ase EC. 3.1.1.5	Peroxiredoxin-6	<i>Homo sapiens</i>	P30041	Moonlighting
	Lysophospholipase	<i>Homo sapiens</i>	P47712	NOGD
	Lysophospholipase	<i>Rattus norvegicus</i>	O88202	NOGD
	Lysophospholipase	<i>Escherichia coli</i>	P07000	NOGD

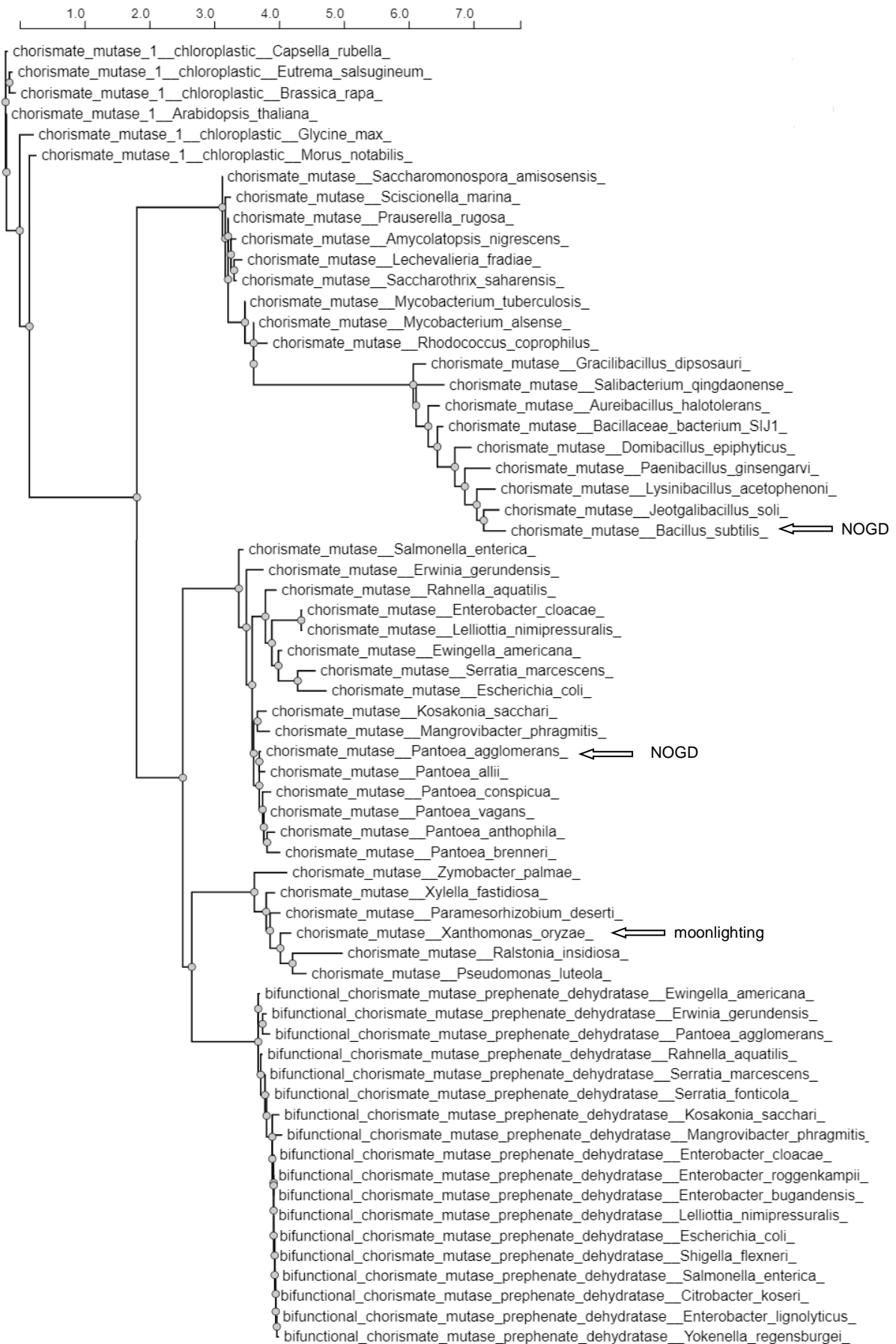
Anexo II. Árboles filogenéticos y alineamientos realizados con varios ejemplos de proteínas multifuncionales y NOGD que realizan la misma función.



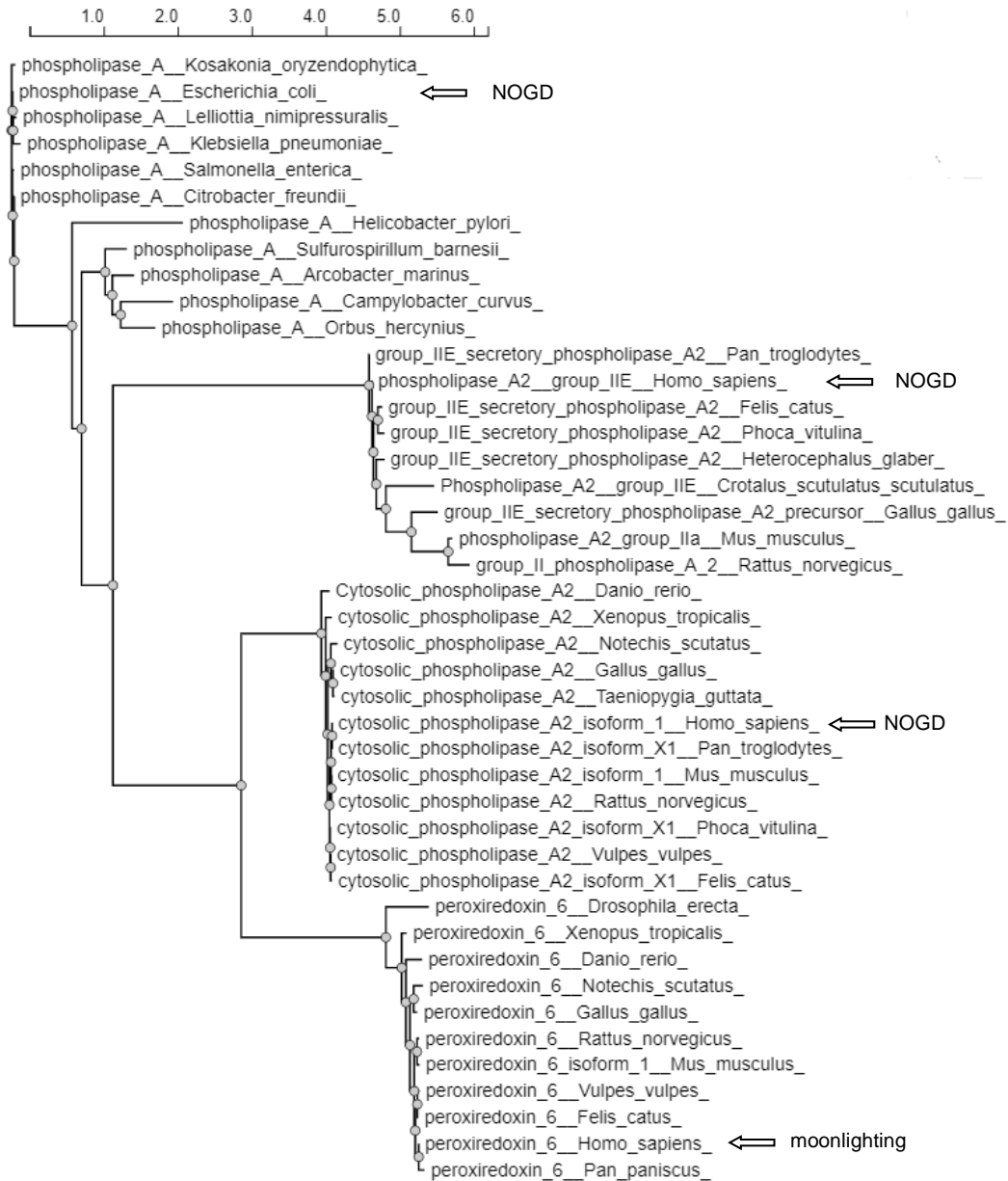
Árbol filogenético de proteínas con función óxido nítrico sintasa



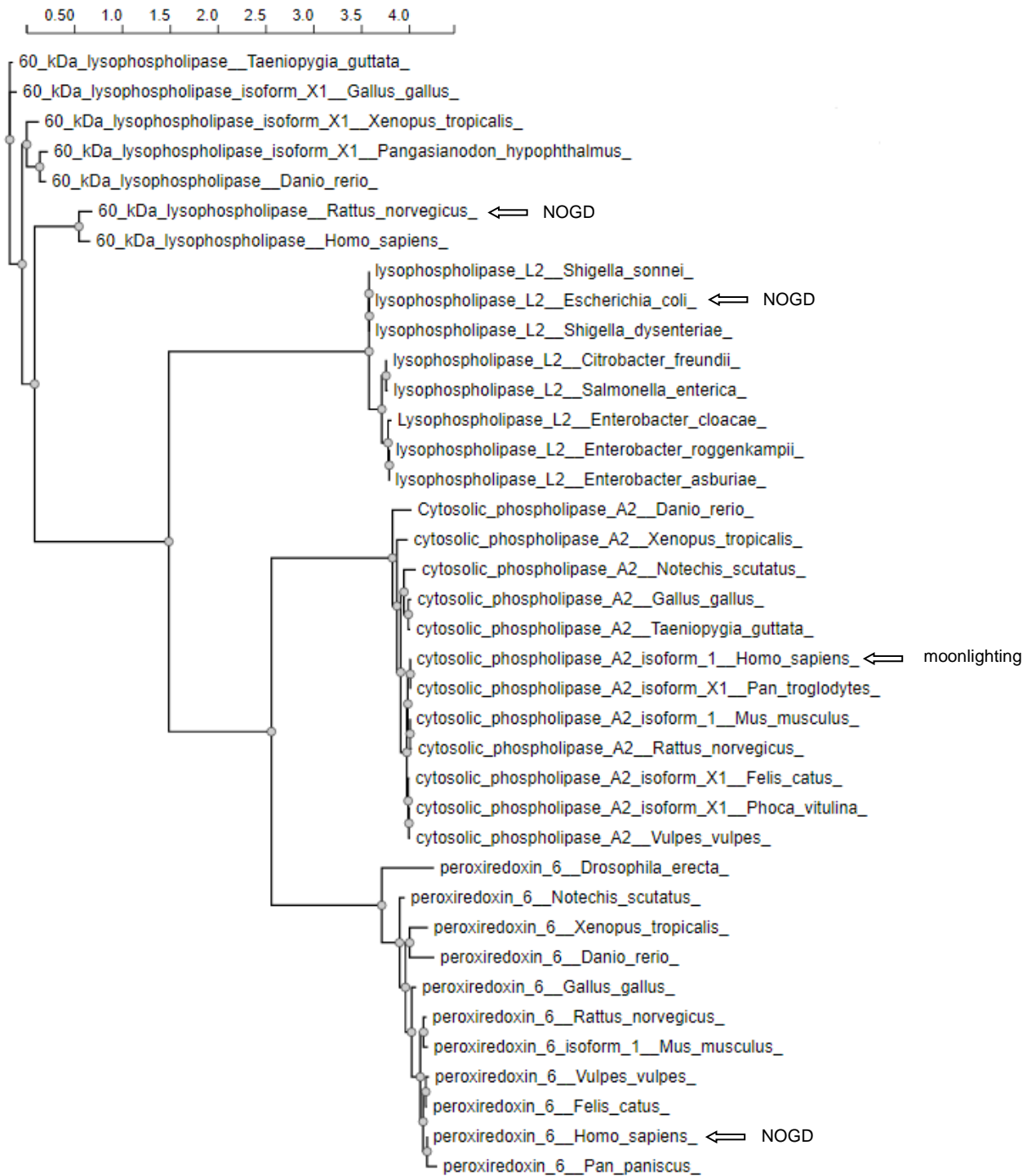
Árbol filogenético de proteínas con función corismato mutasa.



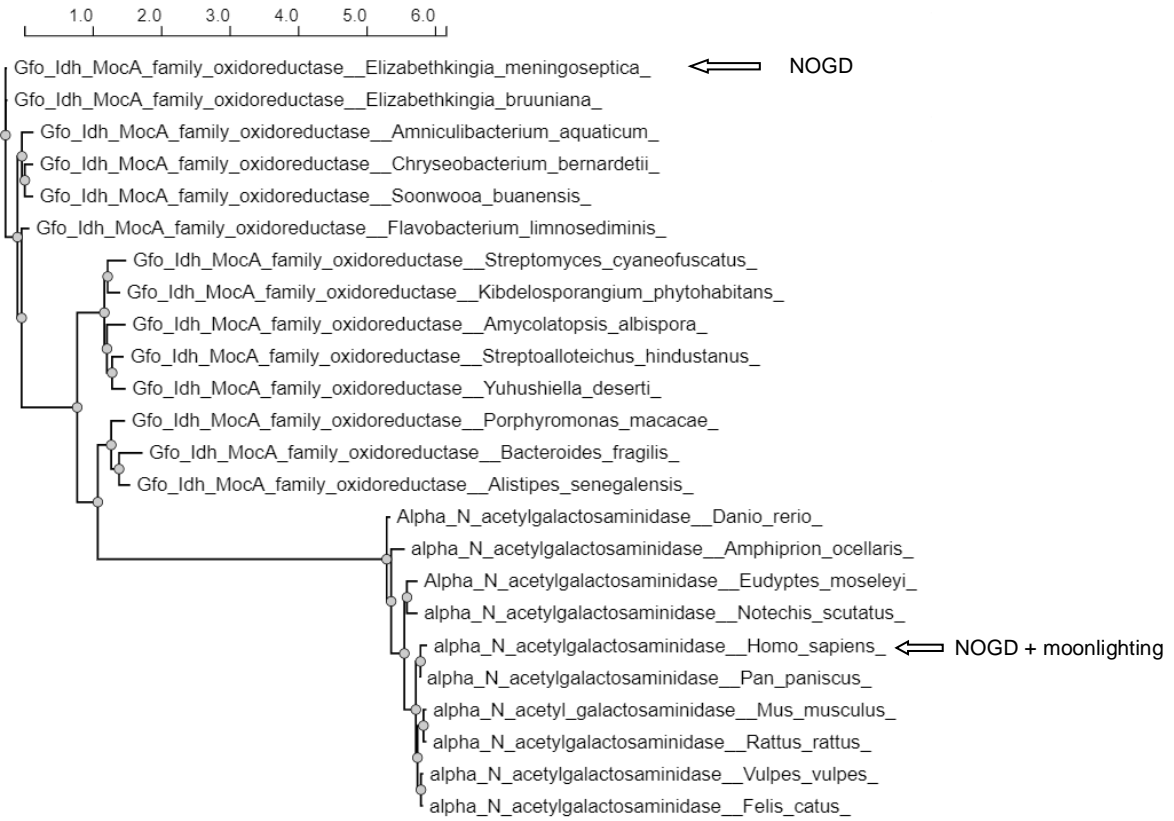
Árbol filogenético de proteínas con función fosfolipasa A2.



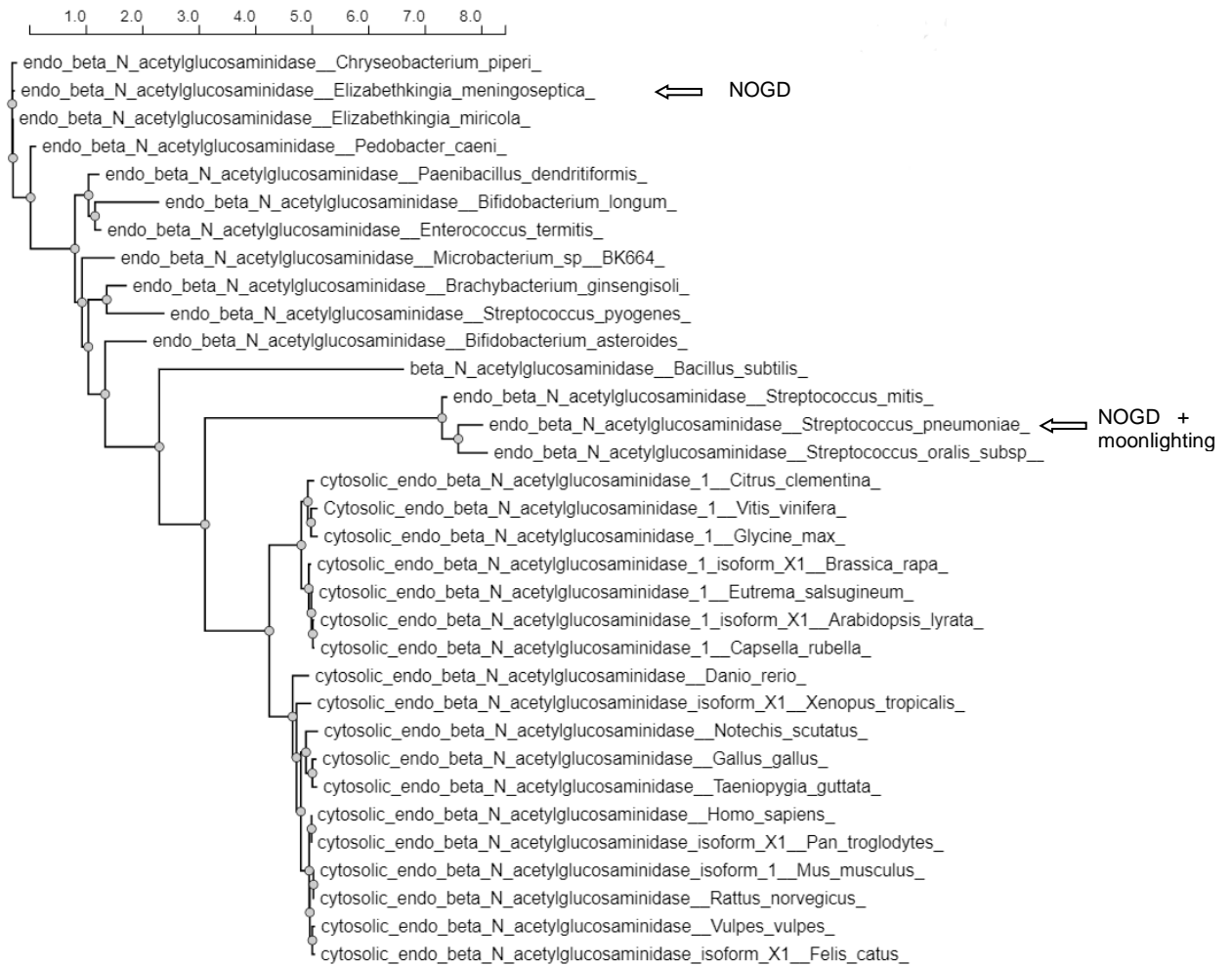
Árbol filogenético de proteínas con función lisofosfolipasa.



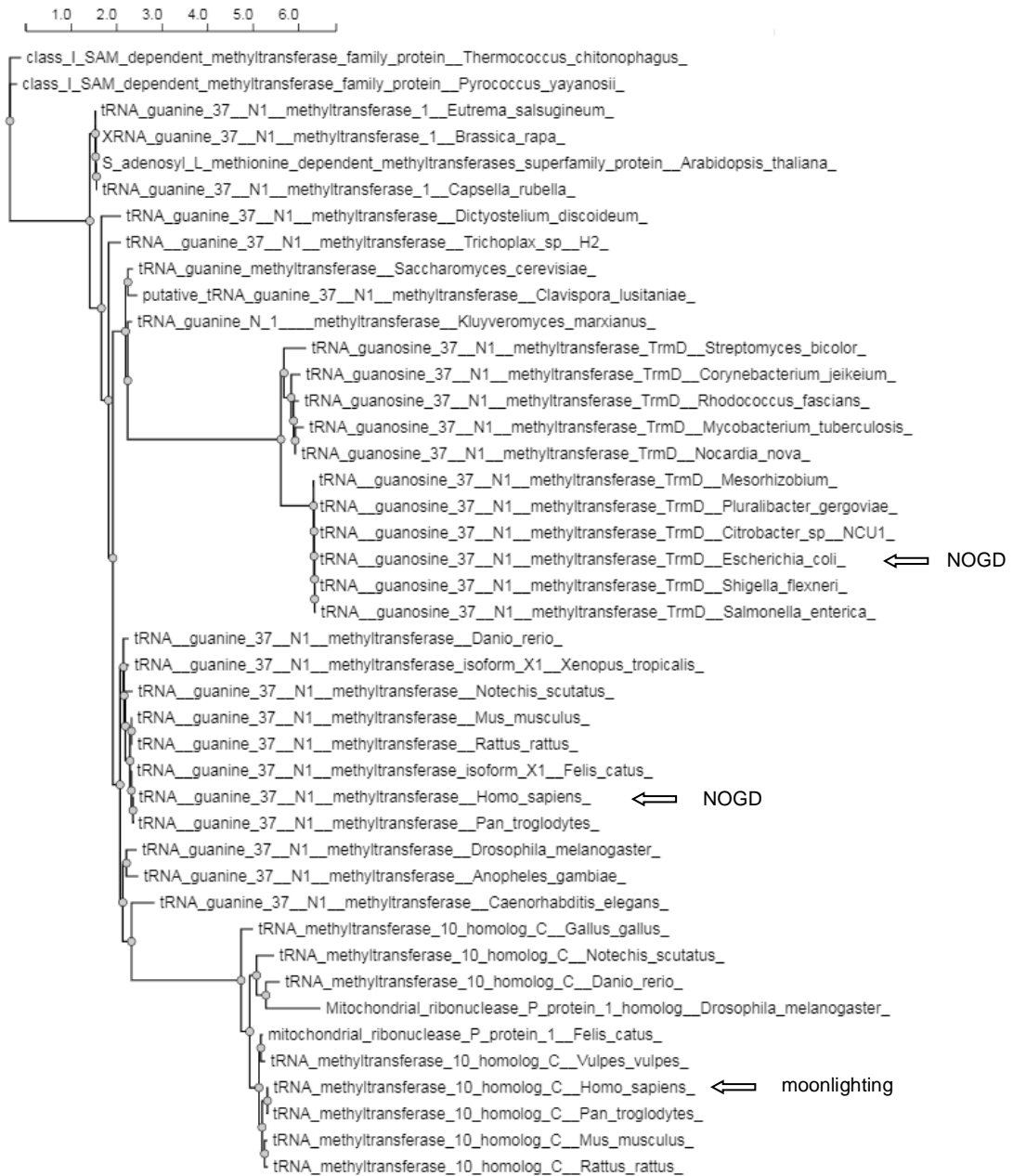
Árbol filogenético de proteínas con función α -N-acetilgalactosaminidasa.



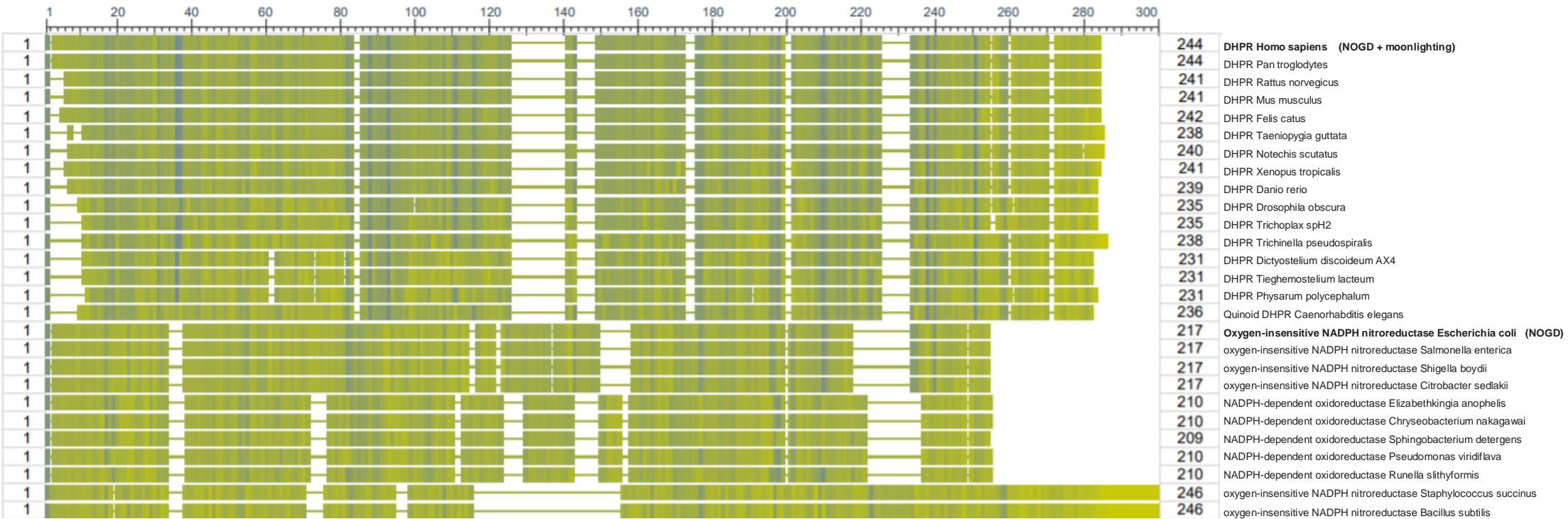
Árbol filogenético de proteínas con función endo-β-N-acetilglucosaminidasa



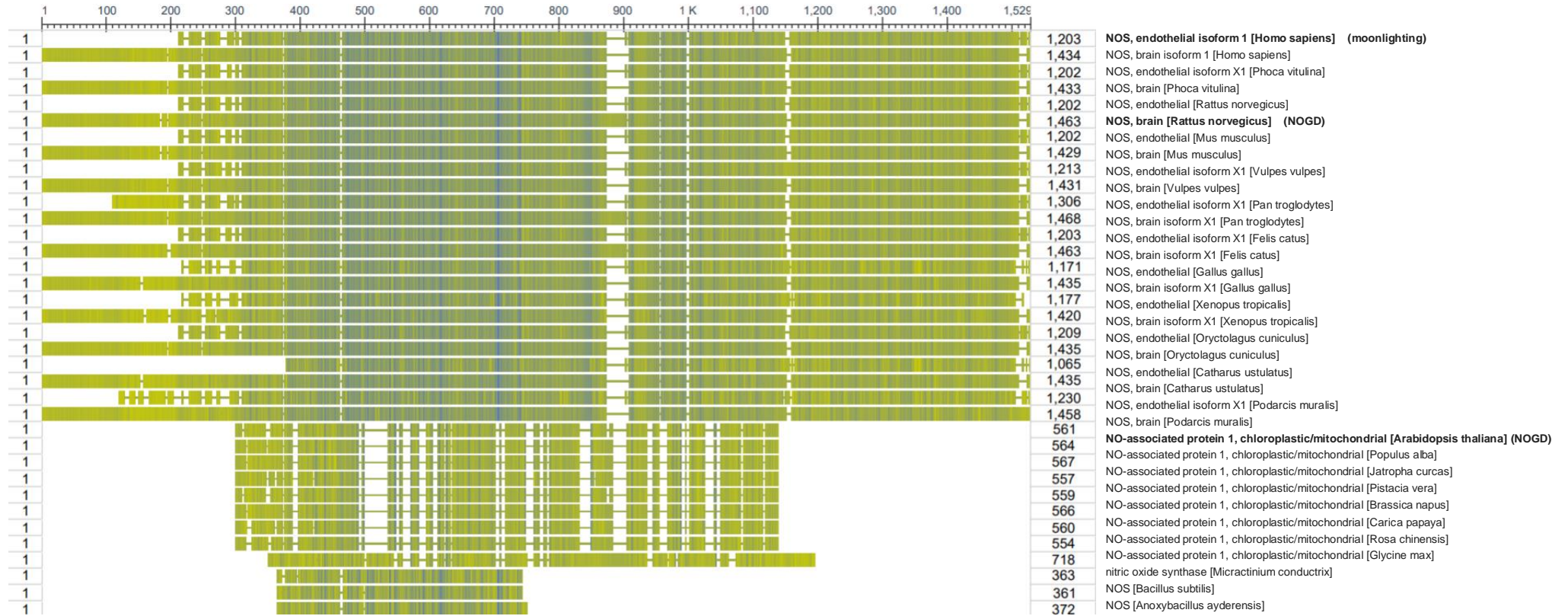
Árbol filogenético de proteínas con función ARNt metiltransferasa.



Alineamiento entre proteínas con función dihidropteridina reductasa



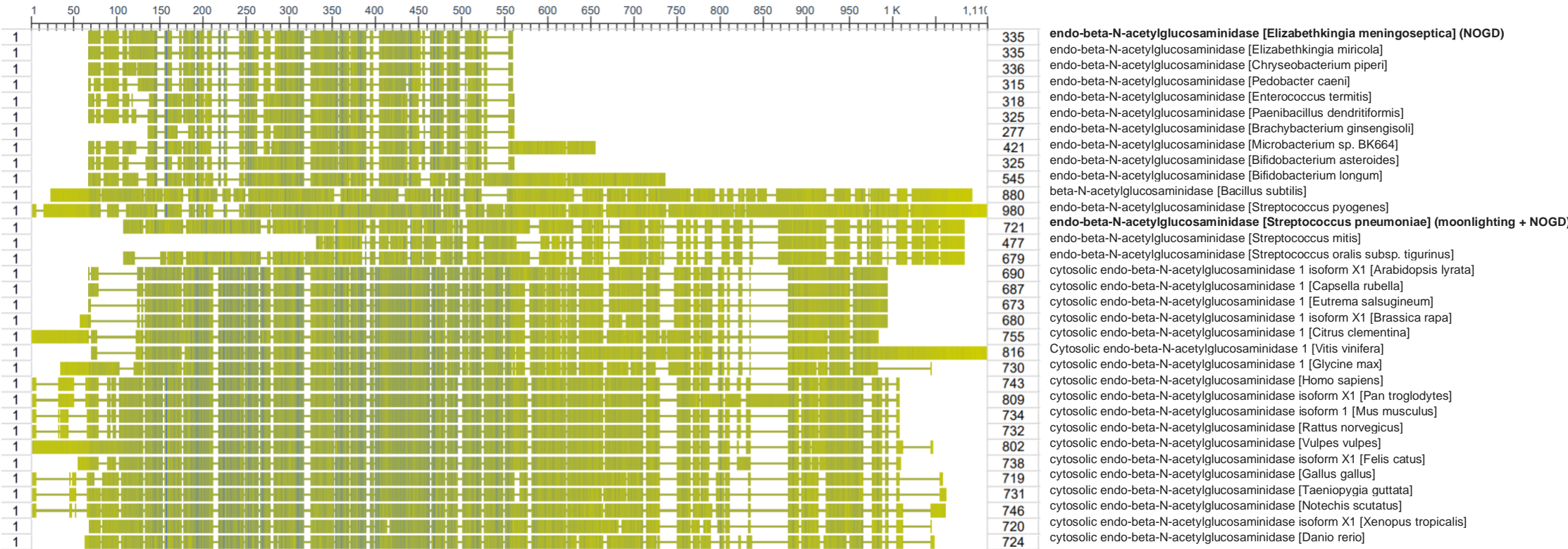
Alineamiento entre proteínas con función óxido nítrico sintasa



Alineamiento entre proteínas con función corismato mutasa



Alineamiento entre proteínas con función Endo-beta-N-acetylglucosaminidase



Alineamiento entre proteínas con función Mitochondrial RNasa P/tRNA (guanine-N1-)-methyltransferase



