



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

RECONSTRUCCIÓN DE LA HISTORIA EVOLUTIVA DE  
LA SUBFAMILIA DE ENZIMAS RELACIONADAS A LA  
N-ACETILNEURAMINATO LIASA (NAL) COMO HERRAMIENTA  
PARA EL DISEÑO DE NUEVAS ACTIVIDADES ENZIMÁTICAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A :

FERNANDA ARTEMISA ESPINOZA HERNÁNDEZ

Director de tesis:  
DR. BERNARDO PÉREZ ZAMORANO

2014



FACULTAD DE CIENCIAS  
UNAM



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno  
Espinoza  
Hernández  
Fernanda Artemisa  
53 06 04 18  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Biología  
306072679
2. Datos del tutor  
Dr.  
Bernardo  
Pérez  
Zamorano
3. Datos del sinodal 1  
Dr.  
Alejandro  
Sosa  
Peinado
4. Datos del sinodal 2  
Dr.  
Arturo Carlos II  
Becerra  
Bracho
5. Datos del sinodal 3  
Quím.  
Viviana  
Escobar  
Sánchez
6. Datos del sinodal 4  
Dr.  
León Patricio  
Martínez  
Castilla
7. Datos del trabajo escrito  
Reconstrucción de la historia evolutiva de la subfamilia de enzimas relacionadas a la N-acetilneuraminato liasa (NAL) como herramienta para el diseño de nuevas actividades enzimáticas  
76 p  
2014

*Dedico esta tesis con especial afecto...*

*...a mis padres*

*... y a mi novio, Gerardo.*

*“Para ser exitoso no tienes que hacer cosas extraordinarias.*

*Haz cosas ordinarias extraordinariamente bien”*

**Jim Rohn**

## *Agradecimientos*

A la Universidad Nacional Autónoma de México, la máxima casa de estudios, por abrirme las puertas para mi desarrollo intelectual y personal. A mi escuela, la Facultad de Ciencias, por haber sido testigo de mi formación académica. A todos mis profesores, que pusieron su granito de arena y sembraron en mí sus conocimientos que perdurarán para toda la vida.

A mi asesor, el Dr. Bernardo Pérez Zamorano, por brindarme un poco de su tiempo y por haber estado presente para resolver mis dudas. De igual manera, a mis sinodales, por haberse tomado el tiempo necesario para revisar mi trabajo y contribuir con sus valiosas observaciones.

Al Laboratorio de Físicoquímica e Ingeniería de Proteínas, por haberme brindado un espacio para poder llevar a cabo la elaboración de mi tesis de manera exitosa. En especial, al Dr. Alejandro Sosa Peinado. También, a mis compañeros del laboratorio, que me aclararon dudas existenciales, por sus palabras de aliento, por su valioso apoyo, por esos momentos tan divertidos y por esos deliciosos desayunos: Renán, Aarón, Andrés “Gansito”, Julio y Tania.

A mi compañera de laboratorio y amiga incondicional Marisol. Gracias por estar en los momentos más difíciles y por haberme apoyado en todo momento: por acompañarme a comprar dulces para el desestrés, por escuchar mis traumas académicos y personales, y simplemente, por tu presencia. Sin ti, mi estadía en el laboratorio no hubiera sido la misma.

A mis amigas de toda la vida, Dalia y Vanessa, que siempre me escucharon y me animaron a seguir adelante. A mis compañeros de la facultad que me acompañaron en distintas etapas de la carrera y que pasamos tiempos memorables juntos.

A mis padres, Ana y Manuel, por su paciencia, dedicación y constancia. Por haber estado durante toda mi vida, por nunca haberme fallado, por criarme de la mejor manera posible, por haberme formado así como soy: una persona capaz de superar cualquier límite, responsable, que odia la mediocridad y que puede llegar a ser muy ambiciosa si se lo propone. A mi hermano Ulises, por su constante fijación por preocuparse por lo que voy a hacer de mi vida, gracias a eso tuve momentos de reflexión profunda. A mis tías, hermanas de mi padre, que siempre estuvieron al pendiente de él debido a su accidente, y de mí, preguntando sobre mi avance académico y deseándome siempre lo mejor.

A Gerardo, mi novio, compañero y mejor amigo, que me hizo mejor persona y me acompañó en todo momento durante la carrera. Por esas tardes de pláticas académicas que ayudaron a despejar mi mente para pensar en mejores ideas. Por creer siempre en mí y ser pieza fundamental en mi vida, por su constante aliento y apoyo incondicional.

## Contenido general

<b>ÍNDICE DE FIGURAS .....</b>	<b>VI</b>
<b>ÍNDICE DE TABLAS .....</b>	<b>VII</b>
<b>GLOSARIO.....</b>	<b>VIII</b>
<b>RESUMEN.....</b>	<b>X</b>
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
1.1. EVOLUCIÓN BIOLÓGICA .....	1
1.2. FILOGENIAS Y EVOLUCIÓN MOLECULAR.....	2
1.3. EVOLUCIÓN Y DISEÑO DE PROTEÍNAS .....	3
1.4. RECONSTRUCCIÓN DE SECUENCIAS ANCESTRALES.....	5
1.5. ALDOLASAS CLASE I, UN EJEMPLO DE UNA SUPERFAMILIA MECANÍSTICAMENTE DIVERSA.....	6
1.5.1. Subfamilia de enzimas relacionadas a la N-acetilneuraminato liasa (NAL).....	8
1.5.1.1. Síntesis y degradación de ácido siálico: N-acetilneuraminato liasa (NAL).....	9
1.5.1.2. Biosíntesis de lisina: 4-hidroxi-tetrahidrodipicolinato sintasa (THDPS).....	10
1.5.1.3. Degradación de hidrocarburos poliaromáticos (PAHs): 4-(2-carboxifenil)-2-oxobut-3- enoato aldolasa (CPOEA) y trans-O-hidroxibencilidenopiruvato hidratasa-aldolasa (HBPHA) .....	11
1.5.1.4. Vías glicolíticas alternas: 2-dehidro-3-deoxi-fosfogluconato aldolasa (KDPGA) y 2- dehidro-3-deoxi-6-fosfogalactonato aldolasa (KDPGalA) .....	13
1.5.1.5. Degradación de hidroxiprolina: 4-hidroxi-2-oxoglutarato aldolasa (HOGA).....	15
1.6. IMPORTANCIA DE LA RECONSTRUCCIÓN ANCESTRAL EN LA INTERCONVERSIÓN DE ACTIVIDADES CATALÍTICAS ENTRE ENZIMAS DE LA SUBFAMILIA DE NAL.....	19
<b>2. JUSTIFICACIÓN Y OBJETIVOS .....</b>	<b>21</b>
2.1. JUSTIFICACIÓN.....	21
2.2. OBJETIVOS .....	21
<b>3. MATERIALES Y MÉTODOS .....</b>	<b>22</b>
3.1. BÚSQUEDA Y ALINEAMIENTO DE SECUENCIAS.....	22
3.2. CONSTRUCCIÓN Y VISUALIZACIÓN DEL ÁRBOL FILOGENÉTICO .....	24
3.3. RECONSTRUCCIÓN DE LAS SECUENCIAS ANCESTRALES .....	26
3.4. ANÁLISIS ESTRUCTURAL DE LAS PROTEÍNAS ANCESTRALES .....	27

<b>4. RESULTADOS.....</b>	<b>28</b>
4.1. NÚMERO DE SECUENCIAS Y DIVERSIDAD TAXONÓMICA .....	28
4.2. ÁRBOL FILOGENÉTICO DE LA SUBFAMILIA DE ENZIMAS RELACIONADAS A LA NAL .....	30
4.3. CARACTERIZACIÓN DE SECUENCIAS ANCESTRALES.....	43
4.3.1. Ancestros de THDPS y NAL .....	43
4.3.2. Ancestros de KDPGA/HOGA y KDPGalA.....	49
<b>5. DISCUSIÓN .....</b>	<b>53</b>
5.1. RELACIONES EVOLUTIVAS ENTRE LOS MIEMBROS DE LA SUBFAMILIA DE ENZIMAS RELACIONADAS A LA NAL .....	53
5.2. IDENTIFICACIÓN DE RESIDUOS DE AMINOÁCIDOS POR MEDIO DE LA RECONSTRUCCIÓN DE PROTEÍNAS ANCESTRALES.....	60
<b>6. CONCLUSIONES Y PERSPECTIVAS .....</b>	<b>61</b>
<b>REFERENCIAS.....</b>	<b>64</b>
<b>APÉNDICE 1: CLASES TAXONÓMICAS.....</b>	<b>74</b>
<b>APÉNDICE 2: ESTRUCTURAS TRIDIMENSIONALES .....</b>	<b>75</b>

## Índice de figuras

Figura 1. Estructura de un árbol filogenético .....	3
Figura 2. Mecanismo de la reacción aldólica.....	6
Figura 3. Mecanismo de activación del sustrato de acuerdo a las dos clases de aldolasas 7	
Figura 4. Diagrama de la estructura terciaria de una subunidad de la NAL .....	9
Figura 5. Ruptura aldólica reversible de Neu5Ac catalizada por la NAL.....	10
Figura 6. Condensación aldólica reversible de THDP catalizada por THDPS .....	10
Figura 7. Vías propuestas de la degradación del fenantreno y naftaleno .....	12
Figura 8. Reacciones catalíticas reversibles de CPOAE y HBPHA .....	13
Figura 9. Esquema general de la condensación aldólica realizada por KDPGA y KDPGaIA .....	13
Figura 10. Reacciones catalíticas reversibles realizadas por KDPGA y KDPGaIA .....	15
Figura 11. Reacción catalítica reversible llevada a cabo por HOGA .....	16
Figura 12. Alineamiento estructural del sitio activo de NAL de <i>E. coli</i> con el sitio activo de THDPS .....	20
Figura 13. Alineamiento múltiple de las aldolasas.....	24
Figura 14. Número de secuencias obtenidas de cada enzima de la subfamilia en cada dominio filogenético .....	28
Figura 15. Clases taxonómicas obtenidas por cada dominio filogenético y porcentaje de secuencias pertenecientes a cada clase .....	29
Figura 16. Árbol filogenético de la subfamilia de enzimas relacionadas a la NAL .....	34
Figura 17. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético .....	36
Figura 18. Árbol filogenético de las secuencias utilizadas para obtener los ancestros de NAL y THDPS.....	43
Figura 19. Alineamiento de las secuencias de THDPS y NAL de <i>E. coli</i> , y las secuencias de los nodos ancestrales .....	45
Figura 20. Residuos conservados del alineamiento entre THDPS y NAL de <i>E. coli</i> , y los nodos ancestrales.....	47
Figura 21. Alineamiento estructural de THDPS y NAL con los nodos ancestrales .....	48

Figura 22. Árbol filogenético de las secuencias utilizadas para obtener los ancestros de KDPGA, KDPGaIA y HOGA.....	49
Figura 23. Alineamiento de las secuencias de KDPGA/HOGA y KDPGaIA de <i>E. coli</i> , y las secuencias de los nodos ancestrales.....	51
Figura 24. Residuos conservados del alineamiento entre KDPGA/HOGA y KDPGaIA de <i>E. coli</i> , y los nodos ancestrales .....	51
Figura 25. Alineamiento estructural de KDPGA y KDPGaIA con los nodos ancestrales ...	52
Figura 26. Propuesta del surgimiento y diversificación de la subfamilia de enzimas relacionadas a la NAL.....	59

## Índice de tablas

Tabla 1. Estructuras tridimensionales y características principales de cada miembro de la subfamilia de enzimas relacionadas a la NAL .....	17
Tabla 2. Secuencias (y números de acceso) utilizadas para realizar la búsqueda bioinformática .....	22
Tabla 3. Matriz de distancias obtenida de la matriz original compuesta por un representante de cada enzima correspondiente al dominio Bacteria .....	26
Tabla 4. Secuencias pertenecientes a la subfamilia de enzimas relacionadas a la NAL con actividad enzimática comprobada experimentalmente.....	30
Tabla 5. Valores globales de RMSD y porcentaje de identidad entre pares de secuencias obtenidos de la comparación de THDPS y NAL de <i>E. coli</i> contra los nodos ancestrales..	47
Tabla 6. Valores globales de RMSD y porcentajes de identidad entre pares de secuencias obtenidos de la comparación de KDPGA y KDPGaIA de <i>E. coli</i> contra los nodos ancestrales.....	50

## Glosario

AAA	vía metabólica del ácido $\alpha$ -aminoadípico
aLRT	Approximate Likelihood-Ratio Test
BLAST	Basic Local Alignment Search Tool
C	número de duplicaciones condicionales (Notung)
cC	costo de duplicaciones condicionales (Notung)
cD	costo de duplicaciones (Notung)
cL	costo de pérdidas de genes (Notung)
CPOEA	4-(2-carboxifenil)-2-oxobut-O-3-enoato aldolasa
D	número de duplicaciones (Notung)
D/L	Duplication/Loss
DAP	vía metabólica del ácido diaminopimélico
DD	vía De Ley-Doudoroff
DHAP	dihidroxiacetona fosfato
DNA	ácido desoxirribonucleico
ED	vía metabólica Entner-Doudoroff
EMP	vía metabólica Embden-Meyerhoff-PRNAass
FastML	servidor que utiliza métodos de máxima verosimilitud para reconstruir secuencias ancestrales
FBPA	fructosa 1,6 bisfosfato aldolasa
G3P	gliceraldehído-3-fosfato
HBPHA	trans-O-hidroxibencilidenopiruvato hidratasa-aldolasa
HOG	4-hidroxi-2-oxoglutarato
HOGA	4-hidroxi-2-oxoglutarato aldolasa
iTOL	Interactive Tree of Life
kDa	kiloDalton
KDG	2-dehidro-3-deoxi-gluconato
KDGal	2-dehidro-3-deoxi-D-galactonato
KDPG	2-dehidro-3-deoxi-D-gluconato-6-fosfato

KDPGA	2-dehidro-3-deoxi-fosfogluconato aldolasa
KDPGal	2-dehidro-3-deoxi-D-galactonato-6-fosfato
KDPGalA	2-dehidro-3-deoxi-6-fosfogalactonato aldolasa
L	número de pérdidas de genes (Notung)
L-ASA	L-aspartato- $\beta$ -semialdehído
LG	modelo empírico mejorado del modelo WAG propuesto por Le y Gascuel que incorpora la variabilidad de las tasas evolutivas de los sitios en el alineamiento, estimado a partir de 3 912 alineamientos de Pfam
MAFFT	Multiple Aligment – fast Fourier transform
ManNAc	N-acetil-D-manosamina
NAL	N-acetilneuraminato liasa
NCBI	National Center for Biotechnology Information
Neu5Ac	ácido N-acetilneuramínico
PAHs	hidrocarburos poliaromáticos
PEP	fosfoenolpiruvato
PhyML	Phylogenetic Maximum Likelihood, software de reconstrucción de árboles filogenéticos que utiliza el método de máxima verosimilitud
Pyr4H2C	1-pirrolina-4-hidroxi-2-carboxilato
RNA	ácido ribonucleico
SH-like	Shimodaira-Hasegawa-like procedure
SPR	Subtree Pruning and Regrafting
THDP	(2S,4S)-4-hidroxi-2,3,4,5-tetrahidrodipicolinato
THDPS	4-hidroxi-tetrahidrodipicolinato sintasa
WAG	modelo empírico de la evolución de proteínas globulares propuesto por Whelan y Goldman, estimado a partir de 182 familias de proteínas usando un método de máxima verosimilitud que toma en cuenta las relaciones evolutivas dentro de cada familia

## Resumen

El rediseño de proteínas es una herramienta de aplicación industrial que ha permitido modificar la actividad catalítica de enzimas con la finalidad de obtener una función deseada; sin embargo, el enfoque horizontal que surge de realizar comparaciones estructurales, que se ha venido utilizando en los últimos años, ha generado varios problemas al trasplantar, por ejemplo, el sitio activo de una enzima a otra, debido a que los residuos de aminoácidos localizados en esas regiones no son los únicos implicados en la catálisis de dichas enzimas. Debido a lo anterior, una de las posibles estrategias para solucionar este problema surge de la idea de utilizar los experimentos que ya ha realizado la evolución biológica a lo largo de millones de años, es decir, realizar estudios verticales que nos permitan indagar en la historia evolutiva de las familias de enzimas.

La reconstrucción de secuencias ancestrales es un método que implica la obtención de ancestros hipotéticos, basado en el hecho de que esta enzima ancestral posee una actividad catalítica incipiente dada la secuencia primaria que la conforma, por lo que podría servir como guía para identificar los cambios, a nivel de residuos de aminoácidos, que se han generado a partir de ésta para la especialización funcional de las enzimas actuales.

En primera instancia, el presente trabajo consistió en reconstruir la historia evolutiva de la subfamilia de enzimas relacionadas a la N-acetilneuraminato liasa que utilizan piruvato para catalizar la ruptura reversible de enlaces carbono – carbono, concluyendo que posiblemente la enzima ancestral sufrió una duplicación génica que dio origen por un lado, a las enzimas tetraméricas THDPS, NAL, KDGA, CPOEA y HBPHA, y por el otro a las enzimas triméricas KPDPGA/HOGA y KDPGAIA. Posteriormente, se propuso una metodología que nos permitió generar ancestros hipotéticos de THDPS, NAL, KDPGA/HOGA y KDPGAIA; de esta manera, se identificaron algunos residuos de aminoácidos con función desconocida y que podrían estar implicados indirectamente en la actividad catalítica o en la estabilidad estructural. Finalmente, las comparaciones estructurales entre las enzimas actuales y los ancestros, permitieron describir algunas características sobresalientes de la estructura secundaria de cada nodo ancestral.

# 1. Introducción

## 1.1. Evolución biológica

La evolución biológica son los cambios en las características de un grupo de organismos o población que ocurren a través de varias generaciones. Las modificaciones se van produciendo a lo largo del tiempo en un mismo linaje de descendientes que tienen un ancestro en común (Ridley, 2004a).

La selección natural es, probablemente, la principal fuerza evolutiva que impulsa el cambio en los seres vivos (Futuyma, 2005). Esta teoría fue propuesta por Charles Darwin en 1859 en su libro “El Origen de las Especies”, y, básicamente, propone que la variabilidad de una característica heredable que poseen algunos individuos de una población puede conferir ventajas sobre otros, dependiendo el ambiente en el que se desarrollan; de esta manera, las características exitosas que van apareciendo, se van heredando, produciendo una acumulación de cambios que transforman a las especies. Un ejemplo claro es la resistencia a antibióticos que presentan bacterias patógenas: de la primera generación, unas cuantas contienen la variante de un gene que les permite sobrevivir al ataque del antibiótico, por lo que son capaces de reproducirse y heredar esa característica a la segunda generación.

Las características fenotípicas que se observan en una población se manifiestan debido a la interacción de un conjunto de alelos, que se van modificando como resultado de mutaciones en el material genético, y debido a la interacción con el ambiente (Smith, 1970). Las mutaciones pueden ocurrir de varias maneras: cuando la DNA o RNA polimerasa comete errores durante la replicación del material genético e introduce nucleótidos no complementarios (transiciones o transversiones), o bien, introduce o elimina nucleótidos que generan indeles (inserciones o deleciones); o por modificaciones a los nucleótidos debido a factores ambientales como la luz ultravioleta o agentes químicos (Vandamme, 2009).

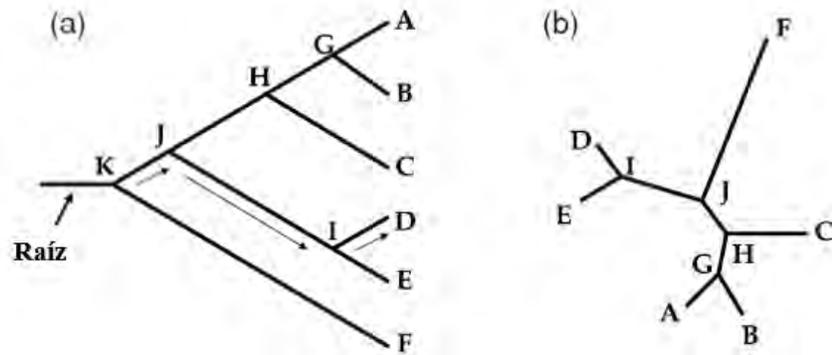
El resultado de las mutaciones en el material genético altera la información necesaria para la expresión de proteínas, y de acuerdo a la Teoría Neutralista de la Evolución Molecular que propuso Mooto Kimura en 1968, estos cambios en su mayoría son neutros y fijados azarosamente en una población por deriva génica; sin embargo, hay mutaciones que pueden resultar favorables provocando un aumento en su frecuencia alélica en la población debido a la selección natural (Ridley, 2004b).

Las fuerzas evolutivas que se encargan de fijar los caracteres en una población son la selección natural y la deriva génica aunque, por otro lado, se encuentran los mecanismos que generan la variabilidad genotípica en las poblaciones. Estos mecanismos son, entre otros: 1) la **recombinación genética**, que describe el intercambio genético de partes de alelos en una sola hebra de DNA; 2) la **duplicación génica**, donde se genera una copia de un gen que puede acumular mutaciones dando origen a un nuevo gen funcional; y 3) la **transferencia horizontal de genes**, que es un evento frecuente en bacterias.

## **1.2. Filogenias y evolución molecular**

La herramienta principal de los biólogos evolutivos para establecer las relaciones evolutivas entre especies es la filogenia. Desde el punto de vista clásico, el uso de caracteres morfológicos ha servido para este fin; sin embargo, la utilización de datos moleculares, es decir, secuencias de nucleótidos y aminoácidos, ha ido ganando terreno a la hora de realizar estudios evolutivos (Vandamme, 2009). La evolución molecular es un nuevo campo de investigación que ha sido desarrollado durante las últimas décadas para analizar los procesos evolutivos de los genes y proteínas comparando sus propiedades y las secuencias de varios organismos (Futuyma, 2005). En general, las filogenias asumen que el grupo de secuencias analizadas son homólogas, es decir, que comparten un ancestro en común.

Una filogenia o árbol filogenético es una representación que describe las relaciones evolutivas entre especies, genes, poblaciones o individuos. Está formado por partes que recuerdan a un árbol común: ramas, nodos, hojas y, en algunos casos, una raíz (Figura 1). Los nodos externos u hojas representan a los organismos existentes, mientras que los nodos internos representan los ancestros extintos de dichos organismos. La raíz es el ancestro de todos los organismos analizados en un árbol dado, y si el árbol no la tiene, no se puede establecer una dirección del proceso evolutivo (Z. Yang, 2006a). El largo de las ramas representa el número de sustituciones de nucleótidos o aminoácidos, y su patrón, es decir, el orden de los nodos, se conoce como la topología del árbol (Vandamme, 2009).



**Figura 1. Estructura de un árbol filogenético. Las letras indican los nodos del árbol conectados por ramas: A, B, C, D, E, y F son los nodos externos o los organismos actuales; G, H, I, J, y K (nodo raíz) son los nodos internos o los organismos ancestrales hipotéticos. a) Las flechas indican la dirección de la evolución; b) Árbol sin raíz. Vandamme, A. M. (2009). Basic concepts of molecular evolution [Imagen].**

Para reconstruir una filogenia se utilizan métodos basados en distancia, o bien, métodos basados en caracteres. En los métodos basados en distancia se calcula una matriz a partir de la distancia filogenética obtenida por la comparación de cada par de secuencias; posteriormente esta matriz es utilizada para construir el árbol, a través de un algoritmo (Z. Yang, 2006a).

Los métodos basados en caracteres comparan todas las secuencias del alineamiento considerando un carácter a la vez, es decir, un sitio en el alineamiento, al mismo tiempo que calculan un puntaje para cada árbol. Estos métodos incluyen la máxima parsimonia, la máxima verosimilitud y los métodos de inferencia Bayesiana, que generan un árbol de inicio por medio de un algoritmo para después mejorar el puntaje a través de rearrreglos locales (Z. Yang & Rannala, 2012).

### 1.3. Evolución y diseño de proteínas

Al llevar a cabo comparaciones estructurales entre proteínas que realizan la misma función en diversos organismos, se observa que presentan homología estructural a pesar de que su secuencia difiere en determinada cantidad de aminoácidos.

A través del campo de estudio de la evolución de proteínas se pueden conocer: 1) los factores genéticos implicados en la evolución de estructuras de enzimas homólogas; 2) los factores que dan lugar a diversas estructuras manifestando nuevas especificidades y funciones que intervienen de manera importante en el desarrollo de nuevas vías metabólicas; y, 3) las funciones específicas de residuos de aminoácidos en el mecanismo de acción y especificidad de las enzimas en relación a su conformación (Smith, 1970).

El estudio de estos factores ha permitido re-diseñar algunas enzimas para que catalicen nuevas reacciones. Las enzimas son de particular importancia en la industria debido a que son utilizadas como catalizadores químicos, siendo capaces de catalizar una gran gama de reacciones químicas con cierta especificidad y selectividad (Dwyer, Looger, & Hellinga, 2004).

El diseño de proteínas se refiere a la habilidad de modificar la estructura de una proteína para obtener la función deseada. Los métodos de diseño pueden ser categorizados en dos estrategias. La evolución dirigida es una herramienta que simula la evolución darwiniana en un tubo de ensaye e involucra rondas iterativas de creación de diversidad genética seguida por un proceso de selección. Por el otro lado, se encuentra el diseño racional que es un proceso que utiliza conocimientos previos acerca de la enzima en cuestión, tales como su estructura o secuencia; de esta manera, se seleccionan aminoácidos para llevar a cabo mutaciones que podrían afectar la actividad de la enzima (Z. Chen & Zhao, 2006).

Los enfoques tradicionales que se han utilizado para diseñar proteínas no han sido tan exitosos como se esperaba. La idea principal fue tomada de la propia evolución natural al estudiar con detenimiento la homología estructural de las diversas familias de enzimas que catalizan reacciones diferentes, es decir, al conocer las estructuras nativas y el mecanismo de acción de las enzimas, se tendría la capacidad para modificarlas para intercambiar su actividad catalítica entre distintos miembros de la propia familia, o bien, a partir de predicciones computacionales, crear nuevos andamios que permitan la actividad catalítica deseada. Sin embargo, la identidad que comparten los miembros de una sola familia puede ser menor al 40%, lo que sugiere que para obtener la nueva actividad podrían ser requeridas muchas sustituciones; aunque, por otro lado, se han podido generar actividades nuevas con una sola mutación (Gerlt & Babbitt, 2009; O'Brien & Herschlag, 1999).

Un concepto que ha sido de utilidad para el rediseño es la promiscuidad catalítica que presentan las enzimas. Esta propiedad les confiere el potencial de realizar reacciones catalíticas distintas a las que usualmente llevan a cabo, por lo que ha servido como punto de inicio para la creación de nuevas actividades, mejorando las actividades incipientes (Hult & Berglund, 2007).

Debido al poco éxito que han tenido estos enfoques en el diseño de proteínas, ya que algunas de las sustituciones de residuos de aminoácidos pueden desestabilizar la estructura, y que el método para escoger a la enzima “progenitora” es de ensayo y error, se han buscado alternativas como la reconstrucción de ancestros a partir de filogenias que proporcionen la estabilidad necesaria para un rediseño robusto.

#### **1.4. Reconstrucción de secuencias ancestrales**

Uno de los objetivos principales del estudio de la evolución molecular es obtener información sobre la estructura y función de las proteínas que han existido a lo largo del tiempo utilizando como herramienta la reconstrucción de secuencias ancestrales (Chang, Ugalde, & Matz, 2005).

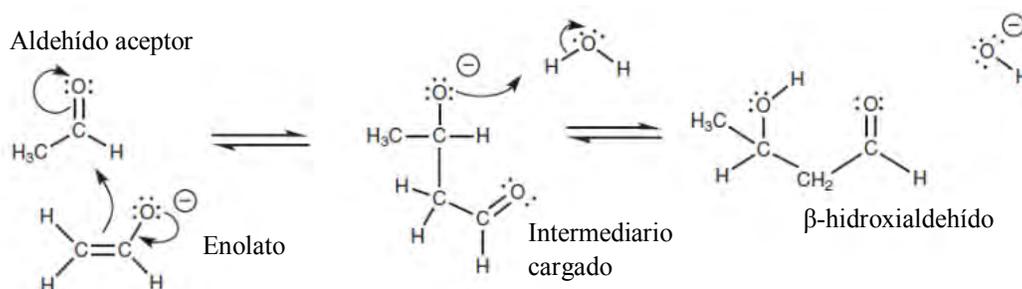
El método de reconstrucción de ancestros involucra inferir las secuencias proteicas de los nodos internos del árbol filogenético para, posteriormente, sintetizarlas, subclonaras, expresarlas y purificarlas para caracterizar su estabilidad y función. La primera técnica para reconstruir las secuencias ancestrales fue utilizar métodos de consenso que consistían simplemente en muestrear las secuencias escogidas y asumir que el estado más frecuente de cada posición correspondía al estado ancestral; sin embargo, la inferencia de las secuencias ancestrales dependía de cuáles secuencias eran elegidas desde un principio, por lo cual no era considerado un procedimiento robusto (Thornton, 2004).

En la actualidad existen dos métodos filogenéticos que permiten la inferencia de secuencias ancestrales. Los métodos de parsimonia minimizan la cantidad de cambios evolutivos a lo largo de las ramas del árbol filogenético, asumiendo tasas de cambio lentas y constantes. La limitación principal de estos métodos es que entre mayor sea la diferencia entre las secuencias, la inferencia es menos eficaz; además, debido a que minimizan los cambios de las ramas, no toman en cuenta las tasas desiguales de sustitución tanto de nucleótidos como de aminoácidos. Por otro lado, los métodos de máxima verosimilitud y bayesianos permiten el acoplamiento de un modelo evolutivo que mejor describa a los datos analizados, incluyendo tasas de sustitución variables de nucleótidos y aminoácidos que resultan en un enfoque más realista; además, la inferencia se puede calificar por un puntaje probabilístico. Los tipos de datos que emplean los métodos de máxima verosimilitud y los bayesianos se pueden clasificar en tres grupos: nucleótidos, aminoácidos y codones.

La reconstrucción de secuencias ancestrales permite poner a prueba tanto la hipótesis de la especialización a partir de un rol general, así como la pérdida de una actividad ancestral específica (Chang & Donoghue, 2000). Por ejemplo, Jermann y colaboradores demostraron que los ancestros de las ribonucleasas de algunos rumiantes (artiodáctilos) poseían, además de la actividad catalítica normal sobre RNA unicatenario (presente en las ribonucleasas digestivas de estos mamíferos), una actividad catalítica incrementada sobre RNA de doble cadena, al igual que la actividad catalítica de las ribonucleasas no digestivas de estos y otros mamíferos; de esta manera, se concluyó que las ribonucleasas digestivas se originaron a partir de ancestros que no participaban en la digestión, y se especializaron perdiendo la actividad sobre el RNA bicatenario (Jermann, Opitz, Stackhouse, & Benner, 1995).

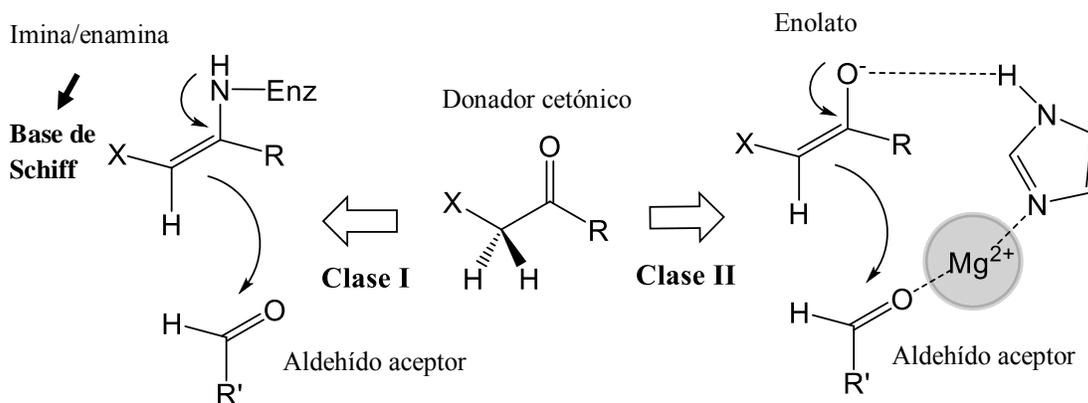
### 1.5. Aldolasas clase I, un ejemplo de una superfamilia mecanísticamente diversa

Las aldolasas son un grupo de liasas que catalizan la formación usualmente reversible de enlaces carbono (C-C), adicionando estereoselectivamente un donador cetónico a un aldehído aceptor (Takayama, McGarvey, & Wong, 1997), creando nuevos centros estereogénicos. Estas enzimas han sido utilizadas en la industria química para la síntesis de nuevos estereoisómeros (síntesis asimétrica) que son evaluados para su utilización farmacéutica, principalmente. El uso de enzimas en el campo de la síntesis orgánica es una alternativa a la metodología química clásica debido a su alta eficiencia catalítica y que permiten un control quiral completo (Fessner & Walter, 1997). La reacción aldólica involucra tres pasos generales: 1) la activación del donador cetónico (enolato o análogo); 2) la adición nucleofílica al aldehído aceptor para formar un intermediario cargado; y, 3) la protonación del intermediario cargado (Figura 2). El producto final de esta reacción son compuestos  $\beta$ -hidroxicarbonilos, aunque una deshidratación puede ocurrir después de la reacción aldólica obteniendo productos carbonilos  $\alpha,\beta$ -insaturados (Patrick, 2004).



**Figura 2. Mecanismo de la reacción aldólica. El ion enolato realiza un ataque nucleofílico al aldehído aceptor formando el enlace C-C; posteriormente, el intermediario cargado es protonado por una molécula de agua para obtener el producto final.** Patrick, G. L. (2004). Aldehydes and ketones [Imagen].

Las aldolasas se pueden dividir en cuatro grupos principales de acuerdo al sustrato que actúa como componente nucleofílico en la reacción catalítica, adicionando el grupo cetona al grupo carbonilo del aldehído: las liasas dependientes de piruvato o fosfoenolpiruvato (PEP sintasas); las liasas dependientes de dihidroxiacetona fosfato (DHAP liasas); las liasas dependientes de acetaldehído; y las liasas dependientes de glicina (Takayama et al., 1997). Por otro lado, se clasifican en dos clases con base en el mecanismo de activación de los sustratos que ocurre por medio de una desprotonación estereoespecífica: las aldolasas clase I unen sus sustratos covalentemente al residuo de lisina del sitio activo por medio de una base de Schiff para comenzar a formar o romper el enlace vía la formación de imina/enamina; por otro lado, las aldolasas clase II utilizan iones metálicos que facilitan la desprotonación del donador cetónico formando un ion enolato (Figura 3) (Fessner & Walter, 1997).



**Figura 3. Mecanismo de activación del sustrato de acuerdo a las dos clases de aldolasas. La activación puede ser formando una base de Schiff (clase I) o utilizando un ion metálico (clase II).** Fessner, W. D., & Walter, C. (1997). *Enzymatic C-C bond formation in asymmetric synthesis* [Imagen].

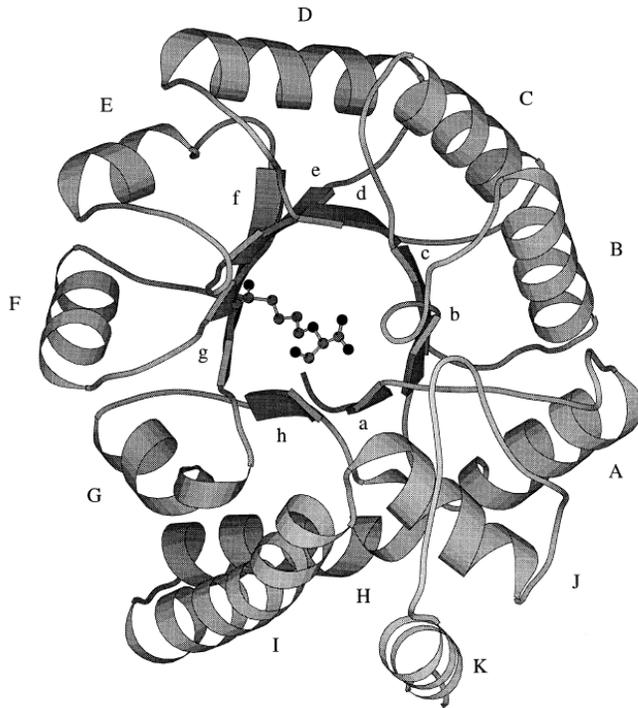
Las superfamilias de enzimas mecanísticamente diversas son grupos de proteínas homólogas que conservan un atributo mecanístico, como una reacción parcial específica, la formación de un intermediario, o un estado de transición, para catalizar diversas reacciones con diferentes sustratos a distintas especificidades. Los integrantes de estas superfamilias usualmente presentan un bajo nivel de identidad en un alineamiento de secuencias; los elementos conservados que conforman el sitio activo realizan la función mecanística común de todos los miembros; y exhiben el mismo plegamiento general (Glasner, Gerlt, & Babbitt, 2006). Las enzimas que forman parte de la familia de aldolasas clase I comprenden una superfamilia mecanísticamente diversa que se define por la

formación de una base de Schiff en su reacción de catálisis y por compartir la estructura de barriles  $(\beta/\alpha)_8$ . La primera enzima en la que se describió la formación de la base de Schiff fue la fructosa 1,6 bisfosfato aldolasa (FBPA) que se encuentra ampliamente distribuida en eucariontes y arqueas (Choi et al., 2006).

#### **1.5.1. Subfamilia de enzimas relacionadas a la N-acetilneuraminato liasa (NAL)**

Dentro de la superfamilia de aldolasas clase I se encuentra la subfamilia de enzimas relacionadas a la N-acetilneuraminato liasa (NAL) que comprende aldolasas de clase I dependientes de piruvato. Sus integrantes comparten tanto la estructura tridimensional de barriles  $(\beta/\alpha)_8$  y la formación de una base de Schiff al llevar a cabo su reacción catalítica participando en diferentes vías metabólicas (Barbosa et al., 2000; Lawrence et al., 1997). Los miembros identificados de esta subfamilia son: la N-acetilneuraminato liasa (NAL; EC 4.1.3.3), la 4-hidroxi-tetrahidrodipicolinato sintasa (THDPS; EC 4.3.3.7), la 4-(2-carboxifenil)-2-oxobut-3-enoato aldolasa (CPOEA; EC 4.1.2.34), la trans-O-hidroxibencilidenopiruvato hidratasa-aldolasa (HBPHA; EC 4.1.2.45), la 2-dehidro-3-deoxi-fosfogluconato aldolasa (KDPGA; EC 4.1.2.14), la 2-dehidro-3-deoxi-6-fosfogalactonato aldolasa (KDPGalA; EC 4.1.2.21) y la 4-hidroxi-2-oxoglutarato aldolasa (HOGA; E.C.4.1.3.16). La enzima D-5-ceto-4-deoxiglucarato dehidratasa (KDGDH; 4.2.1.41) que Barbosa *et al.* coloca en esta subfamilia no se consideró para este trabajo debido a que no utiliza piruvato.

La estructura terciaria de las subunidades de los miembros de esta subfamilia está conformada por dos dominios principales: el dominio N-terminal es un barril  $(\beta/\alpha)_8$ , que comprende ocho hebras  $\beta$  en el interior y ocho hélices  $\alpha$  en el exterior; y el dominio C-terminal que consiste en una cantidad variable de hélices  $\alpha$  (Figura 4) (Lawrence et al., 1997). Todas las aldolasas que forman una base de Schiff contienen un residuo de lisina específico conservado que se encuentra en el “piso” del sitio activo localizado en la sexta hebra  $\beta$  (hebra f en la Figura 4) del barril  $(\beta/\alpha)_8$  donde sobresale el grupo  $\epsilon$ -amino que reacciona con el grupo  $\alpha$ -ceto del piruvato. Además, hay otros residuos conservados que forman el motivo GxxGE localizado en la hebra b, asociado con la unión del grupo  $\alpha$ -ceto del piruvato, y los residuos encargados de unir el dominio C-terminal con el barril  $(\beta/\alpha)_8$  (Barbosa et al., 2000). La Tabla 1 (pág. 17) muestra las estructuras tridimensionales disponibles de los miembros de la subfamilia, así como el resumen de sus características principales. Las estructuras de CPOEA y HBPHA fueron modeladas con RaptorX (Källberg et al., 2012).

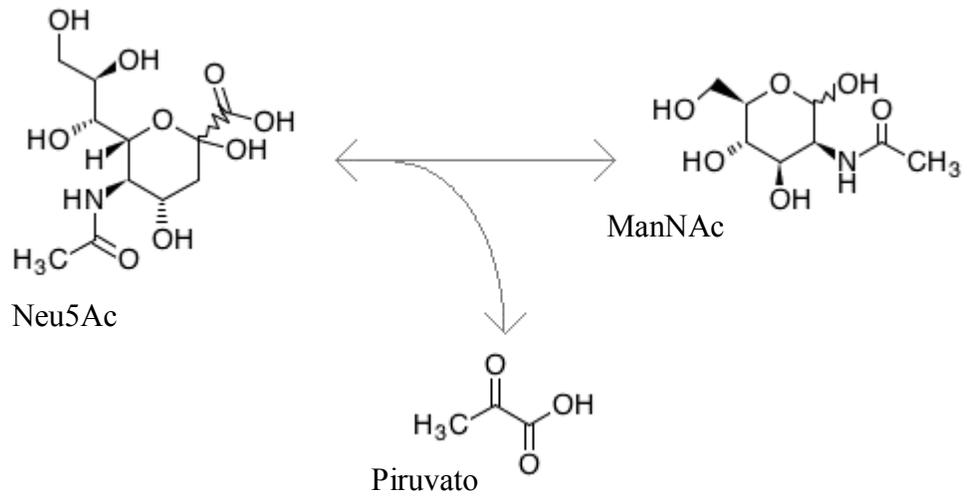


**Figura 4. Diagrama de la estructura terciaria de una subunidad de la N-acetilneuraminato liasa de *Escherichia coli*. Las hebras  $\beta$  que se encuentran en el interior del barril ( $\beta/\alpha$ )<sub>8</sub> son denotadas como a, b, c, d, e, f, g, h. Las hélices  $\alpha$  que se encuentran en el exterior del barril son denotadas como A, B, C, D, E, F, G, H, mientras que las hélices extra que se encuentran en el dominio C-terminal son denotadas como I, J, K. La lisina catalítica (Lys165) se encuentra en la hebra f, donde se muestra en complejo con hidroxipiruvato. Lawrence, M. C., Barbosa, J. A., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C., & Marcuccio, S. M. (1997). Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminatylase [Imagen].**

#### 1.5.1.1. Síntesis y degradación de ácido siálico: N-acetilneuraminato liasa (NAL)

NAL cataliza la ruptura reversible del ácido N-acetilneuramínico (Neu5Ac), un tipo de ácido siálico, a piruvato y N-acetil-D-manosamina (ManNAc) (Figura 5) (Izard, Lawrence, Malby, Lilley, & Colman, 1994; Kanehisa, Goto, Sato, Furumichi, & Tanabe, 2012; Kanehisa, 2000). Es un homotetrámero que se encarga de regular la concentración intracelular de este tipo de ácido siálico y es la primera enzima en la vía de degradación de este azúcar en bacterias. El residuo de lisina catalítico que forma la base de Schiff con el piruvato es la Lys165.

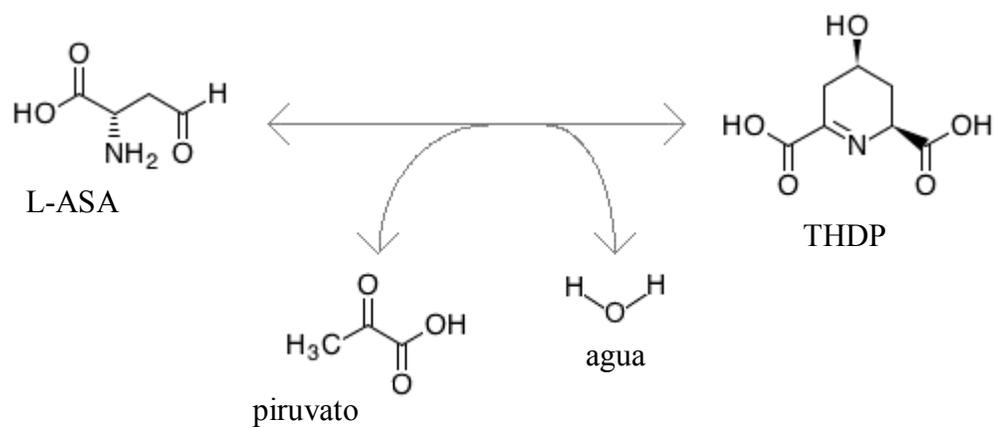
El Neu5Ac es un ácido siálico que forma parte de las cadenas de polisacáridos de glicoproteínas y glicolípidos de las cubiertas celulares de tejidos animales (Aisaka & Uwajima, 1986), además es utilizado por bacterias patógenas, como *Clostridium perfringens*, *Corynebacterium diphtheriae*, *Pasteurella multocida* y *Vibrio cholerae*, para evadir la respuesta inmune del hospedero (Aisaka, Igarashi, Yamaguchi, & Uwajima, 1991). Por otro lado, algunas bacterias como *Escherichia coli* pueden utilizar este ceto azúcar como única fuente de carbono (Vimr & Troy, 1985).



**Figura 5. Ruptura aldólica reversible de Neu5Ac catalizada por la NAL.** Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes [Imagen].

1.5.1.2. Biosíntesis de lisina: 4-hidroxi-tetrahidropicolinato sintasa (THDPS)

THDPS es la primera enzima en la ruta de biosíntesis de la lisina conocida como vía del ácido diaminopimélico (DAP), que se presenta en procariontes, algunos protozoarios y plantas vasculares. Su reacción catalítica consiste en formar (2S,4S)-4-hidroxi-2,3,4,5-tetrahidropicolinato (THDP) a partir de L-aspartato-β-semialdehído (L-ASA) y piruvato como sustratos (Figura 6) (Blickling et al., 1997; Kanehisa et al., 2012; Kanehisa, 2000). La estructura nativa es un homotetrámero. El residuo de lisina catalítico es la Lys161 (Mirwaldt, Korndörfer, & Huber, 1995).



**Figura 6. Condensación aldólica reversible de THDP catalizada por THDPS.** Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes [Imagen].

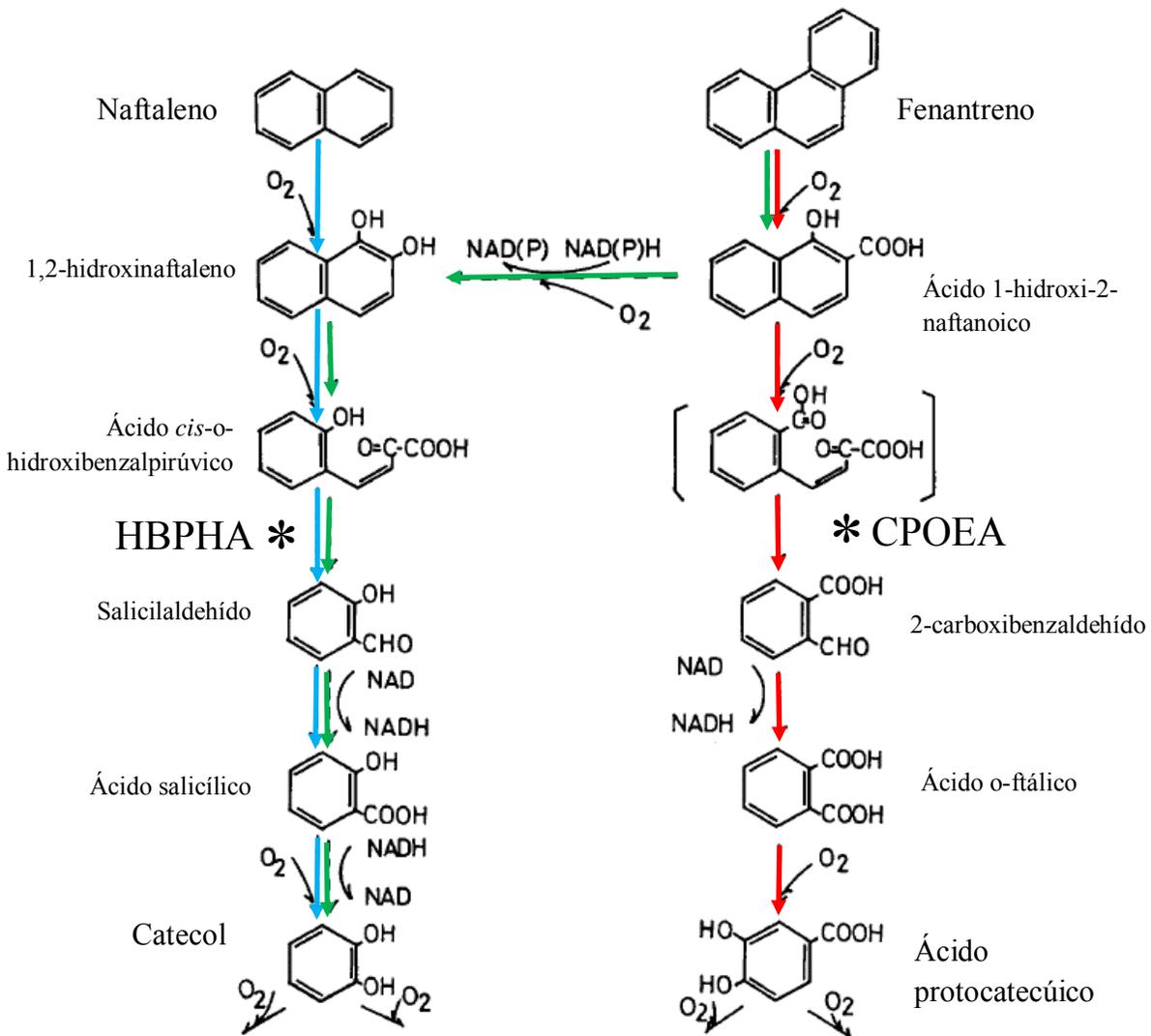
### 1.5.1.3. Degradación de hidrocarburos poliaromáticos (PAHs): 4-(2-carboxifenil)-2-oxobut-3-enoato aldolasa (CPOEA) y trans-O-hidroxibencilidenopiruvato hidratasa-aldolasa (HBPHA)

Los hidrocarburos poliaromáticos (PAHs) son hidrocarburos conformados por dos o más anillos aromáticos que son parte de los contaminantes ubicuos productos de la industria del petróleo, la quema de combustible fósil y la incineración de desechos (Seo, Keum, & Li, 2012). Estos compuestos son degradados en condiciones aeróbicas por bacterias y algunos hongos, donde ocurre un mecanismo de oxidación debido a la adición de O<sub>2</sub>, así como en condiciones anaeróbicas, donde los PAHs se degradan gracias a la reducción de los anillos. En ambos casos, el oxígeno debe ser añadido para que los PAHs puedan ser catabolizados. Bacterias sulfato-reductoras y desnitrificantes adicionan una molécula de CO<sub>2</sub> al anillo para formar derivados de ácido carboxílico; mientras que bajo condiciones aeróbicas, el oxígeno es añadido en etapas tempranas de la vía por oxigenasas que después de varias reacciones catalíticas producen catecol o protocatehuato; finalmente, son degradados a moléculas que pueden entrar al ciclo de Krebs (Madigan, Martinko, Stahl, & Clark, 2012).

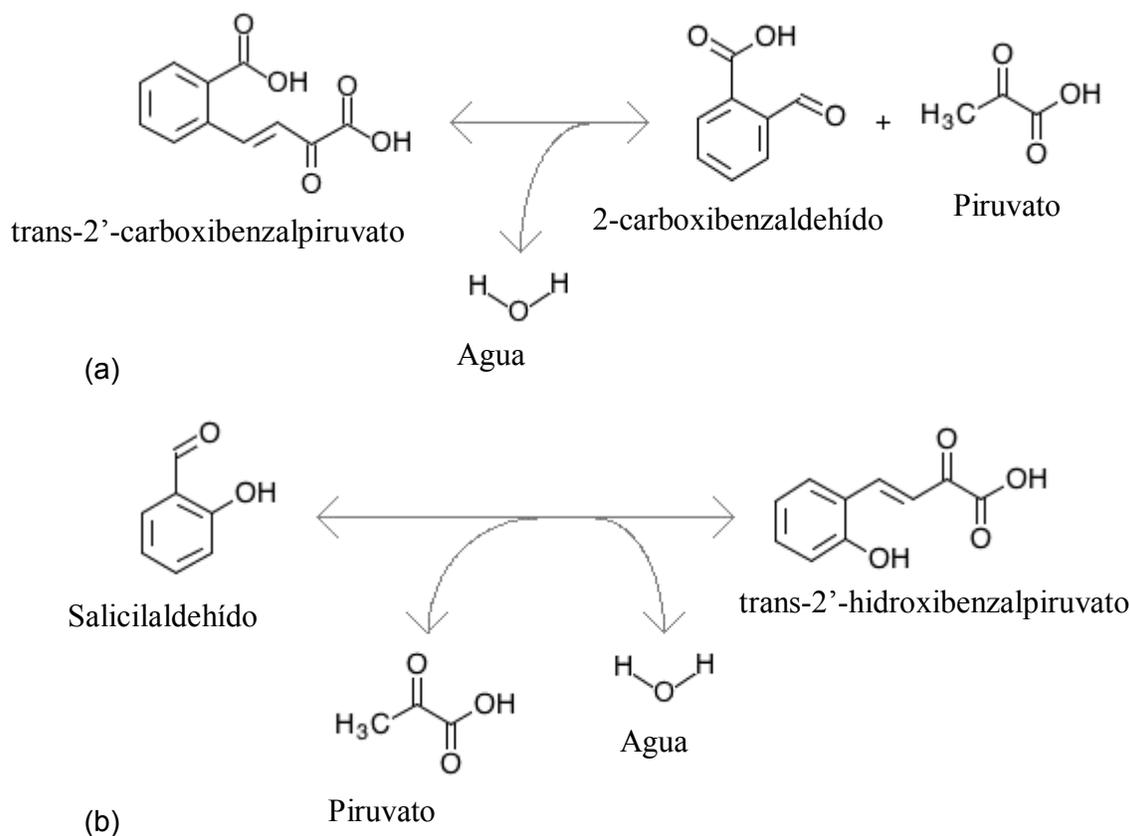
Las aldolasas CPOEA y HBPHA son enzimas involucradas en el catabolismo aeróbico del fenantreno y del naftaleno, respectivamente. Fueron caracterizadas como trímeros compuestos por subunidades de alrededor de 38 kDa (Iwabuchi & Harayama, 1998; Kuhm, Knackmuss, & Stolz, 1993), aunque no hay estructuras tridimensionales disponibles.

Las vías catabólicas del fenantreno y el naftaleno se encuentran relacionadas entre sí (Figura 7). Se ha propuesto que el fenantreno puede ser degradado en sus últimas etapas mediante reacciones pertenecientes a la vía degradativa del naftaleno; sin embargo, existen bacterias que sólo pueden crecer en alguno de los dos compuestos, por lo que el fenantreno se degrada utilizando su propia vía (H. Kiyohara & Nagao, 1978). *Aeromonas* sp. es una bacteria capaz de crecer en ambos sustratos (Hohzoh Kiyohara, Nagao, & Nomi, 1976).

CPOAE está presente en bacterias Gram-negativas del suelo que son capaces de utilizar el fenantreno como única fuente de carbono y energía, y cataliza la reacción reversible de (3E)-4-(2-carboxifenil)-2-oxobut-3-enoato (trans-2'-carboxibenzalpiruvato) a 2-formilbenzoato (2-carboxibenzaldehído) y piruvato (Figura 8a) (Barnsley, 1983; Kanehisa et al., 2012; Kanehisa, 2000). Por otro lado, HBPHA cataliza la ruptura de trans-2'-hidroxibenzalpiruvato a salicilaldehído y piruvato (Figura 8b), que es parte del metabolismo de degradación del naftaleno en bacterias (Kuhm et al., 1993).



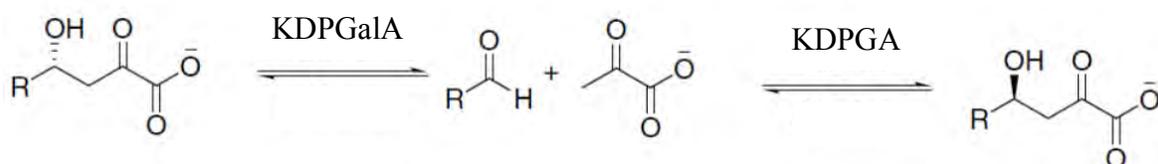
**Figura 7. Vías propuestas de la degradación del fenantreno y naftaleno. Las flechas azules representan la vía catabólica del naftaleno; las flechas rojas representan la vía catabólica del fenantreno; las flechas verdes representan la vía degradativa que puede tener el fenantreno siguiendo la vía del naftaleno. Kiyohara, H., & Nagao, K. (1978). The Catabolism of Phenanthrene and Naphthalene by Bacteria [Imagen].**



**Figura 8. Reacciones catalíticas reversibles de (a) CPOAE y (b) HBPFA. Ambas reacciones son precedidas por reacciones catalizadas por dioxigenasas en sus respectivas vías metabólicas.** Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes [Imagen].

#### 1.5.1.4. Vías glicolíticas alternas: 2-dehidro-3-deoxi-fosfogluconato aldolasa (KDPGA) y 2-dehidro-3-deoxi-6-fosfogalactonato aldolasa (KDPGalA)

Estas dos aldolasas catalizan una reacción idéntica que difiere en la especificidad por el sustrato en la configuración de un solo estereocentro (Figura 9) (Walters et al., 2008). Ambas se encuentran presentes en vías glicolíticas alternas a la canónica vía Embden-Meyerhoff-PRNAass (EMP).



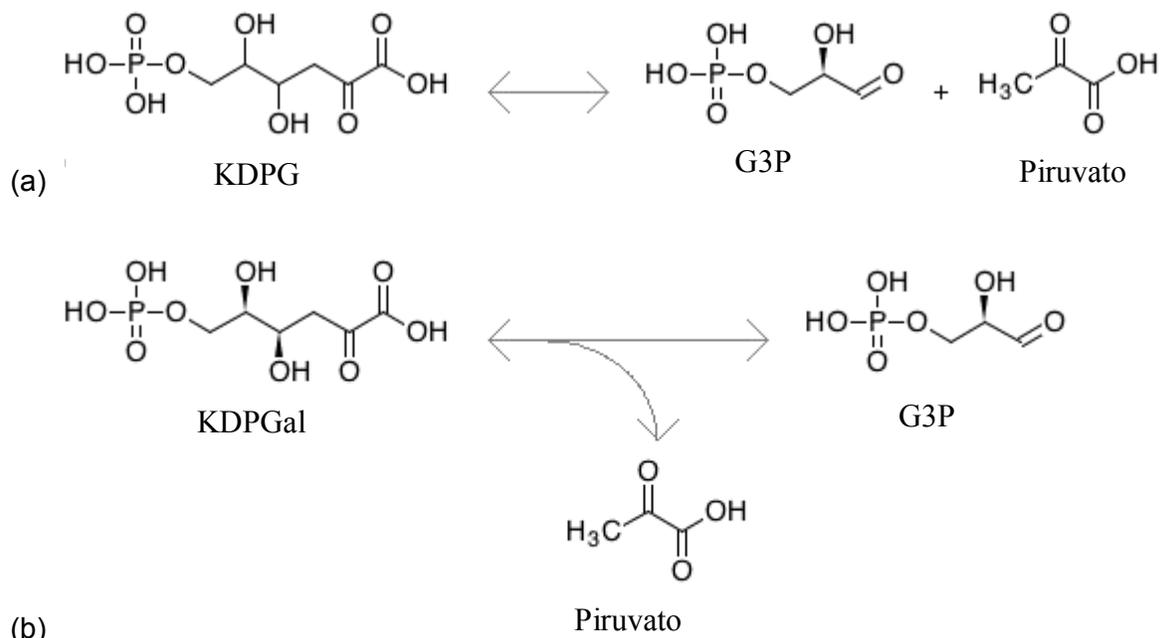
**Figura 9. Esquema general de la condensación aldólica realizada por KDPGA y KDPGalA. La diferencia radica en la especificidad por el sustrato en un estereocentro del sustrato.** Walters, M. J., Srikannathasan, V., McEwan, A. R., Naismith, J. H., Fierke, C. A., & Toone, E. J. (2008). Characterization and crystal structure of Escherichia coli KDPGal aldolase [Imagen].

KDPGA forma parte del catabolismo del gluconato de la vía Entner-Doudoroff (ED), catalizando la ruptura aldólica del sustrato 2-dehidro-3-deoxi-D-gluconato-6-fosfato (KDPG) en gliceraldehído-3-fosfato (G3P) y piruvato (Figura 10a) (Kanehisa et al., 2012; Kanehisa, 2000). En un principio, se consideraba que la vía ED se limitaba al metabolismo de algunas bacterias Gram negativas, como *Pseudomonas saccharophila*, aunque ahora se sabe que se encuentra presente en los tres dominios de la vida (Conway, 1992).

La vía ED es conocida por operar de distintas maneras de acuerdo al organismo en el que se encuentre. Puede presentarse de forma constitutiva en *Zymomonas mobilis*, un organismo anaerobio estricto que crece en asociación con plantas que tienen un alto contenido de azúcar, y que es de gran interés en la industria porque puede convertir la glucosa a etanol y CO<sub>2</sub> de tres a cuatro veces más rápido que las levaduras (García Ortega & Ponce Rivas, 2003). En bacterias entéricas como *E. coli*, esta vía es inducible por ciertos carbohidratos como el gluconato, y sólo juega un papel periférico en el metabolismo central. Esta bacteria obtiene el gluconato y otros carbohidratos cuando se encuentra colonizando el intestino de los mamíferos, a partir de los tejidos del músculo y de la comida que el hospedero consume (Peekhaus & Conway, 1998).

Existen dos versiones de la vía ED que consisten en la formación de intermediarios no fosforilados. En varias especies de *Clostridium* y en *Halobacterium saccharovororum*, hay una conversión de gluconato a 2-dehidro-3-deoxi-gluconato (KDG) por una deshidratasa específica; posteriormente, este sustrato es fosforilado por una cinasa para seguir con la ruptura aldólica por la KDPGA. Por otro lado, en arqueas termoacidófilas como *Sulfolobus solataricus* y *Thermoplasma acidophilum*, el gluconato es convertido directamente a KDG, obteniendo como productos finales al gliceraldehído y al piruvato (Buchanan, Connaris, Danson, Reeve, & Hough, 1999).

Además, de catalizar la interconversión de KDPG a piruvato y G3P, KDPGA puede catalizar la ruptura reversible de HOG a piruvato y glioxilato, así como la  $\beta$ -descarboxilación de oxalacetato dando lugar a piruvato más CO<sub>2</sub>, lo que indica que esta aldolasa es una enzima multifuncional, al menos en *E. coli* (Egan et al., 1992).



**Figura 10. Reacciones catalíticas reversibles realizadas por (a) KDPGA y (b) KDPGalA. Los productos obtenidos de ambas reacciones son G3P y piruvato.** Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes [Imagen].

KDPGalA se encuentra en la vía de degradación de la galactosa de algunas bacterias del suelo como *Azotobacter vineelandii*, *Rhizobium meliloti*, *P. saccharophila* y *P. fluorescens*, conocida como De Ley-Doudoroff (DD) (Wong & Yao, 1994). También, está presente en bacterias no patógenas del género *Mycobacterium* que son capaces de utilizar la galactosa como fuente de carbono (Szumilo, 1981). Como primer paso en la vía, la galactosa es convertida a galactonato para después ser deshidratada y fosforilada a 2-dehidro-3-deoxi-D-galactonato-6-fosfato (KDPGal). Posteriormente, la enzima realiza la ruptura aldólica de KDPGal a piruvato y G3P (Figura 10b) (Kanehisa et al., 2012; Kanehisa, 2000). Algunos hongos del género *Aspergillus* pueden degradar el galactonato a través de una vía no fosforilativa utilizando como intermediario al sustrato 2-dehidro-3-deoxi-D-galactonato (KDGal) para formar piruvato y gliceraldehído (Elshafei & Abdel-Fatah, 1991).

#### 1.5.1.5. Degradación de hidroxiprolina: 4-hidroxi-2-oxoglutarato aldolasa (HOGA)

Cataliza la ruptura reversible de 4-hidroxi-2-oxoglutarato (HOG) para formar piruvato y glioxilato (Figura 11) (Kanehisa et al., 2012; Kanehisa, 2000). Esta reacción es el último paso en la vía de degradación de 4-hidroxi-L-prolina en mamíferos. Los residuos de este



Tabla 1. Estructuras tridimensionales y características principales de cada miembro de la subfamilia de enzimas relacionadas a la NAL.

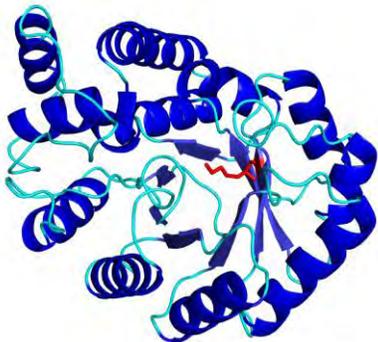
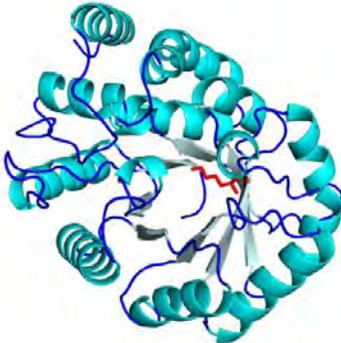
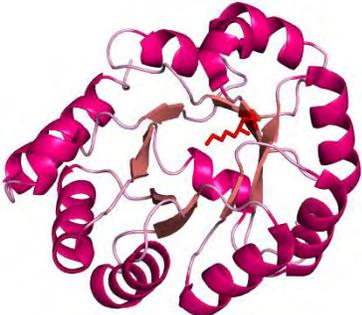
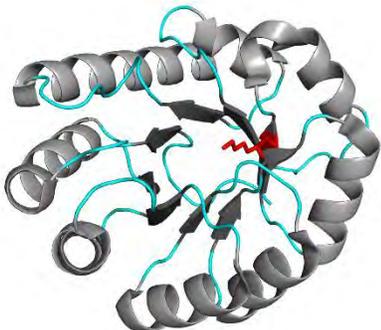
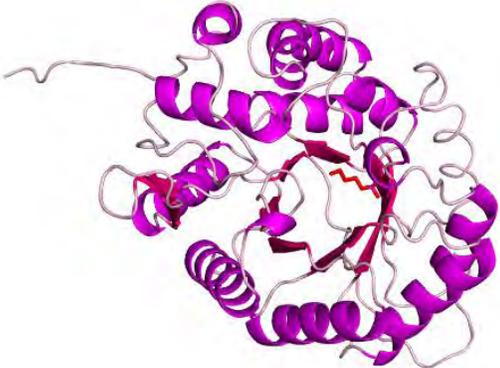
Estructura	Características principales
	<p>N-acetilneuraminato liasa(NAL; EC 4.1.3.3)  <b>PDB:</b> 1NAL  <b>Organismo:</b> <i>E. coli</i>  <b>Longitud de cadena:</b> 297 residuos  <b>Estado oligomérico:</b> tetramero  <b>Lisina catalítica:</b> K165 (rojo)  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub>  <b>Dominio C-terminal:</b> tres hélices <math>\alpha</math></p>
	<p>4-hidroxi-tetrahidrodipicolinato sintasa (THDPS; EC 4.3.3.7)  <b>PDB:</b> 1DHP  <b>Organismo:</b> <i>E. coli</i>  <b>Longitud de cadena:</b> 292 residuos  <b>Estado oligomérico:</b> tetramero  <b>Lisina catalítica:</b> K161 (rojo)  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub>  <b>Dominio C-terminal:</b> tres hélices <math>\alpha</math></p>
	<p>4-hidroxi-2-oxoglutarato aldolasa (HOGA; E.C.4.1.3.16)  <b>PDB:</b> 3S5N  <b>Organismo:</b> <i>Homo sapiens</i>  <b>Longitud de cadena:</b> 304 residuos  <b>Estado oligomérico:</b> tetramero  <b>Lisina catalítica:</b> K196 (rojo)  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub>  <b>Dominio C-terminal:</b> tres hélices <math>\alpha</math></p>
	<p>2-dehidro-3-deoxi-fosfogluconato aldolasa (KDPGA; EC 4.1.2.14)  <b>PDB:</b> 1EUN  <b>Organismo:</b> <i>E. coli</i>  <b>Longitud de cadena:</b> 213 residuos  <b>Estado oligomérico:</b> trímero  <b>Lisina catalítica:</b> K133 (rojo)  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub> + hélice <math>\alpha</math>  <b>Dominio C-terminal:</b> ---</p>

Tabla 1 (cont.). Estructuras tridimensionales y características principales de cada miembro de la subfamilia de enzimas relacionadas a la NAL.

Estructura	Características principales
	<p>2-dehidro-3-deoxi-6-fosfogalactonato aldolasa (KDPGalA; EC 4.1.2.21)  <b>PDB:</b>2V81  <b>Organismo:</b> <i>E. coli</i>  <b>Longitud de cadena:</b>212 residuos  <b>Estado oligomérico:</b> trímero  <b>Lisina catalítica:</b> K126 (rojo)  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub>  <b>Dominio C-terminal:</b>---</p>
	<p>4-(2-carboxifenil)-2-oxobut-3-enoato aldolasa (CPOEA; EC 4.1.2.34)  <b>Organismo:</b> <i>Nocardioidea</i> sp.  <b>Longitud de cadena:</b>332 residuos  <b>Estado oligomérico:</b> trímero putativo (Iwabuchi &amp; Harayama, 1998)  <b>Lisina catalítica:</b>K178  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub>  <b>Dominio C-terminal:</b> tres hélices <math>\alpha</math></p>
	<p>trans-O-hidroxibencilidenopiruvato hidratasa-aldolasa (HBPHA; EC 4.1.2.45)  <b>Organismo:</b> <i>Pseudomonas putida</i>  <b>Longitud de cadena:</b>331 residuos  <b>Estado oligomérico:</b> trímero putativo (Eaton, 1994; Kuhm et al., 1993)  <b>Lisina catalítica:</b>K183  <b>Dominio N-terminal:</b> barril (<math>\beta/\alpha</math>)<sub>8</sub>  <b>Dominio C-terminal:</b> tres hélices <math>\alpha</math></p>
	<p>Estructura modelada en RaptorX (véase Apéndice 2)</p>

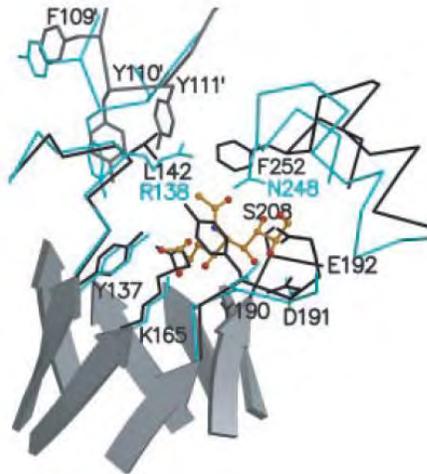
## **1.6. Importancia de la reconstrucción ancestral en la interconversión de actividades catalíticas entre enzimas de la subfamilia de NAL**

El enfoque clásico del estudio de la función de enzimas utilizando a las superfamilias mecanísticamente diversas consiste en identificar los residuos claves que determinan la actividad catalítica a través de comparaciones estructurales entre los miembros de estas familias para después cambiarlos por otros residuos que potencialmente pudieran intercambiar la actividad de una enzima por otra por medio de mutagénesis dirigida (estudio horizontal). Sin embargo, la modificación de la función usualmente requiere de múltiples mutaciones que podrían generar proteínas no funcionales debido a dos problemas principales (Harms & Thornton, 2010): 1) los residuos que son seleccionados pueden no tener relevancia para la función, aunque sean diferentes entre ambas secuencias; o bien, 2) pueden haber residuos que no se identifiquen a simple vista en la secuencia (por no estar cerca del sitio activo, por ejemplo) que impidan el cambio en la función entre los miembros de la familia, debido a que existe una interdependencia entre ellos.

Reconstruir la historia evolutiva de las familias de enzimas (estudio vertical) podría ayudar al enfoque exclusivo de las mutaciones que se fueron acumulando a lo largo del tiempo y que dieron origen a su diversificación funcional. Para llevar a cabo este estudio, es necesaria la reconstrucción de secuencias ancestrales debido a que la identificación de las mutaciones necesarias para cambiar la actividad catalítica podría obtenerse gracias a la caracterización estructural y a nivel de secuencia de las enzimas representadas en los nodos ancestrales elegidos del árbol filogenético para, posteriormente, compararlas con las enzimas actuales.

El diseño racional puede funcionar utilizando la promiscuidad catalítica de las enzimas. Por ejemplo, la NAL fue escogida como punto de inicio para obtener la actividad de THDPS, otro miembro de la misma subfamilia, principalmente, debido a que no discrimina entre diferentes aldehídos como sustratos (Joerger, Mayer, & Fersht, 2003). Un residuo de leucina fue modificado en NAL por un residuo de arginina para obtener la actividad de THDPS (L142R), ya que éste último está involucrado en la unión de L-ASA y se cree que ayuda en la catálisis (Figura 12). El resultado de esta modificación fue una actividad incrementada de THDPS, comprometiendo la actividad de NAL. Sin embargo, la interconversión de actividades entre THDPS y NAL no es lineal en ambos sentidos, ya que a pesar de modificar el residuo de arginina en THDPS (R138), no se obtiene la

actividad de NAL (Pérez-Zamorano, 2007); por lo que el enfoque vertical puede brindar información sobre la posible interacción entre las mutaciones que van de la mano con R138 para obtener la actividad de NAL en THDPS.



**Figura 12. Alineamiento estructural del sitio activo de NAL de *E. coli* (negro), en complejo con hidroxipiruvato, con el sitio activo de THDPS (azul). Se observa que L142 se encuentra en la misma posición que R138. Joerger, A. C., Mayer, S., & Fersht, A. R. (2003). Mimicking natural evolution in vitro: an N-acetylneuraminatelyase mutant with an increased dihydrodipicolinate synthase activity [Imagen].**

Otro ejemplo de diseño racional que se ha llevado a cabo en la subfamilia de NAL, ha sido el análisis de la estereoespecificidad de la enzima bifuncional KDPGA/HOGA y la KDPGaIA de bacterias. Ambas enzimas catalizan la misma reacción que difiere en la especificidad por dos estereoisómeros. Los residuos a los cuales se le han atribuido esta diferencia han sido la treonina de la posición 161 en KDPGA, y la valina de la posición 154 en KDPGaIA, por lo que se realizó una mutación puntual en ambas enzimas: T161V en KDPGA y V154T en KDPGaIA (Walters et al., 2008). El resultado de ambas mutaciones fue que la especificidad por cada estereoisómero se vio disminuida en cada enzima mutante; sin embargo, se concluye que, a pesar de que estos residuos juegan un papel muy importante en la estereoselectividad de ambas enzimas, harían falta otras sustituciones para mejorar la eficiencia de KDPGA y KDPGaIA con especificidad alterada por sus sustratos. En este caso, la reconstrucción de ancestros podría permitirnos identificar aquellos residuos de aminoácidos que pudieran estar involucrados, además de los ya descritos, en la estereoselectividad de la enzima.

## 2. Justificación y objetivos

### 2.1. Justificación

La subfamilia de enzimas relacionadas a la NAL representa un buen ejemplo de la diversificación metabólica debido a que todos sus miembros, a pesar de que comparten un ancestro en común, tanto la estructura de barriles ( $\beta/\alpha$ )<sub>8</sub> y la reacción aldólica, forman parte de distintas rutas metabólicas. Por otro lado, se ha observado que la promiscuidad catalítica está presente en estas enzimas, por lo que es importante su estudio estructural de manera vertical, es decir, reconstruir secuencias ancestrales para identificar cuáles son los residuos de aminoácidos involucrados indirectamente en la actividad catalítica y que se han modificado a lo largo de su historia evolutiva, y de esta manera generar una herramienta que permita el diseño de nuevas actividades enzimáticas mediante un enfoque evolutivo como alternativa al estudio horizontal que se ha empleado en los últimos años.

### 2.2. Objetivos

El objetivo general de este trabajo es reconstruir la historia evolutiva de las enzimas relacionadas a la NAL mediante la creación e interpretación de un árbol filogenético, con la finalidad de realizar comparaciones entre las distancias evolutivas obtenidas a partir del árbol para obtener ancestros. Los objetivos particulares son:

- Correlacionar la divergencia evolutiva de las enzimas de esta subfamilia con el nicho ecológico de las especies en las que se encuentran a nivel de clase taxonómica.
- Obtener las secuencias ancestrales de los pares de enzimas THDPS – NAL, y KDPGA/HOGA – KDPGaIA de *E. coli*, a partir de una filogenia generada de las secuencias más cercanas entre sí utilizando la matriz de distancias evolutivas del árbol filogenético original.
- Obtener y caracterizar las secuencias y estructuras teóricas de los ancestros de THDPS – NAL, y KDPGA/HOGA – KDPGaIA identificando las diferencias que podrían contribuir en su actividad catalítica.

### 3. Materiales y métodos

#### 3.1. Búsqueda y alineamiento de secuencias

Se realizaron búsquedas bioinformáticas de secuencias de cada enzima usando Basic Local Alignment Search Tool (BLAST) (Altschul, Gish, Miller, Myers, & Lipman, 1990) en la base de datos de proteínas no redundantes del National Center for Biotechnology Information (NCBI). Las secuencias semilla que se utilizaron se enlistan en la Tabla 2. Cada BLAST fue delimitado a mil secuencias por cada dominio de la vida (Archaea, Bacteria y Eukarya), y en dado caso que no existiera secuencia de esa enzima para algún dominio, se utilizó una secuencia de otro dominio. En algunos casos se utilizaron dos secuencias para la búsqueda de un mismo dominio debido a que en una sola búsqueda no salían algunas secuencias reportadas en la literatura.

**Tabla 2. Secuencias semilla (y números de acceso) utilizadas para realizar la búsqueda bioinformática.**

Enzima	Dominio	Secuencia
CPOEA	<b>B</b>	<i>Nocardioides</i> sp. (Q79EM8.1) <sup>1</sup>
THDPS	<b>A</b>	<i>Methanosaeta thermophila</i> (YP_843255.1)
	<b>B</b>	<i>Escherichia coli</i> (NP_416973.1)
	<b>E</b>	<i>Arabidopsis thaliana</i> (NP_182068.1) y <i>Phytophthora infestans</i> (AAN31469.1)
HBPHA	<b>B</b>	<i>Pseudomonas putida</i> (NP_863078.1) <sup>1</sup>
KDPGA	<b>A</b>	<i>Sulfolobus solfataricus</i> (1W37_A)
	<b>B</b>	<i>Clostridium acetobutylicum</i> (NP_347034.1) y <i>Zymomonas mobilis</i> (7436628)
	<b>E</b>	<i>Phaeodactylum tricornutum</i> (EEC50314.1)
KDPGalA	<b>B</b>	<i>Escherichia coli</i> (YP_026238.1) <sup>2</sup>
	<b>E</b>	<i>Ceratitis capitata</i> (XP_004533112.1)
HOGA	<b>A</b>	<i>Halorubrum lacusprofundi</i> (YP_002564826.1)
	<b>B</b>	<i>Escherichia coli</i> (CAA48732.1)
	<b>E</b>	<i>Bos taurus</i> (Q0P5I5.1)
NAL	<b>B</b>	<i>Escherichia coli</i> (NP_289793.1) <sup>2</sup>
	<b>E</b>	<i>Sus scrofa</i> (NP_999236.1) <sup>3</sup> y <i>Trichomonas vaginalis</i> (XP_001323296.1)

Claves: **A**: Archaea; **B**: Bacteria; **E**: Eukarya. \*Secuencia utilizada para realizar la búsqueda en:

<sup>1</sup> Archaea y Eukarya

<sup>2</sup> Archaea

<sup>3</sup> Bacteria

Los criterios de depuración fueron eliminar las secuencias con un E-value mayor a  $1 \times 10^{-11}$ , (aunque posteriormente se tomaron en cuenta sólo las secuencias que se encontraran por debajo de un umbral delimitado por un gran salto en órdenes de magnitud de este valor); con un Query Coverage menor a 60%; y observando los valores de Score, que toman en cuenta tanto la similitud y la longitud de la secuencia. Las secuencias resultantes se organizaron y se renombraron utilizando un nombre clave para cada enzima en el programa Mesquite, basándose en las iniciales anteriores de las enzimas, de la siguiente manera: **NAL** para N-acetilneuraminato liasa, **DHDPS** para 4-hidroxi-tetrahidropicolinato sintasa, **CBPHA** para 4-(2-carboxifenil)-2-oxobut-3-enoato aldolasa, **HBPHA** para trans-O-hidroxibencilidenopiruvato hidratasa-aldolasa, **KDGA** para 2-dehidro-3-deoxi-fosfogluconato aldolasa, **KDPG** para 2-dehidro-3-deoxi-6-fosfogalactonato aldolasa y **KHG** para 4-hidroxi-2-oxoglutarato aldolasa (Maddison & Maddison, 2011).

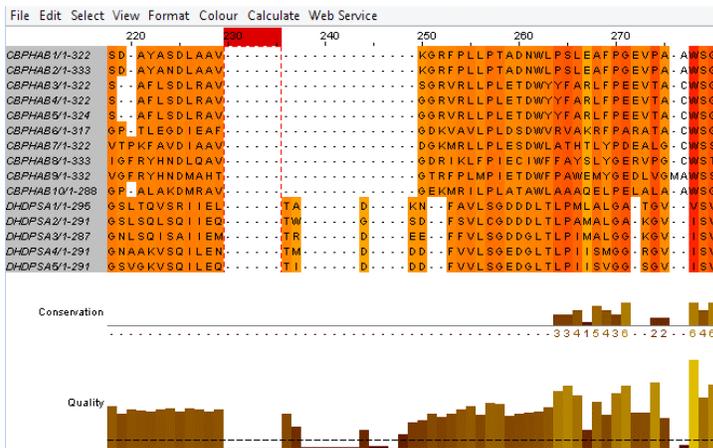
Inicialmente, se obtuvieron 2 205 secuencias tomando en cuenta los criterios de depuración ya mencionados, aunque debido al gran poder de cómputo que requería su análisis y al difícil manejo, se optó por tomar las primeras cien secuencias de cada dominio filogenético de cada enzima. Por otro lado, se especificó la clase taxonómica de cada secuencia de acuerdo a las bases de datos de List of Prokaryotic names with Standing in Nomenclature (Euzéby, 1997), Naturalista de Conabio (Naturalista, 2013) y The Taxonomicon (Brands, 2013), evitando que las especies pertenecieran a sólo unas cuantas clases, de esta manera, se intentó buscar que existiera variedad para cada enzima en cada dominio filogenético, tomando en cuenta secuencias del conjunto de datos original (2 205 secuencias) si es que el puntaje de E-value lo permitía.

Para llevar a cabo el alineamiento de las secuencias, se utilizó el programa MAFFT especializado en realizar alineamientos múltiples de secuencias de residuos de aminoácidos. El algoritmo se basa en identificar de manera rápida las regiones homólogas de las secuencias por medio de una transformación de Fourier, en la cual los aminoácidos son convertidos a valores que dependen de su volumen y polaridad (Kato, Misawa, Kuma, & Miyata, 2002).

Después, se procedió a eliminar las regiones problemáticas del alineamiento para mejorar el soporte probabilístico de las ramas y la topología del árbol filogenético resultante. Estas regiones introducen muchos gaps al momento de alinear debido a que su contenido no está bien conservado, resultado de posiciones no homólogas o divergentes y una gran cantidad de sustituciones múltiples. En primer lugar se consideró utilizar Gblocks, un

programa especializado en seleccionar bloques conservados, carentes de gaps y una cantidad permitida de posiciones flanqueantes no conservadas (Castresana, 2000); sin embargo, debido a que se basa en parámetros estrictos de conservación, sobreestima la aparición de regiones divergentes y descarta posiciones con gaps, lo que podría resultar en una pérdida de información a la hora de construir el árbol (Kück et al., 2010).

Debido a lo anterior, se estableció un método de eliminación manual de regiones problemáticas del alineamiento utilizando los criterios de conservación y calidad de cada posición del alineamiento del programa Jalview (Waterhouse, Procter, Martin, Clamp, & Barton, 2009). El puntaje de calidad de Jalview se define como la probabilidad de observar una mutación desfavorable (es decir, que un residuo cambie por otro con diferentes propiedades fisicoquímicas) en esa posición; este valor es inversamente proporcional al costo promedio de todas las mutaciones observadas en la posición, por lo que un valor alto de calidad es igual a no observar mutaciones u observar mutaciones favorables. De esta manera, se eliminaron las posiciones que cumplieran con tres criterios: un valor por debajo de 150.0 en calidad, que presentaran gaps y que su puntaje de conservación fuera cero (Figura 13).



**Figura 13. Parte del alineamiento múltiple de las aldolasas visto en Jalview (Waterhouse et al., 2009) donde se seleccionan con rojo las posiciones a eliminar de acuerdo a los criterios mencionados en el texto. Las posiciones están coloreadas en tonalidades naranja de acuerdo al grado de conservación. La línea punteada mostrada en la sección de Calidad representa el umbral de 150.**

### 3.2. Construcción y visualización del árbol filogenético

Para la construcción del árbol, se empleó el programa PhyML (Phylogenetic Maximum Likelihood) basado en el método de máxima verosimilitud. Este programa está diseñado especialmente para la búsqueda del árbol óptimo de conjuntos de datos grandes de hasta 4 mil secuencias y 2 millones de posiciones (Guindon & Gascuel, 2003).

Los parámetros que se toman en cuenta en el análisis filogenético son la **topología** del árbol, el conjunto de la **longitud** de las ramas en una topología determinada y el **modelo** de sustitución que permite maximizar la verosimilitud de los datos. En este caso, para explorar el espacio de topologías se utilizó el método de rearrreglo topológico SPR (subtree pruning and regrafting), que consiste en “cortar” un subárbol y colocarlo en una posición diferente del árbol (movimientos globales); de esta manera se evitan los óptimos locales malos permitiendo una búsqueda más exhaustiva (Hordijk & Gascuel, 2005). El método para evaluar el soporte de ramas usado fue el aLRT SH-like (approximate likelihood-rate test Shimodaira-Hasegawa-like procedure), una alternativa más rápida al análisis de bootstrap, que permite identificar la ganancia de verosimilitud comparando la topología actual de la rama de interés contra nuevos rearrreglos locales; si existe una ganancia de verosimilitud, el nuevo rearrreglo es tomado en cuenta desechando el anterior (Anisimova & Gascuel, 2006). En cuanto al modelo, se utilizó la matriz de sustitución de aminoácidos LG para construir el árbol. Esta matriz es una versión mejorada de la matriz WAG propuesta por Whelan y Goldman en 2001, aunque a diferencia de esta, la matriz de sustitución LG incorpora la variabilidad de las tasas de sustitución de los sitios de la secuencia de aminoácidos (Le & Gascuel, 2008).

Posteriormente, se emplearon dos criterios para colocar la raíz en el árbol obtenido. El primero consistió en usar el programa Notung que se basa en un puntaje D/L (Duplication/Loss) que calcula el programa al reconciliar el árbol de genes con un árbol de especies obtenido de la base de datos del NCBI. El puntaje D/L se define como:  $c_L L + c_D D + c_C C$ , donde L es el número de pérdidas, D es el número de duplicaciones y C es el número de duplicaciones condicionales; mientras que el costo de las pérdidas,  $c_L$ , el costo de las duplicaciones,  $c_D$ , y el costo de las duplicaciones condicionales,  $c_C$ , reflejan la importancia relativa de las pérdidas de genes, las duplicaciones y las duplicaciones condicionales (casos donde no es posible determinar si hubo una duplicación génica debido a la reconciliación con árboles no binarios) del árbol reconciliado. El lugar donde la raíz es más probable, que se ve reflejado en el puntaje D/L, da como resultado un árbol con menos pérdidas de genes y menos duplicaciones génicas (principio de parsimonia), ya que funciona bajo el supuesto de que estos eventos son raros (K. Chen, Durand, & Farach-Colton, 2000). El segundo criterio fue usar el método de la rama más larga (mid-point rooting method) que consiste en colocar la raíz en el punto medio de la rama que une a los dos nodos externos más lejanos. Este método asume que los cambios

evolutivos son proporcionales al tiempo y que la tasa de mutación es igual en todos los linajes (reloj molecular) (Bryant, 2001; Carr, 2009).

Se utilizaron dos programas para visualizar y editar el árbol. El primero de ellos es un servidor en línea llamado iTOL (Interactive Tree of Life) donde se representaron principalmente los dominios de la vida a los que correspondían las secuencias (Letunic & Bork, 2007, 2011); y el segundo es Archaeopteryx, donde se apreciaron mejor las distancias evolutivas de cada secuencia (Han & Zmasek, 2009).

### 3.3. Reconstrucción de las secuencias ancestrales

El árbol filogenético resultante fue convertido a una matriz de distancias utilizando el servidor T-rex (Boc, Diallo, & Makarenkov, 2012). A partir de la matriz se seleccionaron las secuencias que, al compararlas contra cada par secuencias de *E. coli* (NAL – THDPS y KDPGA – KDPGalA), tuvieron una distancia menor o igual comparada con la distancia entre ellas, es decir, el valor obtenido de la distancia evolutiva entre NAL – THDPS y KDPGA – KDPGalA fue el máximo valor que debieron tener las distancias para poder seleccionar la secuencia (Tabla 3). Las secuencias escogidas y sin editar se alinearon utilizando MAFFT y, posteriormente, a partir de cada alineamiento, se reconstruyó un árbol filogenético. Los alineamientos y los árboles resultantes fueron utilizados por el servidor del programa FastML para calcular las secuencias ancestrales de todos los nodos del árbol (Ashkenazy et al., 2012). Debido a que este servidor soporta una cantidad menor a 200 secuencias en el alineamiento inicial por análisis, se utilizó un segundo criterio de selección que consistió en dejar la secuencia de una especie por cada género, permitiendo la inclusión de sólo cinco secuencias por cada clase taxonómica, de esta manera los grupos de secuencias se redujeron para poder ser analizados por FastML.

**Tabla 3. Matriz de distancias obtenida de la matriz original compuesta por un representante de cada enzima correspondiente al dominio Bacteria. Los valores marcados en rojo son las distancias evolutivas entre NAL – THDPS y KDPGA – KDPGalA de *E. coli*.**

	KDPGB1	KDGAB1	KHGB1	CBPHAB1	HBPHAB1	NALB1	DHDPSB1
KDPGB1	0	4.11699	2.88558	7.3541	7.10881	6.05983	6.13147
KDGAB1	4.11699	0	3.82406	8.92435	8.67905	7.63008	7.70172
KHGB1	2.88558	3.82406	0	7.69294	7.44764	6.39867	6.47031
CBPHAB1	7.3541	8.92435	7.69294	0	1.55871	4.37114	4.271
HBPHAB1	7.10881	8.67905	7.44764	1.55871	0	4.12585	4.02571
NALB1	6.05983	7.63008	6.39867	4.37114	4.12585	0	3.14851
DHDPSB1	6.13147	7.70172	6.47031	4.271	4.02571	3.14851	0

\*B1 corresponde a la secuencia número 1 del dominio Bacteria.

Por otro lado, también se utilizó la herramienta de reconstrucción de secuencias ancestrales del servidor de Datamonkey (Delport, Poon, Frost, & Kosakovsky Pond, 2010; Pond, Frost, & Muse, 2005; Pond & Frost, 2005) para comparar las secuencias resultantes contra las obtenidas de FastML. Ambos servidores usan dos formas de reconstrucción: la reconstrucción “joint” consiste en encontrar la ruta de cambios de estado de carácter más verosímil, tomando en cuenta las probabilidades de sustitución de todos los nodos, mientras que la reconstrucción “marginal” toma en cuenta los estados de carácter más verosímiles de acuerdo al cambio con mayor probabilidad de un solo nodo (Pupko, Pe, Shamir, & Graur, 2000). El primer tipo de reconstrucción es más adecuado cuando se quieren contar los cambios de cada sitio, mientras que el segundo tipo de reconstrucción es más adecuado cuando sólo se quiere obtener la secuencia de un nodo en particular (Z. Yang, 2006b).

Para analizar las secuencias de los ancestros, éstas se alinearon con las secuencias de THDPS y NAL, o KDPGA y KDPGalA de *E. coli* usando MAFFT, para posteriormente, obtener el porcentaje de identidad entre cada uno de los nodos comparado contra las secuencias actuales mediante la herramienta de alineamiento por pares de Jalview. Los alineamientos son presentados en gráficos realizados con la herramienta Color Align Conservation (Stothard, 2000), que analiza los residuos de cada columna del alineamiento para después seleccionar un color de fondo para los residuos que se conservan (negro) o que se intercambian por un residuo similar (gris).

### **3.4. Análisis estructural de las proteínas ancestrales**

Se utilizó el servidor de predicción de estructura de proteínas RaptorX para predecir las estructuras en 3D de las secuencias obtenidas (Källberg et al., 2012). Este servidor predice tanto la estructura secundaria y terciaria de una secuencia dada, así como la accesibilidad al solvente y las regiones desordenadas.

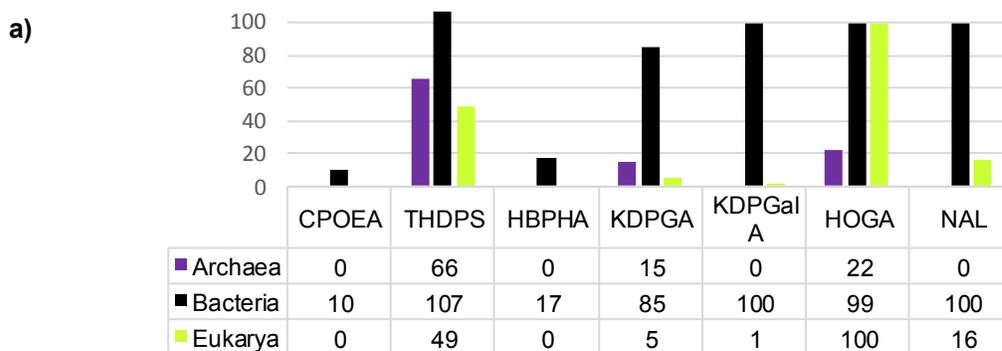
Después de obtener las estructuras tridimensionales de las secuencias de los nodos seleccionados, se llevó a cabo una comparación estructural con los PDBs de secuencias actuales en el programa PyMOL versión 1.5 (Schrödinger, 2010), por medio de la obtención de los valores de RMSD, con la finalidad de identificar las diferencias entre las estructuras actuales de *E. coli* y los ancestros hipotéticos obtenidos.

## 4. Resultados

### 4.1. Número de secuencias y diversidad taxonómica

A partir de la búsqueda y depuración de secuencias se obtuvieron 792 representantes de cada dominio y enzima relacionada a la NAL, de las cuales 518 fueron de bacterias, 103 de arqueas y 171 de eucariontes (Figura 14). Todos los miembros de la subfamilia de NAL tuvieron representantes del dominio Bacteria, mientras que para las enzimas KDPGalA, NAL, CPOEA y HBPHA no se obtuvieron secuencias de arqueas y sólo las dos últimas no tuvieron representantes de eucariontes. Las clases taxonómicas identificadas en cada dominio fueron: ocho de Archaea, de las cuales destaca Halobacteria como la más numerosa; 33 clases de Bacteria, donde Gammaproteobacteria y Alphaproteobacteria fueron las más numerosas; y, 48 de Eukarya, destacando a Mammalia como la clase que más representantes tuvo (Figura 15).

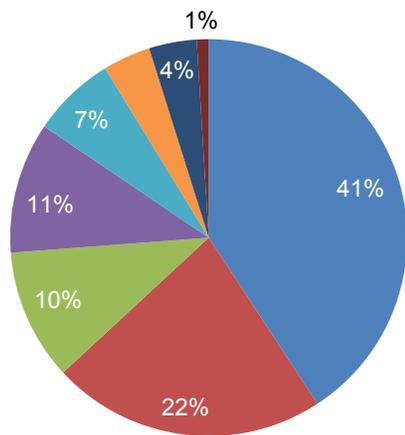
Algunas secuencias presentaron etiquetas que podrían no corresponder adecuadamente a la actividad catalítica de la enzima, nombradas generalmente como THDPS o relacionadas a ésta; por ejemplo, secuencias de hongos que fueron nombradas como THDPS, se agruparon junto a las secuencias de HOGA en la reconstrucción filogenética. Sin embargo, no se puede saber con certeza si corresponden a la actividad descrita por el nombre debido a la falta de datos experimentales. A pesar de esto, la probable actividad catalítica de cada secuencia se asignó de acuerdo al límite de E-value de cada BLAST para cada enzima.



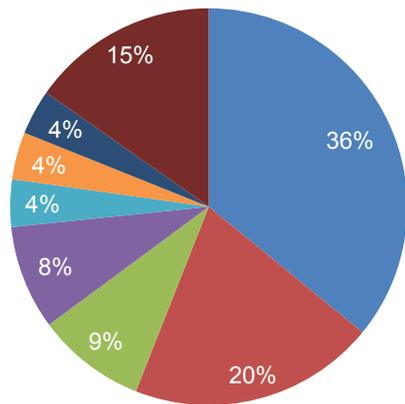
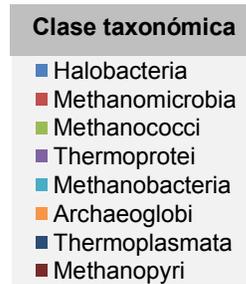
b)

Enzima	Archaea	Bacteria	Eukarya	Total
CPOEA	0	10	0	10
THDPS	66	107	49	222
HBPHA	0	17	0	17
KDPGA	15	85	5	105
KDPGalA	0	100	1	101
HOGA	22	99	100	221
NAL	0	100	16	116
Total	103	518	171	792

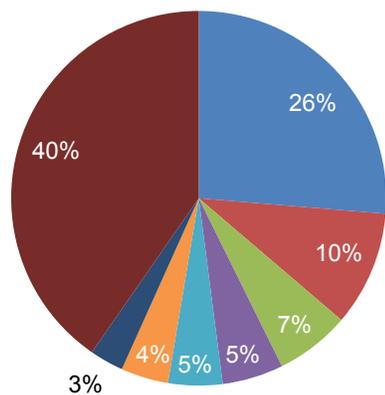
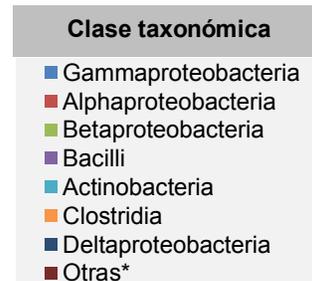
**Figura 14. Número de secuencias obtenidas de cada enzima de la subfamilia en cada dominio filogenético. a) Representación gráfica del número de secuencias por enzima y por dominio. b) Tabla donde se presentan los totales de secuencias por enzima y por dominio.**



a) Archaea



b) Bacteria



c) Eukarya



**Figura 15. Clases taxonómicas en las cuales se encontraron secuencias de la subfamilia de NAL, obtenidas por cada dominio filogenético; y porcentaje de secuencias pertenecientes a cada clase. a) Se muestran las ocho clases correspondientes de arqueas; b) y c) Se representan las siete clases con mayor número de secuencias de bacterias y eucariontes. \*El apartado 'Otras' representa clases con un número menor a 5 secuencias (véase Apéndice 1).**

## 4.2. Árbol filogenético de la subfamilia de enzimas relacionadas a la NAL

La mayor parte de las secuencias elegidas no tenía un respaldo reportado el cual dijera que realizaran la actividad enzimática del nombre anotado de la base datos del NCBI, sin embargo, se encontró al menos un reporte de actividad para cada enzima, aunque no para cada dominio (Tabla 4).

**Tabla 4. Secuencias pertenecientes a la subfamilia de enzimas relacionadas a la NAL con actividad enzimática comprobada experimentalmente.**

Dominio	Enzima	No. Acceso	Organismo	Referencia
Bacteria	CPOEA	Q79EM8.1	<i>Nocardioides sp.</i>	(Iwabuchi & Harayama, 1998)
	THDPS	NP_416973.1	<i>Escherichia coli</i>	(Blickling et al., 1997)
		BAC92753.1	<i>Methylophilus methylotrophus</i>	(Gunji et al., 2004)
		YP_079075.2	<i>Bacillus licheniformis</i>	(Halling & Stahly, 1976)
		YP_005493283.1	<i>Bacillus megaterium</i>	(Webster & Lechowich, 1970)
		ZP_17378547.1	<i>Bacillus cereus</i>	(Hoganson & Stahly, 1975)
		ZP_07078487.1	<i>Lactobacillus plantarum</i>	(Cahyanto, Kawasaki, Nagashio, Fujiyama, & Seki, 2006)
		ZP_12897052.1	<i>Staphylococcus aureus</i>	(Burgess et al., 2008)
	HBPFA	YP_002887426.1	<i>Pseudomonas fluorescens</i>	(Ferrara, Mapelli, Sello, & Di Gennaro, 2011)
		AAD45417.1	<i>Sphingobium xenophagum</i>	(Kuhm et al., 1993)
	KDPGA	YP_162732.1	<i>Zymomonas mobilis</i>	(Ryu, Park, Lee, & Kim, 2010)
	KDPGaIA	YP_026238.1	<i>Escherichia coli</i>	(Walters et al., 2008)
	HOGA	NP_288287.1	<i>Escherichia coli</i>	(Grady, Wang, & Dekker, 1981)
		YP_002798762.1	<i>Azotobacter vinelandii</i>	(Taha & Deits, 1994)
	NAL	NP_289793.1	<i>Escherichia coli</i>	(Aisaka et al., 1991)
		EHP45801.1	<i>Clostridium perfringens</i>	(Nees, Schauer, & Mayer, 1976)
		YP_002635477.1	<i>Staphylococcus carnosus</i>	(García García, Sola Carvajal, García Carmona, & Sánchez Ferrer, 2012)
		YP_003064435.1	<i>Lactobacillus plantarum</i>	(Sánchez-Carrón et al., 2011)
		YP_005178230.1	<i>Haemophilus influenzae</i>	(Lilley, Barbosa, & Pearce, 1998)

**Tabla 4 (cont.). Secuencias pertenecientes a la subfamilia de enzimas relacionadas a la NAL con actividad enzimática comprobada experimentalmente.**

Dominio	Enzima	No. Acceso	Organismo	Referencia
Eukarya	THDPS	NP_182068.1	<i>Arabidopsis thaliana</i>	(Griffin et al., 2012)
		ACU23971.1	<i>Glycine max</i>	(Mazelis, Whatley, & Whatley, 1977)
		P24846.1	<i>Triticum aestivum</i>	(Kumpaisal, Hashimoto, & Yamada, 1987)
		ACF78377.1	<i>Zea mays</i>	(Mazelis et al., 1977)
	HOGA	Q0P515.1	<i>Bos taurus</i>	(Dekker & Kitson, 1992)
		NP_001099825.1	<i>Rattus norvegicus</i>	(Anderson, Scholtz, & Schuster, 1985)
	NAL	NP_999236.1	<i>Sus scrofa</i>	(Schauer & Wember, 1996)
		NP_001039443.1	<i>Bos taurus</i>	(Sirbasku & Binkley, 1970)
		XP_001323296.1	<i>Trichomonas vaginalis</i>	(Meysick, Dimock, & Garber, 1996)
Archaea	KDPGA	YP_005642780.1	<i>Sulfolobus solfataricus</i>	(Williams, Pollock, Blackburne, & Goldstein, 2006)
		YP_004892875.1	<i>Thermoproteus tenax</i>	(Pauluhn et al., 2008)
		YP_023804.1	<i>Picrophilus torridus</i>	(Reher, Fuhrer, Bott, & Schönheit, 2010)

En el árbol final obtenido se observa que las secuencias se distribuyeron de acuerdo a su probable actividad enzimática asignada con base en los criterios mencionados en la sección de materiales y métodos, lo que sugiere que potencialmente podrían cumplir la función asociada a la posición del árbol (Figura 16a). Por otra parte, el programa Notung encontró 45 ramas con el puntaje D/L = 1 681.5, que fue el mejor valor encontrado en las ramas seleccionadas como probables raíces, por lo que no se pudo colocar una raíz. Este puntaje fue obtenido de 241 duplicaciones, 182 duplicaciones condicionales y 1 320 genes perdidos. Generalmente, los mejores valores D/L se presentaron en las ramas más profundas del árbol (Figura 16b; 17c; 17g; 17h), y el clado que presentó una mayor cantidad de probables raíces fue el de las secuencias de KDPGaIA (25; Figura 17g). No obstante, el programa Archaeopteryx indicó que la raíz podría ser colocada en la rama con aLRT = 1 (Figura 16b, flecha roja), que fue escogida mediante el criterio de la rama más larga.

El soporte probabilístico de las ramas del árbol fue variable, ya que se presentaron valores de aLRT de 0.347 en la rama principal que agrupa a NAL y valores cercanos a 0.5 en otras dos ramas principales (Figura 16b); aunque también se obtuvieron buenos valores en otras ramas principales que van desde 0.8 a 1, por ejemplo, en la rama que agrupa a todas las secuencias de THDPS se obtuvo un valor de 0.999, mientras que en la rama que agrupa a las secuencias de KDPGA, HOGA y KDPGaIA se obtuvo un aLRT de

1. También se obtuvieron algunos valores de cero (10) aunque se localizaron cerca de las hojas del árbol.

En primera instancia, se observa una bifurcación indicada con una flecha de color rojo en la rama más larga (Figura 16b), donde por un lado se engloban a todas las secuencias de KDPGalA de bacterias, secuencias de KDPGA de Bacteria y Eukarya, así como las secuencias de HOGA de Archaea. Este grupo está conformado por enzimas triméricas. Por otro lado, se observa que HOGA de eucariontes se encuentra más relacionada con las secuencias de CPOEA y HBPHA de bacterias, y a su vez, este grupo comparte un ancestro en común con las secuencias de THDPS, mientras que NAL y KDPGA de arqueas se encuentran más relacionadas entre sí. Los paneles de la Figura 17 muestran con mayor detalle las clases taxonómicas que se representan en cada subárbol de acuerdo a cada enzima de la subfamilia. En la Figura 17a se muestra el subárbol que incluye a las secuencias de HOGA de bacterias y unas cuantas secuencias de KDPGA del mismo dominio. Se observa que la mayor parte de las enzimas se encuentran en organismos pertenecientes al phylum Proteobacteria, donde la clase Gammaproteobacteria tiene una mayor cantidad de representantes que son patógenos Gram-negativos, aunque también se encuentran secuencias pertenecientes a especies aeróbicas de vida libre que se desarrollan principalmente en el suelo, pertenecientes a la clase Actinobacteria. Por otro lado, en el subárbol con los representantes de HOGA en eucariontes se observan dos grupos definidos; en una de las ramas se agruparon todas las secuencias de hongos, mientras que en la otra se agruparon algunas secuencias de protozoarios y animales, donde se obtuvo una mayor cantidad de secuencias pertenecientes a mamíferos (Figura 17b). En cuanto al dominio Archaea, se obtuvieron sólo secuencias pertenecientes a especies halófilas extremas de la clase Halobacteria que son capaces de utilizar la luz como fuente de energía (Figura 17c).

De la enzima KDPGA, se obtuvieron mayormente especies de bacterias de las clases Clostridia y Bacilli, mientras que en Eukarya se obtuvieron como representantes a algunas algas (Figura 17c); por otro lado, se encontraron especies de arqueas termófilas anaerobias que pueden crecer en condiciones ligeramente ácidas de la clase Thermoprotei y especies termoacidófilas de la clase Thermoplasmata (Figura 17e). Las secuencias de KDPGalA pertenecieron a especies del phylum Proteobacteria, destacando que en una rama se encontraron únicamente secuencias pertenecientes a géneros de la clase Alphaproteobacteria que establecen relaciones simbióticas con plantas leguminosas,

como *Sinorhizobium*. Dentro de este subárbol, se agrupó la única secuencia de Eukarya perteneciente a la mosca *Ceratitis capitata* (Figura 17g). Las secuencias de CPOEA y HBPHA compartieron un ancestro en común más cercano que con las demás enzimas. Se representaron casi exclusivamente especies de la clase Actinobacteria en el subárbol donde se encuentran las secuencias de CPOEA, mientras que para HBPHA, la mayor parte de las secuencias pertenecieron a especies del phylum Proteobacteria (Figura 17f).

En cuanto al subárbol de NAL, se observa que las secuencias formaron dos agrupaciones principales. En el grupo más grande, se presentó una mayoría de especies de las clases Gammaproteobacteria y Bacilli, así como la secuencia del microorganismo parásito *Trichomonas vaginalis*, un protozoo anaeróbico flagelado. El otro grupo estuvo representado principalmente por bacterias de vida libre de las clases Bacteroidia y Planctomycetacia, además de algunas especies patógenas de Gammaproteobacteria que tuvieron una estrecha relación con especies de vertebrados (Figura 17d).

Finalmente, dentro del subárbol que corresponde a THDPS, se observa que las secuencias de arqueas se agruparon en cuatro ramas donde resalta la clase Methanomicrobia por ser la más numerosa, mientras que las secuencias de bacterias formaron dos clados distintos formados por representantes de las clases Bacilli y Alpha, Beta, Gamma y Deltaproteobacteria. El clado de Eukarya, que también se encuentra dentro de este subárbol, se conformó principalmente de secuencias de especies de algas y plantas de las clases Liliopsida y Magnoliopsida; sin embargo, las secuencias de eucariontes de protozoarios y algunos animales se presentaron de forma aislada a lo largo de los clados de bacterias (Figura 17h).

Enzima	Dominio filogenético
CPOEA	Archaea
THDPS	Bacteria
HBPHA	Eukarya
KDPGA	
KDPGalA	
HOGA	
NAL	

a)

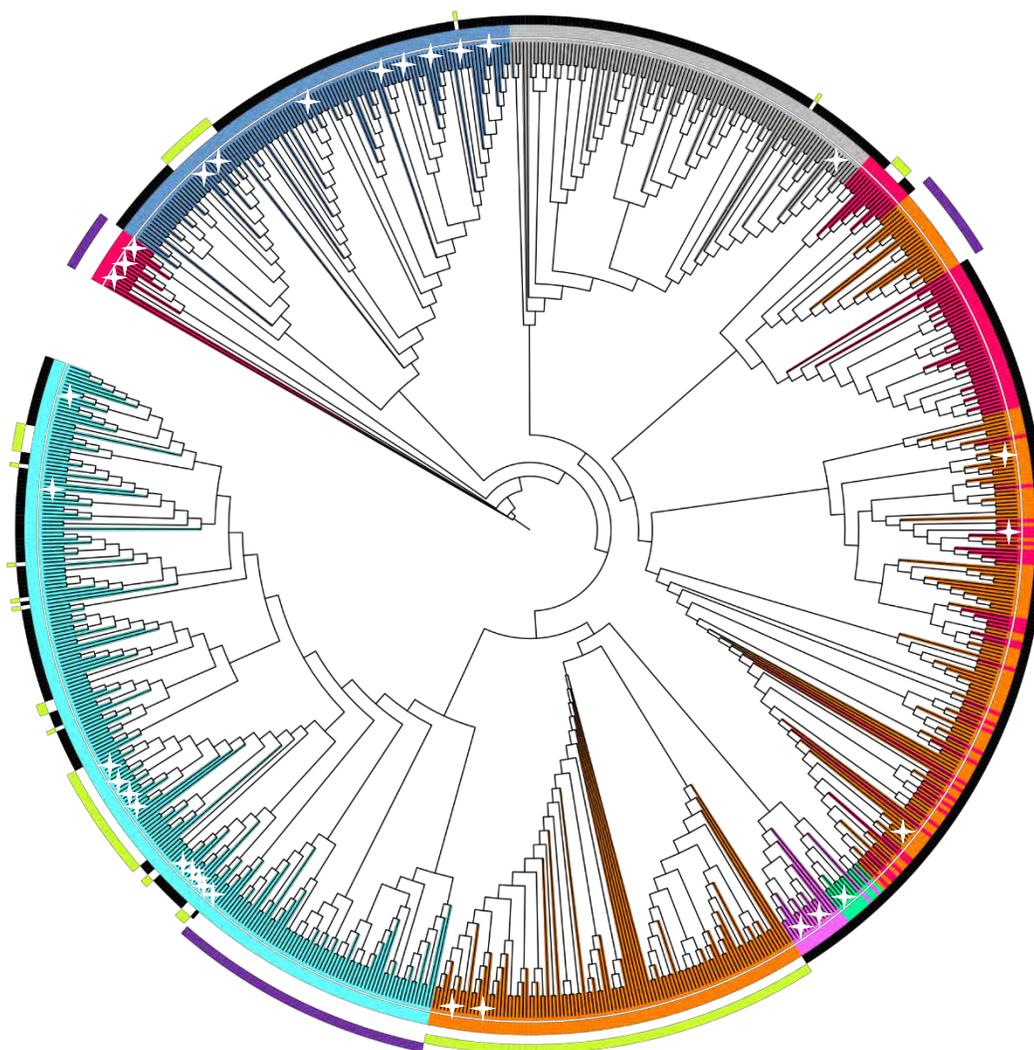


Figura 16a. Árbol filogenético de la subfamilia de enzimas relacionadas a la NAL. a) Vista circular del árbol donde se puede apreciar la distribución de las secuencias de acuerdo a su actividad y al dominio de la vida al cual pertenecen. Las secuencias con actividad enzimática comprobada experimentalmente se localizan con estrellas blancas. Figura realizada en iTOL (Letunic & Bork, 2007, 2011).

Enzima	Clase taxonómica			
	Bacteria		Eukarya	
CPOEA	Acti	Actinobacteria	Lili	Liliopsida
THDPS	Agari	Agaricomycetes	Magn	Magnoliopsida
HBPHA	Alph	Alphaproteobacteria	Mamm	Mammalia
KDPGA	Baci	Bacilli	<b>Archaea</b>	
KDPGalA	Beta	Betaproteobacteria	Halo	Halobacteria
HOGA	Delt	Deltaproteobacteria	Meth	Methanomicrobia
NAL	Gamm	Gammaproteobacteria	Ther	Thermoprotei

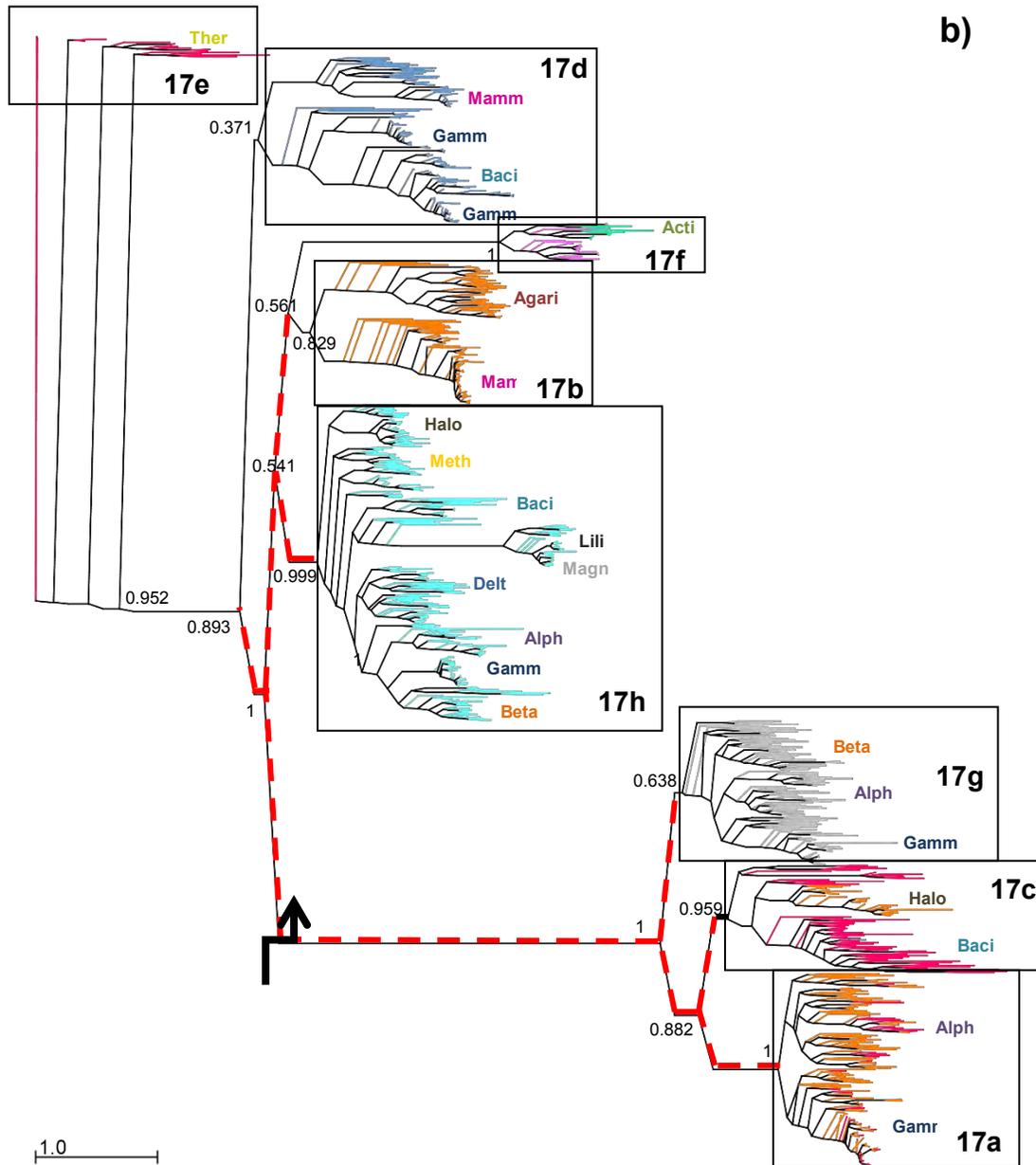
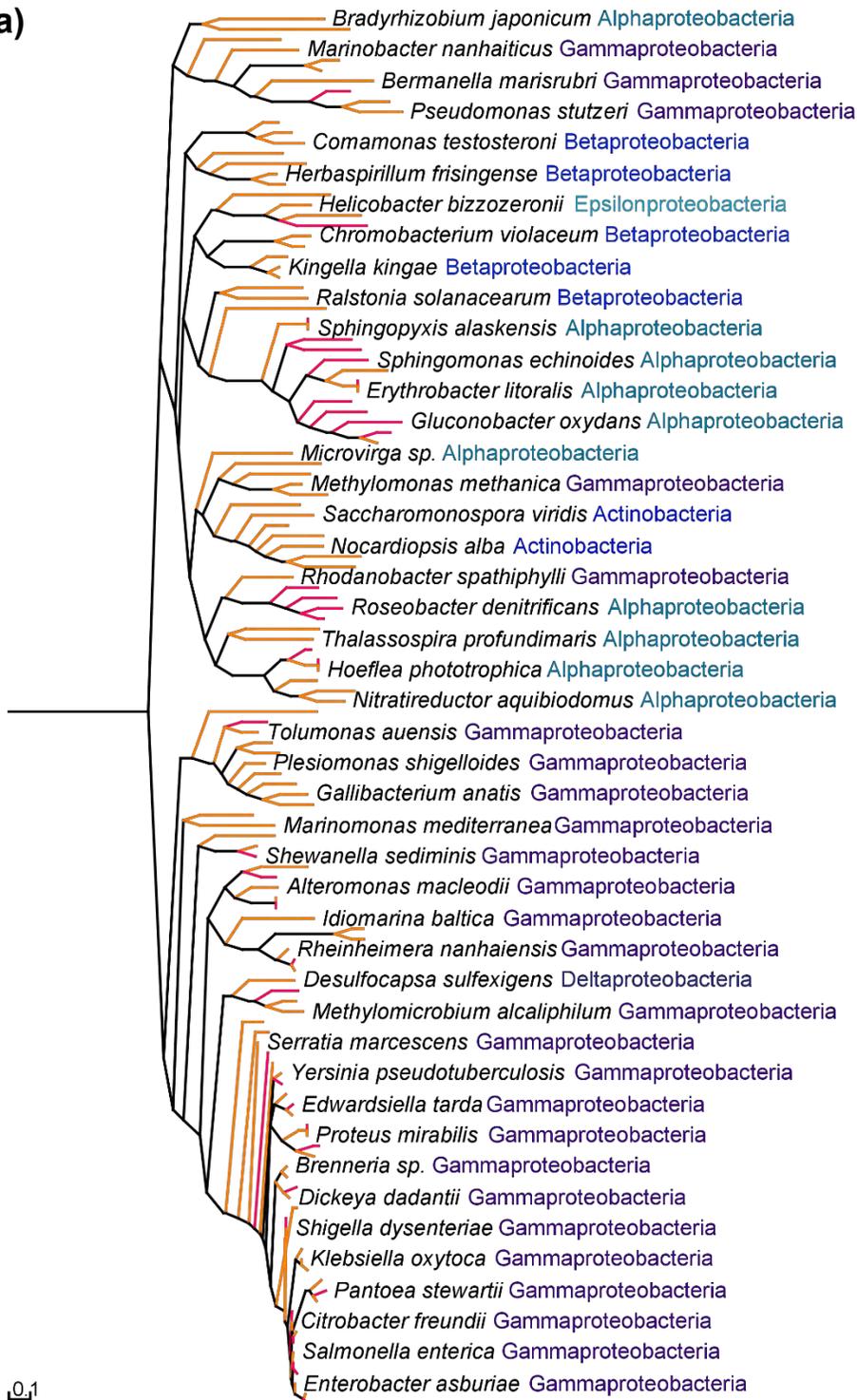


Figura 16b. Árbol filogenético de la subfamilia de enzimas relacionadas a la NAL. b) Vista rectangular del árbol donde se presentan los valores de aLRT de las ramas más importantes, así como las clases taxonómicas más representativas, las distancias evolutivas de cada secuencia y las ramas más profundas que podrían ser potenciales raíces (líneas punteadas rojas). La flecha negra indica la división del árbol de acuerdo al estado oligomérico general de las enzimas. Cada rectángulo corresponde a un panel de la Figura 17. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

a)



0.1

Figura 17a. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. a) Subárbol donde se encuentran las secuencias de HOGA de Bacteria y algunas secuencias de KDPGA de Bacteria. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

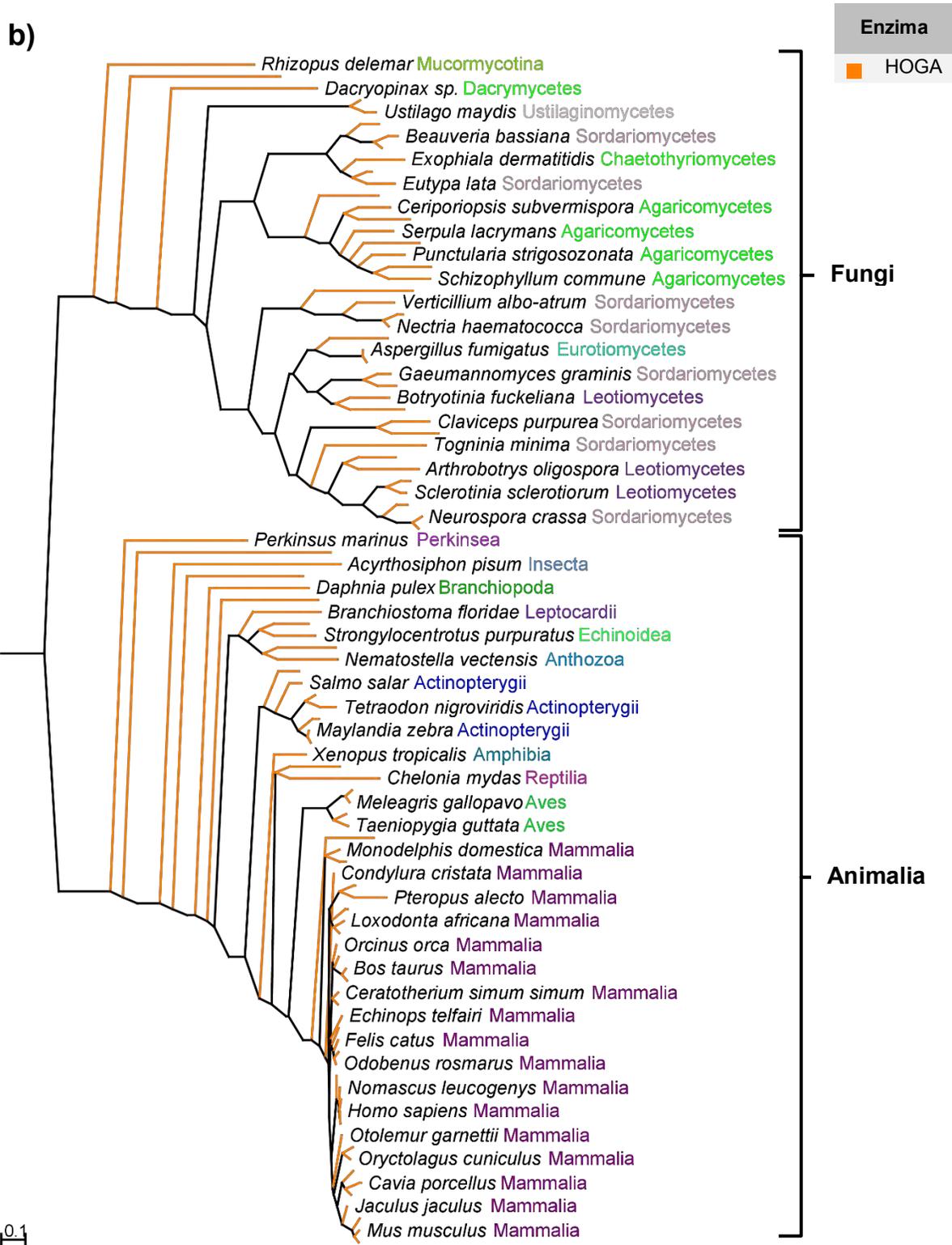


Figura 17b. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. b) Subárbol donde se representan las secuencias de HOGA de Eukarya. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

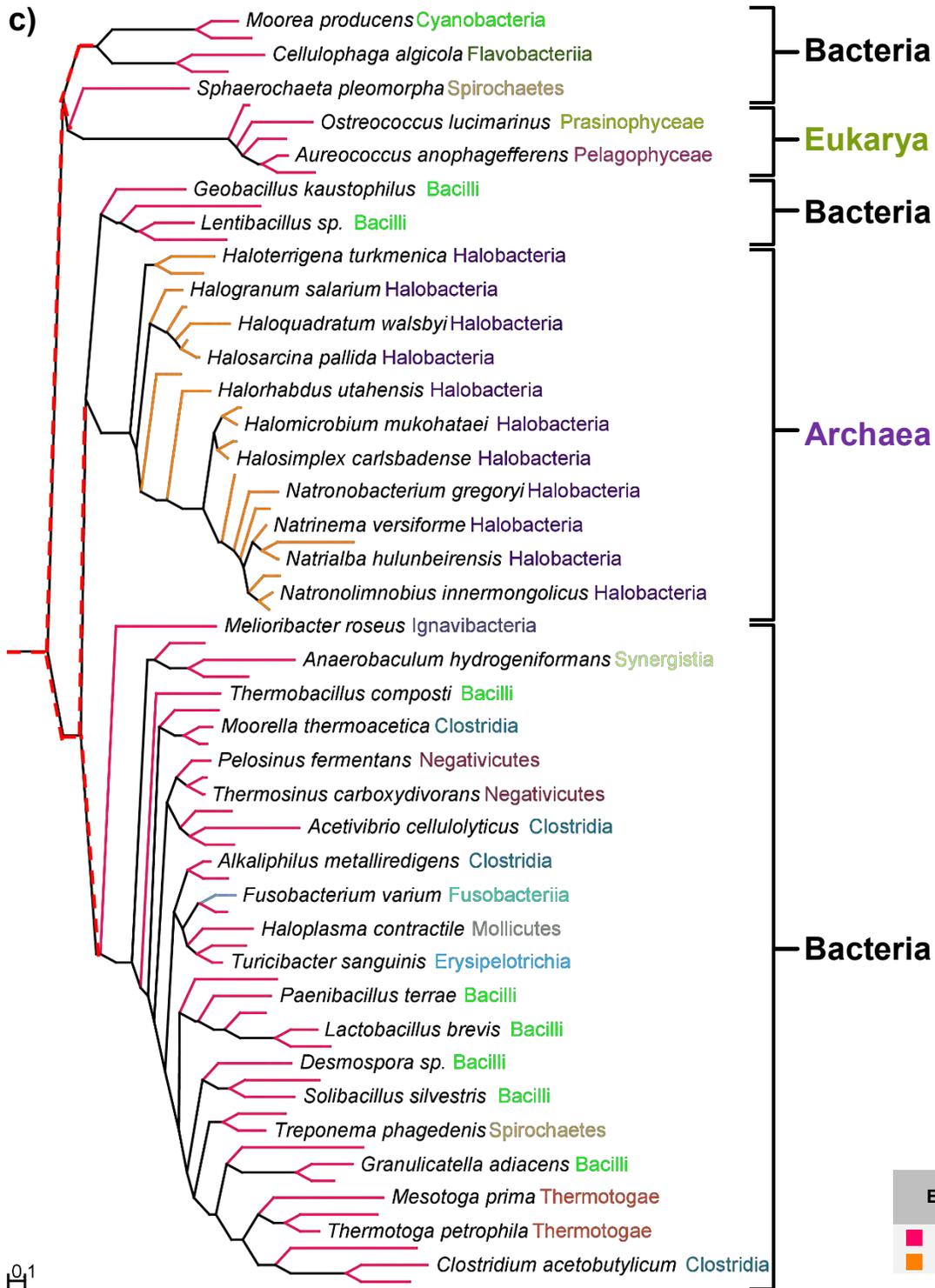


Figura 17c. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. c) Subárbol donde se encuentran secuencias de KDPGA de Bacteria y Eukarya, y las secuencias de HOGA de Archaea. Las ramas marcadas por el programa Notung como probables raíces se resaltan con líneas punteadas de color rojo. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

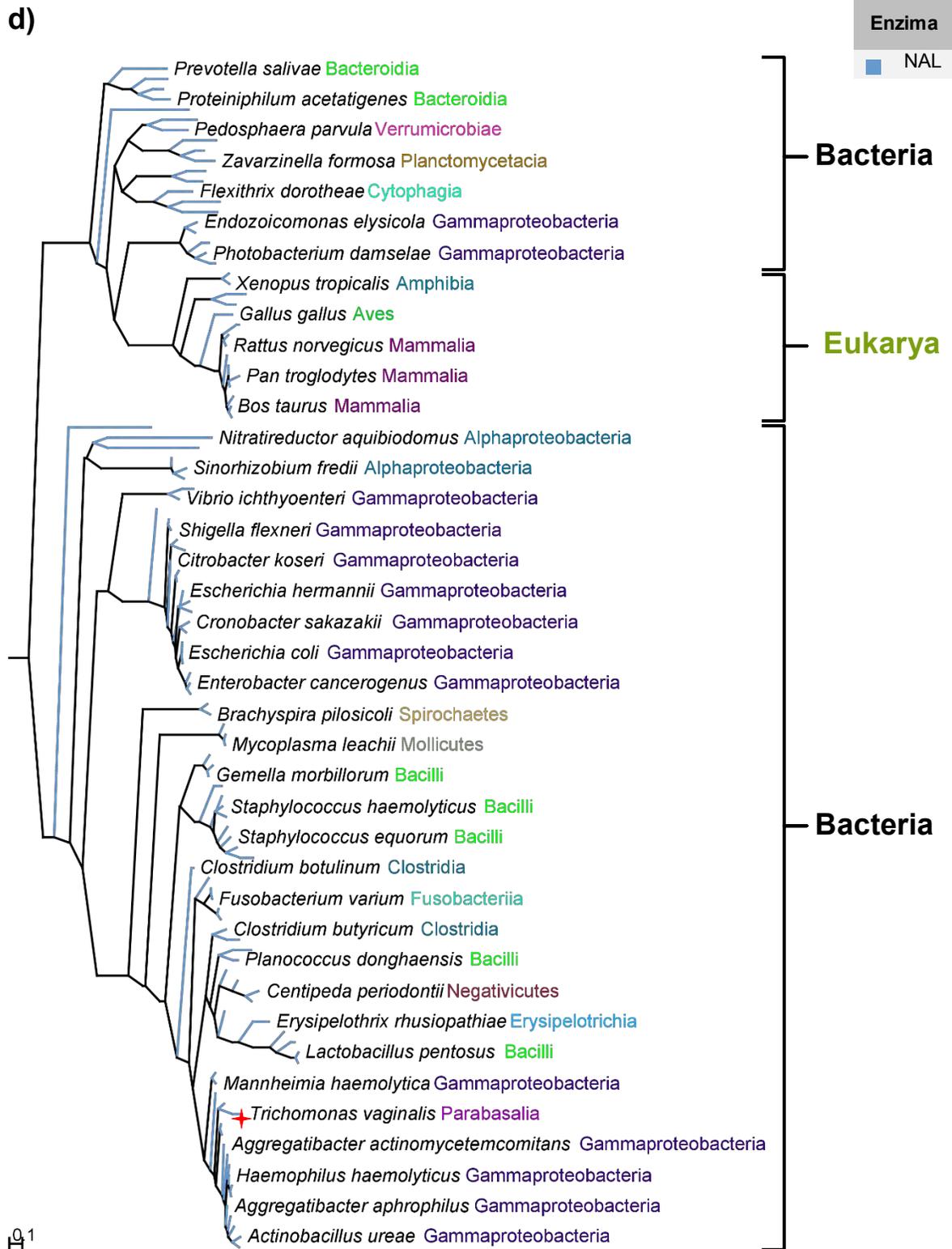
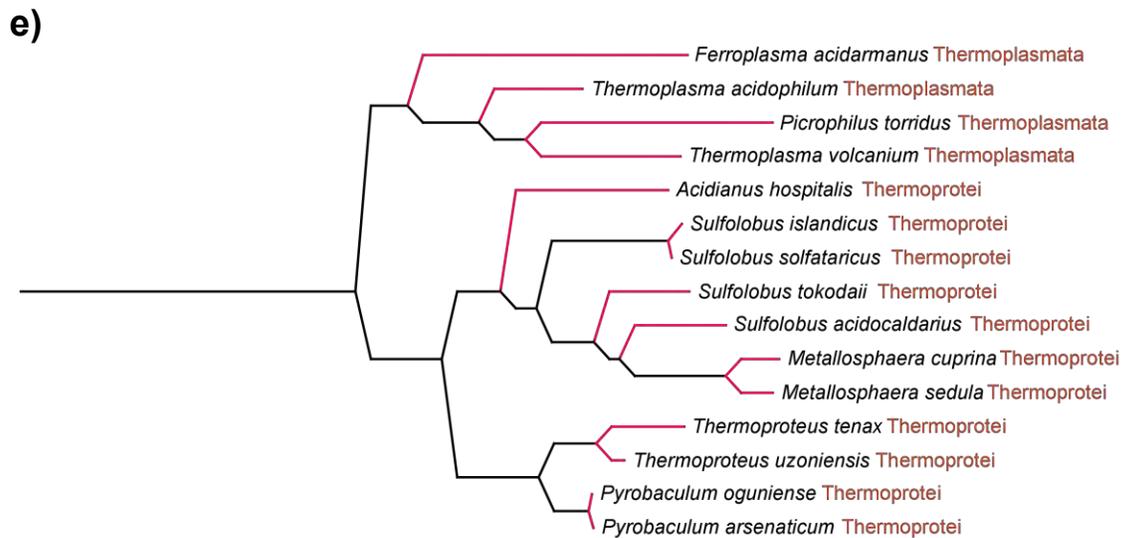
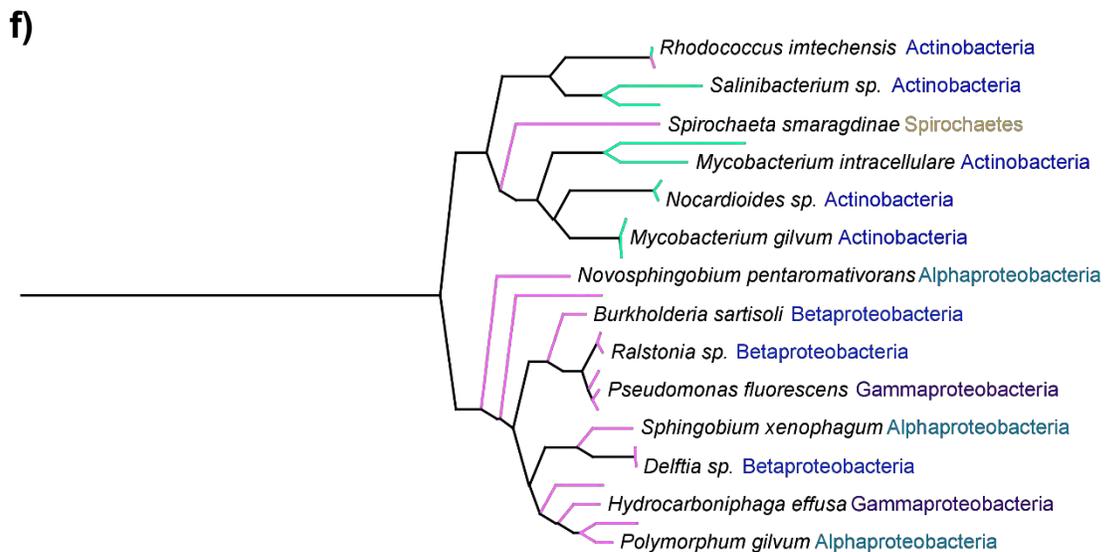


Figura 17d. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. d) Subárbol donde se encuentran las secuencias de NAL de Bacteria y Eukarya. Se resalta con una estrella roja la posición de la secuencia de *T. vaginalis*. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

Enzima	
■	CPOEA
■	HBPFA
■	KDPGA



0.1



0.1

Figura 17e-f. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. Subárboles de e) secuencias de KDPGA de Archaea y de f) CPOEA y HBPFA de Bacteria. Figuras realizadas en Archaeopteryx (Han & Zmasek, 2009).

g)

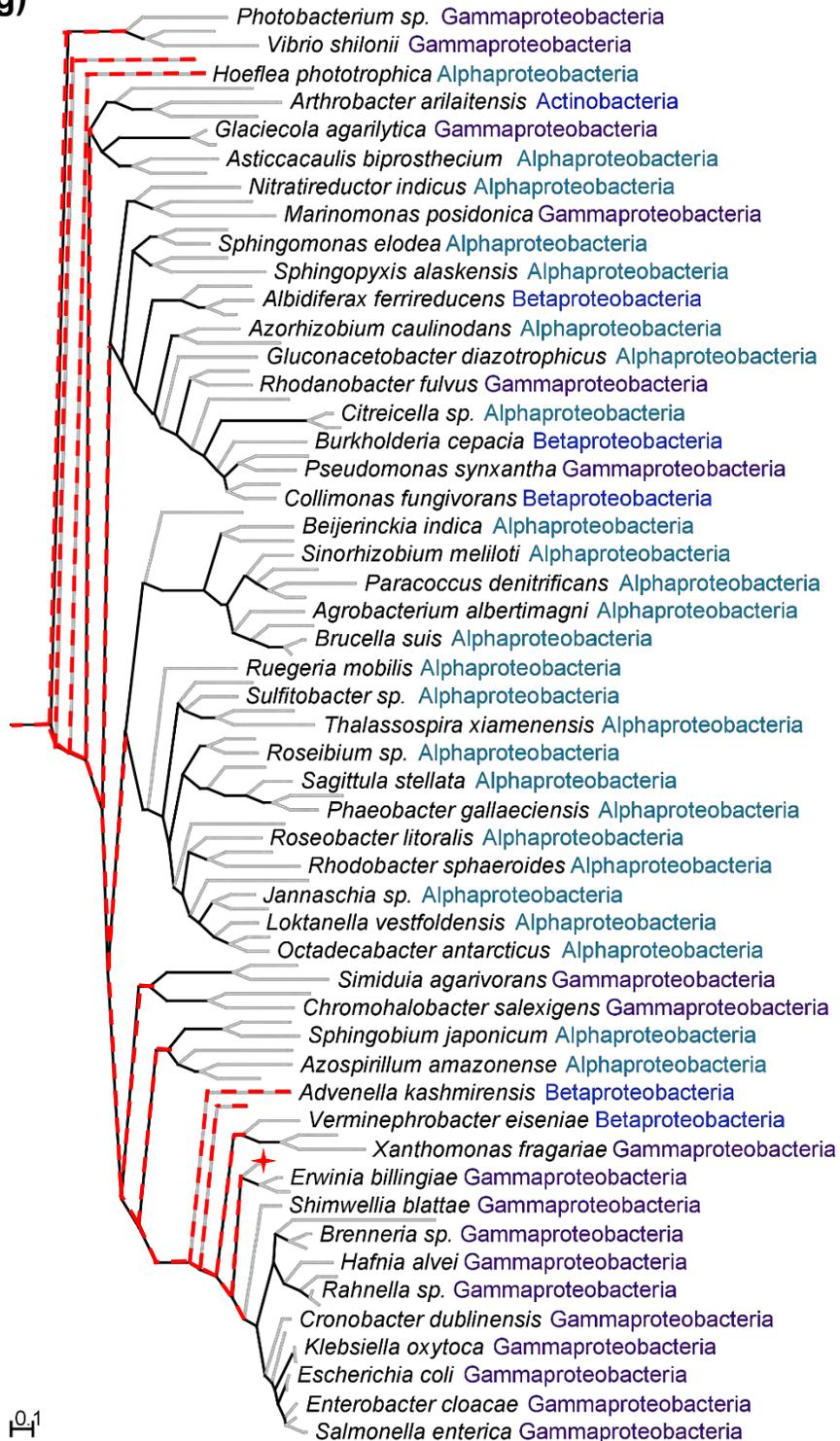


Figura 17g. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. g) Subárbol de las secuencias de KDPGalA. Se encontraron sólo secuencias del dominio Bacteria, a excepción de una secuencia eucarionte perteneciente a *C. capitata* (estrella roja). Las ramas marcadas por el programa Notung como probables raíces se resaltan con líneas punteadas en rojo. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

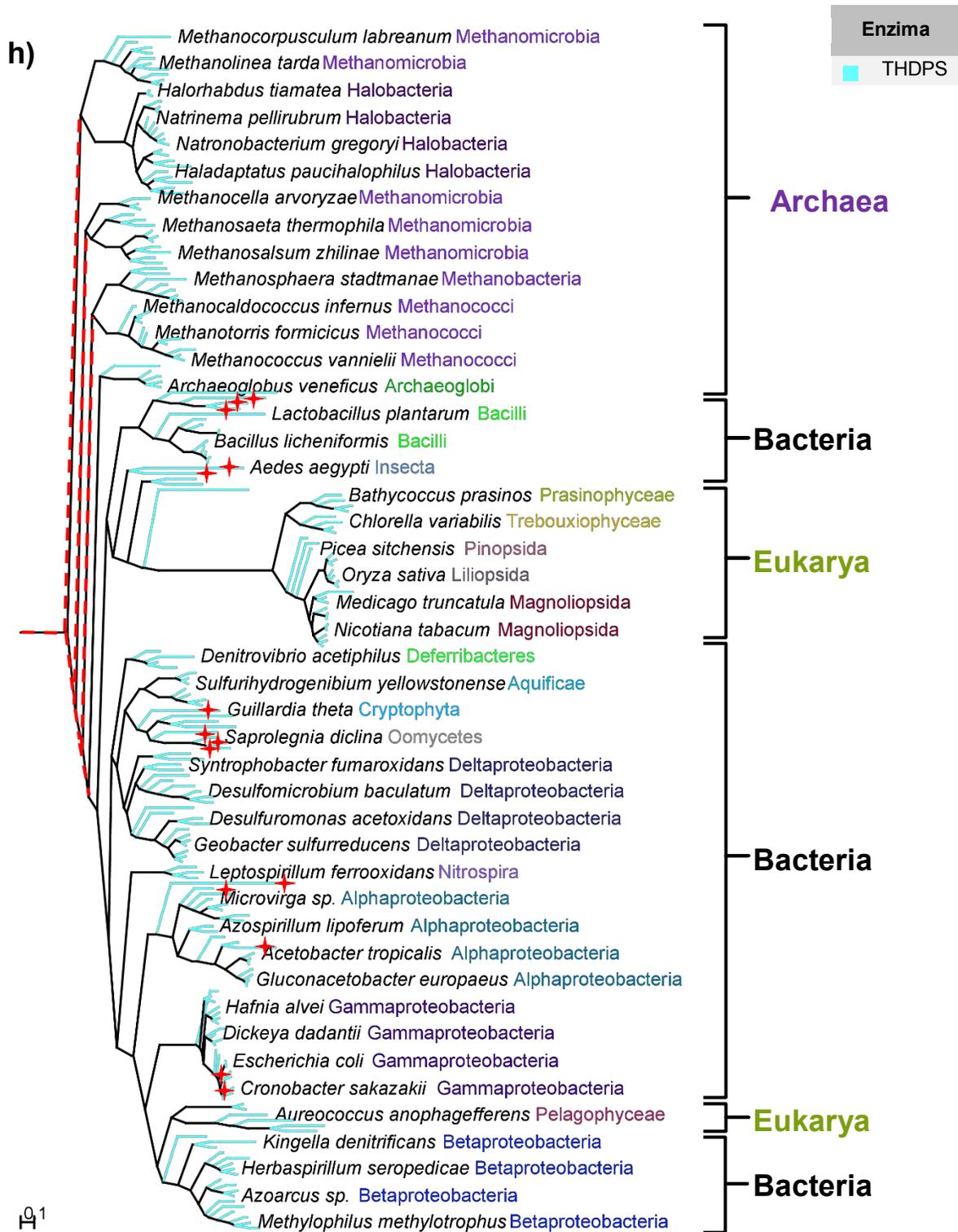


Figura 17h. Clados pertenecientes al árbol original donde se muestran los nombres de algunas especies y clases taxonómicas, y así como el dominio filogenético. h) Subárbol de las secuencias de los tres dominios de THDPS. Se marcan con estrellas rojas las 14 secuencias de Eukarya que se encuentran aisladas entre las secuencias de Bacteria. Las ramas marcadas por el programa Notung como probables raíces se resaltan con líneas punteadas en rojo. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

### 4.3. Caracterización de secuencias ancestrales

#### 4.3.1. Ancestros de THDPS y NAL

Se seleccionaron 195 secuencias que incluyeron THDPS de los tres dominios, NAL de Bacteria y Eukarya, KDGA de Archaea y HOGA de Eukarya, para construir un nuevo árbol filogenético a partir del cual se eligieron los nodos de interés que pudieran tener una actividad catalítica incipiente: N6 representa el ancestro hipotético de todas las secuencias de THDPS, N7 y N8 son los ancestros hipotéticos más profundos del árbol, N9 es el ancestro hipotético de NAL, N104 es el ancestro hipotético de KDGA de Archaea y N109 es el ancestro hipotético de HOGA de Eukarya (Figura 18). El alineamiento estructural se llevó a cabo con las estructuras tridimensionales de THDPS (PDB: 1DHP) y NAL (PDB: 1NAL) de *E. coli*.

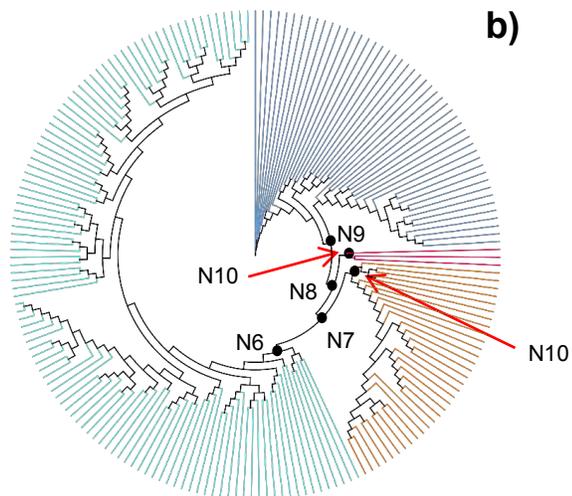
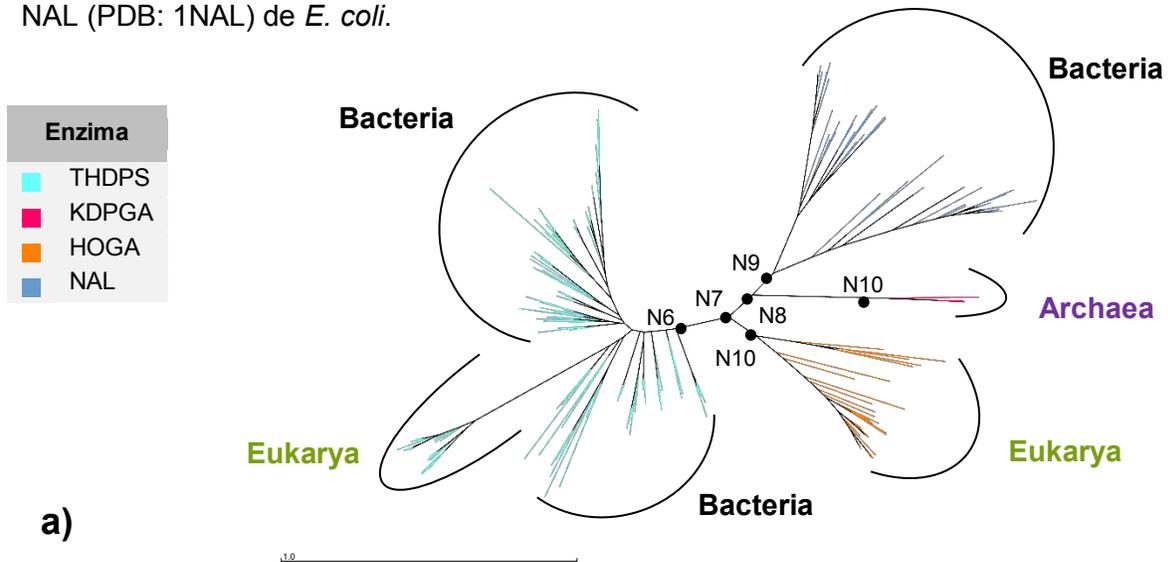
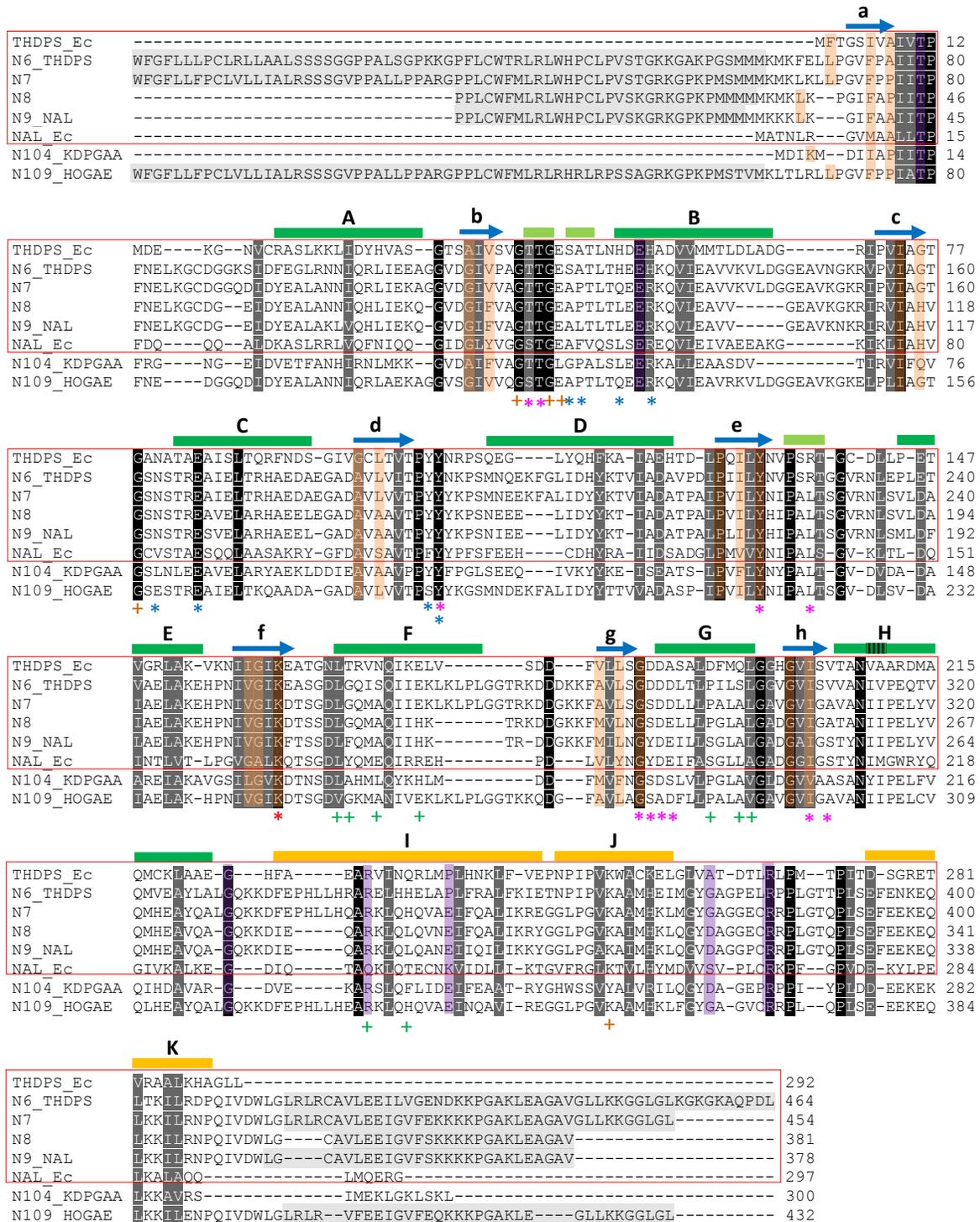


Figura 18. Árbol filogenético de las secuencias utilizadas para obtener los ancestros de NAL y THDPS. a) Vista radial; b) Vista circular. Se marcan los nodos ancestrales de interés. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

Las secuencias que se usaron para el modelaje y análisis final fueron las reconstruidas por medio del método “joint” del servidor DataMonkey, debido a que la reconstrucción “marginal” de este servidor arrojó secuencias con una longitud de 521 residuos de aminoácidos que presentaron regiones de poliprolinas y a que el servidor de FastML nunca terminó el análisis. Por otro lado, la longitud de las secuencias seleccionadas (300 – 464 residuos) se asemejaba de mejor forma a la longitud de THDPS y NAL de *E. coli* (292 y 297, respectivamente). Las regiones iniciales y terminales de las secuencias de los nodos ancestrales no se modelaron debido a que no se encontraron templados con los cuales compararlas.

El alineamiento múltiple (Figura 19) de las secuencias de THDPS y NAL de *E. coli* con los seis nodos ancestrales muestra que las secuencias hipotéticas presentaron tanto el dominio N-terminal que comprende al barril ( $\beta/\alpha$ )<sub>8</sub>, así como el dominio C-terminal que comprende, principalmente, de las tres hélices  $\alpha$ . En primer lugar se observa que la mayor parte de los residuos conservados se presentan en hebras  $\beta$ , debido a que forman parte del núcleo hidrofóbico del barril, en comparación con los residuos que conforman las hélices  $\alpha$ .

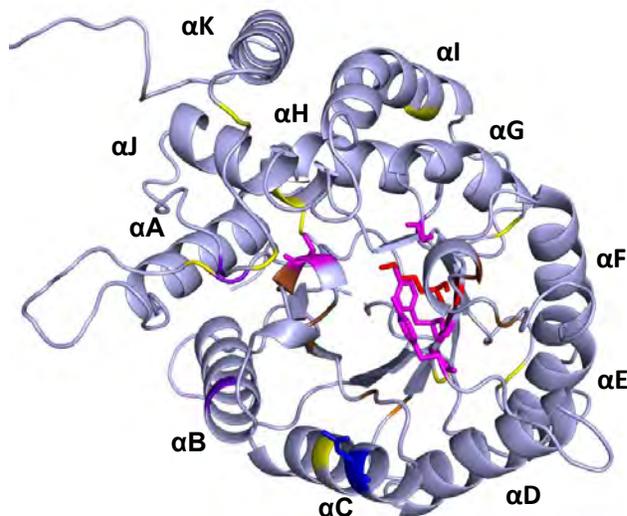
La **Lys** catalítica, al igual que THDPS y NAL, se encontró en la hebra f de todos los nodos ancestrales (Figura 20). Los residuos de la triada catalítica implicados en la unión de piruvato y en el transporte de protones entre el sitio activo y el solvente se conservaron en todas las secuencias (Dobson, Valegård, & Gerrard, 2004): tanto la **Tyr** de la hebra e, como la **Tyr** del loop d-D (que también está presente en el sitio de unión a lisina en THDPS), y la **Thr** presente en el motivo GxxGE. Las **Gly** del motivo GxxGE se conservaron en todos los nodos debido a que juegan un papel estructural importante en la estabilidad del motivo; sin embargo, únicamente en NAL se presentó como un loop, ya que en THDPS y en los nodos ancestrales se encontró en una pequeña hélice  $\alpha$ , e incluso, los ancestros presentaron una hélice “extra” justo después del motivo. El primer residuo del motivo GxxGE correspondió a **Thr** en casi todos los nodos, al igual que THDPS, mientras que fue **Ser** para el caso del nodo N109, al igual que NAL. El residuo **Glu** del motivo se presentó en todos los nodos menos en N104; este residuo forma un puente salino con la **Lys** de la hélice J para estabilizar la hélice/loop (Lawrence et al., 1997), aunque esta última cambió por una **Tyr** en N104.



**Figura 19.** Alineamiento de las secuencias de THDPS y NAL de *E. coli*, y de los nodos ancestrales. La estructura secundaria está representada por bloques (hélices  $\alpha$ ) y flechas (hebras  $\beta$ ). Las regiones sombreadas (inicial y terminal) no fueron modeladas. Los residuos sombreados con naranja pertenecen a las capas del núcleo hidrofóbico, y los residuos sombreados con morado están involucrados en la fijación y movilidad del dominio C-terminal. Los rectángulos rojos resaltan las secuencias involucradas en la vía de interconversión entre THDPS y NAL. Claves: \* lisina catalítica; \* residuos catalíticos; \* residuos involucrados en la unión de lisina; + residuos ligados a características estructurales; + residuos en la zona de interface dímero-dímero. Figura realizada en Color Align Conservation (Stothard, 2000).

Otros residuos importantes e involucrados en la estabilidad de la unión de sustratos son: la **Arg**, importante para la unión de L-ASA en THDPS (Dobson, Griffin, Jameson, & Gerrard, 2005), se localizó en una pequeña hélice  $\alpha$  en el loop e-E que comienza con una **Pro** y termina con una **Gly** conservadas en THDPS y en el nodo N6, y a pesar de que esta hélice estuvo presente en todos los nodos analizados, los demás nodos presentaron **Leu** al igual que NAL (quien no presenta la hélice  $\alpha$ ); la **Gly** de la hebra g, encargada de unirse al grupo carboxilo del respectivo sustrato de cada enzima, se conservó tanto en las secuencias actuales como en los nodos; el par de residuos **Asp** y **Glu** de NAL localizados en la hélice G e importantes para la estabilización de Neu5Ac se presentaron en NAL y en los nodos N8 y N9, mientras que el par de residuos **Asp** que ayudan en la unión de L-ASA sólo se presentaron en THDPS y N6, sin embargo, el residuo **Asp** que está presente en ambas enzimas, se conservó en todos los nodos a excepción de N109; y, finalmente, el residuo **Ile** encontrado en la hebra h, al cual se le ha atribuido la liberación de energía acumulada por los cambios conformacionales que ocurren debido a la unión de L-ASA en THDPS, así como la polarización de una parte del piruvato además de una molécula de agua que realiza un posible ataque nucleofílico que hidroliza la unión de la **Lys** catalítica con el sustrato (Dobson et al., 2008), se conservó en todos los nodos, a excepción de N104 donde fue cambiado por **Val**. La **Ser** importante para que el residuo **Ile** lleve a cabo sus funciones sólo estuvo presente en THDPS, NAL y los nodos N6 y N9.

De los residuos presentes en el sitio de unión a lisina en THDPS (Dobson et al., 2005) destacan la **His** de la hélice B que, al igual que NAL, es reemplazada por **Arg** en casi todos los nodos, y la **Glu** de la hélice C que se conservó en todos los nodos. El dominio C-terminal se encuentra fijado al barril gracias a los residuos **Thr** (loop a-A), **Glu** (hélice B) y **Arg** (loop J-K) conservados en todos los nodos (Barbosa et al., 2000), aunque también participan otros que se encuentran en las hélices I y J. Por otro lado, el residuo **Gly** conservado en todas las secuencias que se encuentra entre las hélices H-I puede proporcionarle movilidad al dominio C-terminal (Mirwaldt et al., 1995). Con respecto a los residuos presentes en la interface dímero-dímero (Dobson et al., 2005), destacan la **Leu** de la hélice F que fue cambiada por **Val** sólo en N109, la **Leu** de la hélice G y la **Arg** de a hélice I presentes en todos los nodos ancestrales. Otros residuos involucrados en algún aspecto estructural son la **Gly** conservada del loop c-C que podría estar involucrada en la formación del canal que va del exterior al sitio activo (Mirwaldt et al., 1995).



**Figura 20.** Residuos conservados del alineamiento entre THDPS y NAL de *E. coli*, y los nodos ancestrales presentados en el modelo de N7 (ancestro hipotético más profundo de la Figura 19). Rojo: lisina catalítica; magenta: residuos implicados en catálisis; azul: residuos presentes en sitio de unión a lisina; morado: residuos relacionados a dominio C-terminal; naranja: residuos del núcleo hidrofóbico; amarillo: residuos sin función aparente.

Los alineamientos estructurales muestran que los ancestros hipotéticos compartieron buenos valores de RMSD con las estructuras de NAL y THDPS (Tabla 5; Figura 21), aunque la mayor parte de los nodos tuvo un mayor parecido estructural con THDPS. Todas las estructuras ancestrales presentaron dos pequeñas hélices  $\alpha$  entre la hebra b y la hélice B, es decir, presentaron una hélice  $\alpha$  extra que la que tiene THDPS en esa región. La posición de la última hélice del barril y las tres hélices del dominio C-terminal fue la principal característica en la que se pudo distinguir el parecido estructural con THDPS o NAL reflejado en los valores globales de RMSD: en las estructuras de N6, N7, N8 y N109 la posición de estas hélices se pareció más a la posición de las mismas hélices de THDPS, donde los valores de RMSD fueron más cercanos a cero al compararlos contra THDPS; en la estructura N9 se pareció más a NAL, mientras que en N104 no hubo parecido con alguna de las dos. Otras características fueron que las terminaciones amino y carboxilo de la cadena se extendieron más (a excepción de N104); a N6, N7 y N104 les hizo falta una hebra diferente a cada uno del barril; y, finalmente, N8 presentó dos hebras extras antes de la hélice A, mientras que N109 tuvo dos hélices antes de la hélice A, una más grande que la otra.

**Tabla 5.** Valores globales de RMSD y porcentaje de identidad entre pares de secuencias obtenidos de la comparación de THDPS y NAL de *E. coli* contra los nodos ancestrales.

	N6		N7		N8		N9		N104		N109	
	RMSD (Å)	% Id										
<b>THDPS</b>	0.832	40.87	0.787	32.17	0.807	30.19	1.363	29.75	1.722	26.76	0.742	29.36
<b>NAL</b>	1.553	26.09	1.474	33.04	1.451	38.94	0.405	42.45	1.497	27.42	1.641	30.54

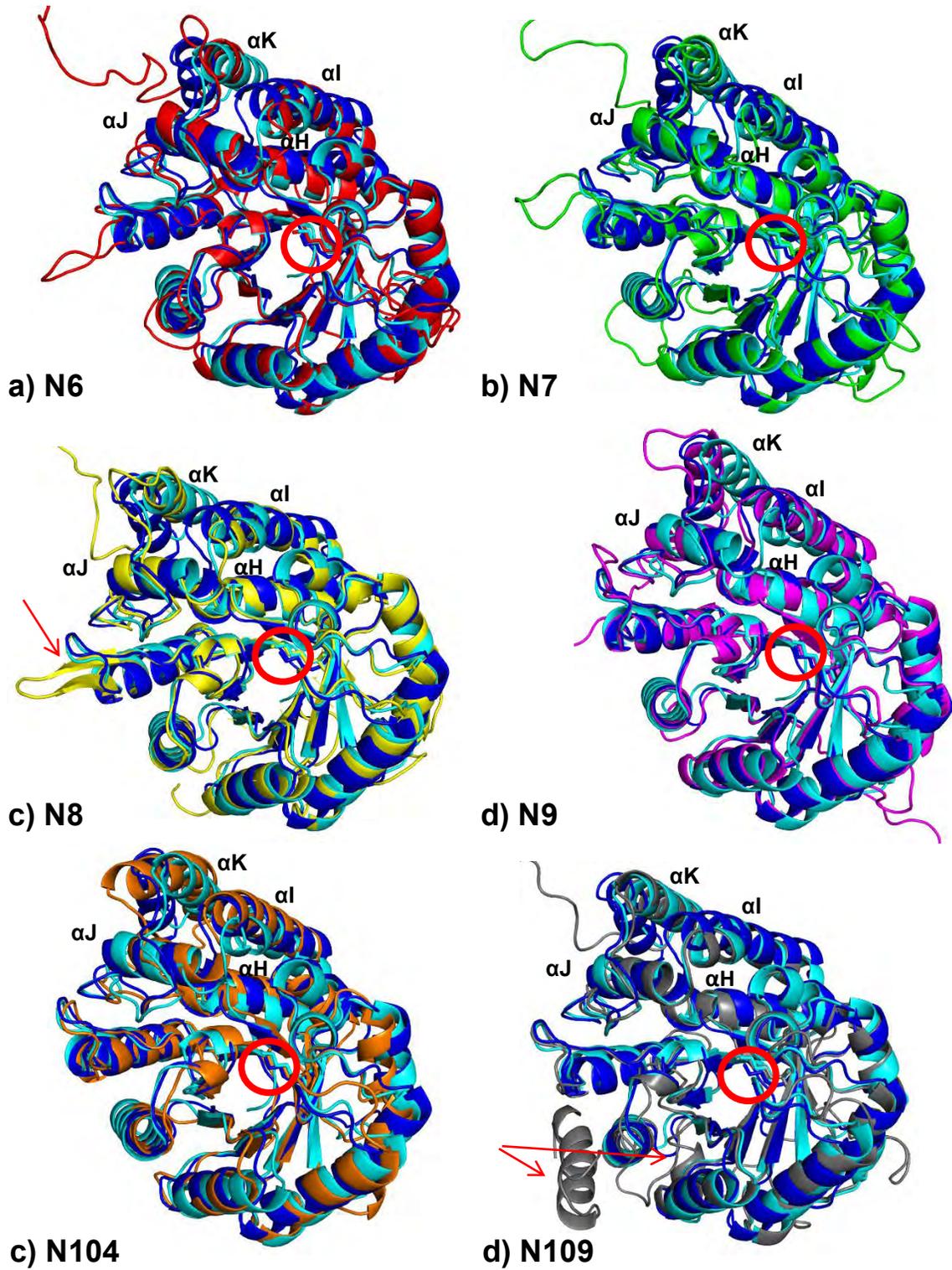
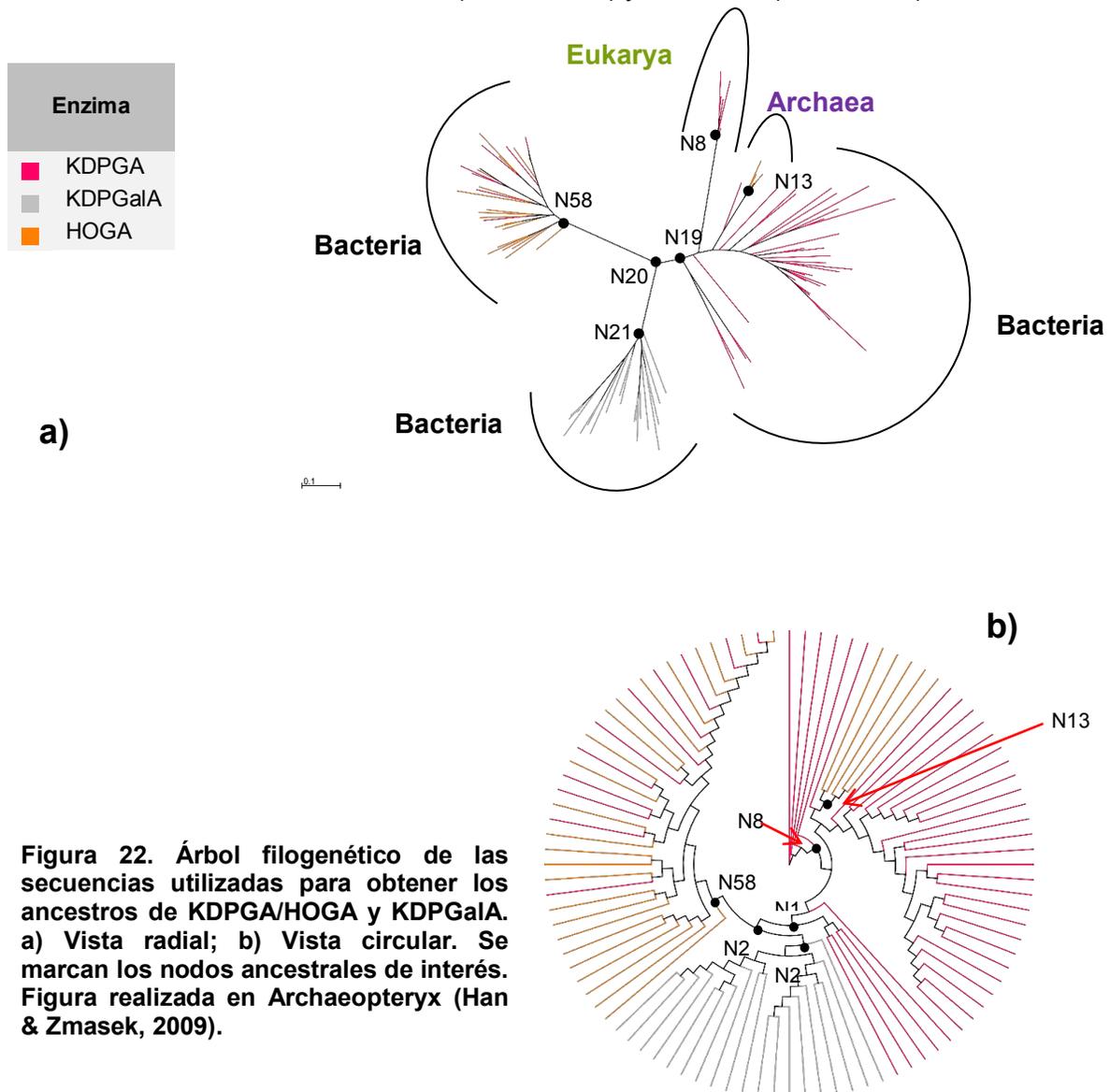


Figura 21. Alineamiento estructural de THDPS y NAL con los nodos ancestrales. La lisina catalítica se señala con un círculo rojo, mientras que las estructuras secundarias únicas de cada estructura con flechas. THDPS: cyan; NAL: azul; N6: rojo; N7: verde; N8: amarillo; N9: magenta; N104: naranja; N109: gris. Figuras realizadas en PyMOL (Schrödinger, 2010).

### 4.3.2. Ancestros de KDPGA/HOGA y KDPGalA

Para construir el árbol filogenético de ancestros se seleccionaron 95 secuencias aplicando la misma metodología descrita para la selección de secuencias de THDPS y NAL. Las secuencias que incluyó el nuevo árbol correspondieron a las enzimas KDPGA/HOGA y KDPGalA de Bacteria, KDPGA de Eukarya, HOGA de Archaea y la única secuencia de Eukarya de KDPGalA. Los nodos elegidos fueron: N8, ancestro hipotético de KDPGA de Eukarya; N19, ancestro hipotético de KDPGA de Bacteria y Eukarya; N20, el ancestro más profundo del árbol; N21, ancestro hipotético de KDPGalA; N58, ancestro hipotético de algunas KDPGA/HOGA de Bacteria; y N134, ancestro hipotético de HOGA de Archaea (Figura 22). El alineamiento estructural se llevó a cabo con las estructuras tridimensionales de KDPGA/HOGA (PDB: 1EUN) y KDPGalA (PDB: 2V81) de *E. coli*.



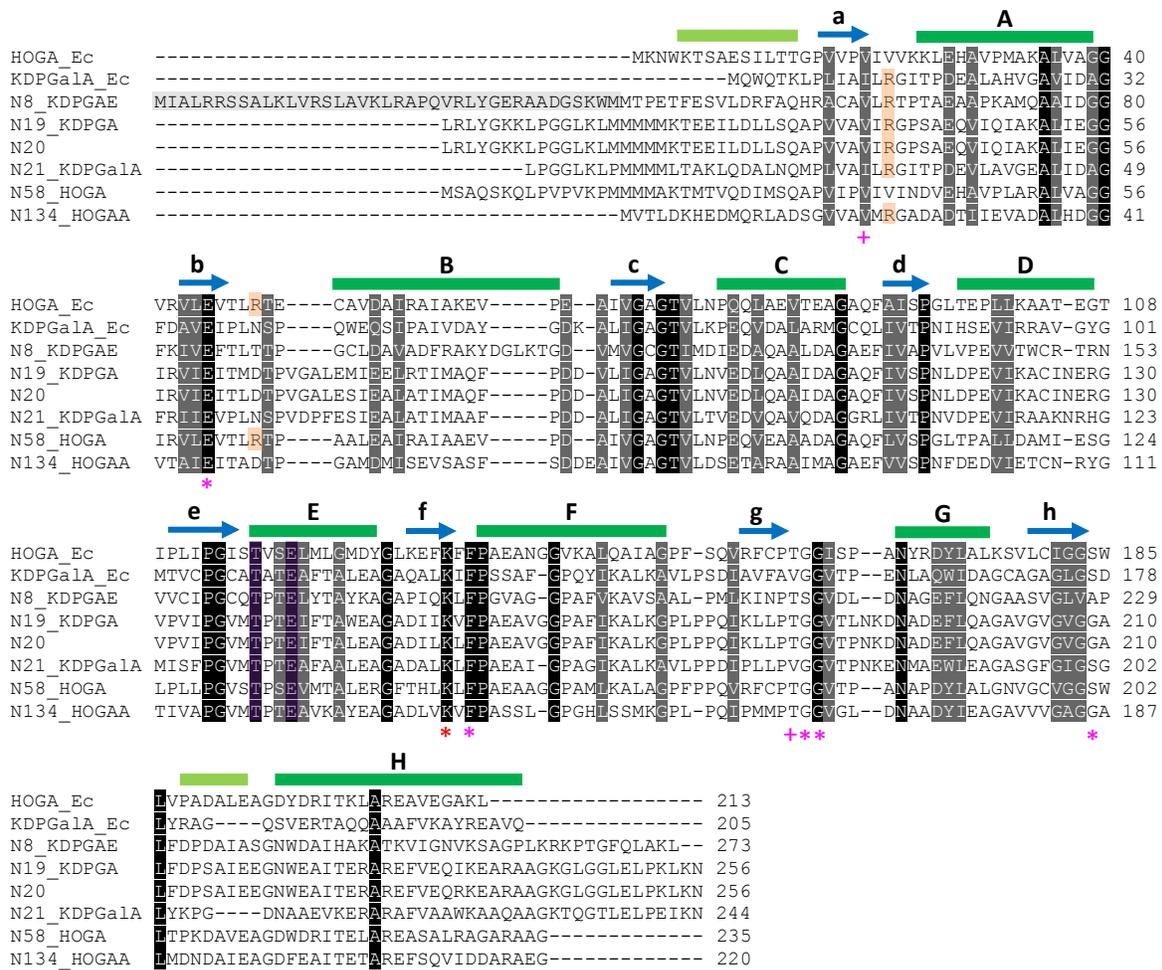
**Figura 22. Árbol filogenético de las secuencias utilizadas para obtener los ancestros de KDPGA/HOGA y KDPGalA. a) Vista radial; b) Vista circular. Se marcan los nodos ancestrales de interés. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).**

Al igual que lo descrito con anterioridad para los ancestros de THDPS y NAL, se eligió la reconstrucción “joint” del servidor DataMonkey, a partir de la cual se obtuvieron longitudes de 220 a 273 residuos de aminoácidos (Figura 23). Las secuencias de todos los nodos ancestrales conservaron los residuos implicados en la catálisis (Allard, Grochulski, & Sygusch, 2001; Walters et al., 2008) (Figura 24): **Glu** en la hebra b, **Thr** en el loop c-C, **Lys** en la hebra f, y **Phe** en el loop f-F; de igual manera, el residuo **Arg** se conservó en todos los nodos aunque en el nodo N58 se localizó en el loop b-B, al igual que KDPGA, mientras que en los demás se encontró en el loop a-A, donde se encuentra en KDPGalA. De los residuos encargados de unir al grupo fosfato, sólo se conservó la **Gly** del loop g-G. Por otro lado, el residuo **Thr** que se encuentra en el loop g-G de KDPGA, encargado potencialmente del control de la estereoselectividad de la enzima, se conservó en todos los nodos, excepto en N21 donde se presenta **Val** en su lugar, al igual que KDPGalA; los residuos **Val** o **Ile**, posiblemente encargados de la estereoselectividad, estuvieron presentes en todos los nodos. Los residuos **Thr** y **Glu** de la hélice E, encargados de estabilizar el trímero (Bell et al., 2003), se conservaron de igual manera.

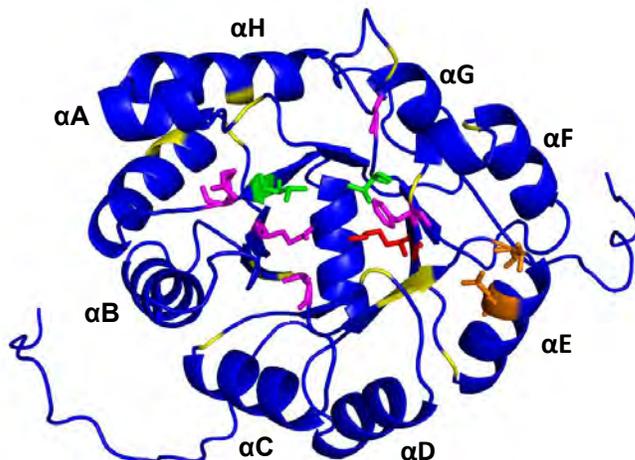
A nivel estructural (Tabla 6; Figura 25), y a diferencia de THDPS y NAL, el protómero, tanto de KDPGA como KDPGalA, está formado por un barril ( $\beta/\alpha$ )<sub>8</sub> que se diferencia principalmente en que KDPGA posee una hélice  $\alpha$  en la parte final del extremo N-terminal que “tapa” el hueco que forma el barril. Todos los nodos ancestrales presentaron esta hélice  $\alpha$ . Por otro lado, todos presentaron una pequeña hélice  $\alpha$  antes de la hélice H, a excepción de N21, donde se presentó antes de la hélice G y donde también se observa que la hélice H está del mismo tamaño que la de KDPGalA. En N8 se presentaron dos hélices antes de la hélice H y una hélice después de ésta. Otro aspecto estructural que destaca en los alineamientos es que se muestra una discontinuación de la hélice F en todos los nodos, a excepción de N8, donde la hélice F es más corta que la hélice F de las estructuras actuales.

**Tabla 6. Valores globales de RMSD y porcentajes de identidad entre pares de secuencias obtenidos de la comparación de KDPGA y KDPGalA de *E. coli* contra los nodos ancestrales.**

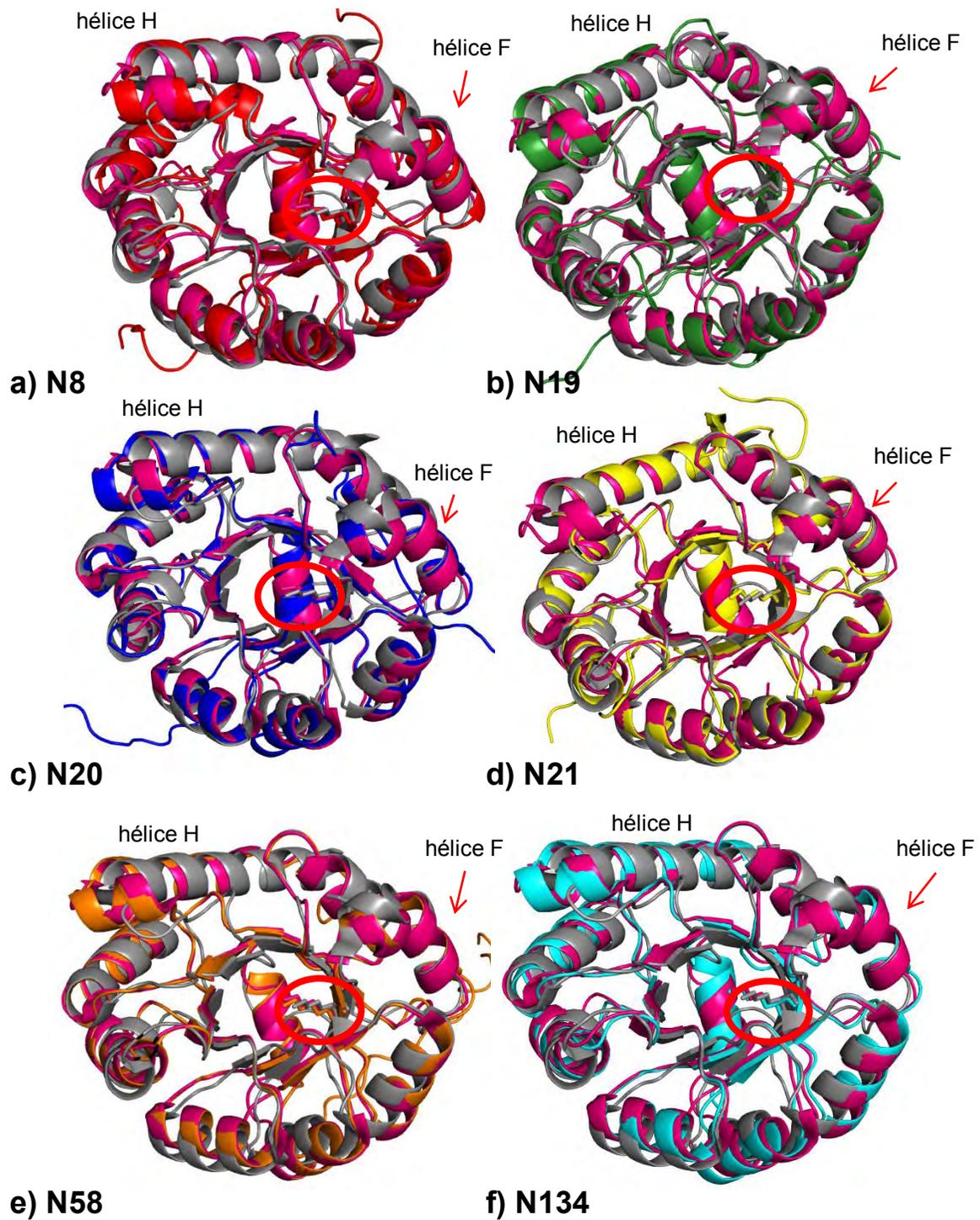
	N8		N19		N20		N21		N58		N134	
	RMSD (Å)	% Id										
<b>KDPGA</b>	1.211	24.09	0.606	37.5	0.574	38.39	1.47	28.76	0.369	60.37	0.737	33.33
<b>KDPGalA</b>	1.341	29.52	1.753	35.48	1.666	37.79	0.338	52.83	1.503	31.6	1.51	34.47



**Figura 23.** Alineamiento de las secuencias de KDPGA/HOGA y KDPGalA de *E. coli*, y de los nodos ancestrales. La estructura secundaria está representada por bloques (hélices  $\alpha$ ) y flechas (hebras  $\beta$ ). Las regiones sombreadas (inicial y terminal) no fueron modeladas. Los asteriscos magenta marcan los residuos relacionados a la catálisis, mientras que el asterisco rojo marca la lisina catalítica. Los residuos sombreados de naranja también participan en la catálisis, y los sombreados de morado están involucrados en la estabilización del trímero. Los residuos marcados con + están relacionados con estereoselectividad de la enzima. Figura realizada en Color Align Conservation (Stothard, 2000).



**Figura 24.** Residuos conservados del alineamiento entre KDPGA/HOGA y KDPGalA de *E. coli*, y los nodos ancestrales presentados en la estructura tridimensional de N20 (ancestro hipotético más profundo de la Figura 21). Rojo: lisina catalítica; magenta: residuos implicados en catálisis; verde: residuos involucrados en la estereoselectividad de la enzima; naranja: residuos involucrados en la estabilización del trímero; amarillo: residuos sin función aparente.



**Figura 25. Alineamiento estructural de KDPGA/HOGA y KDPGalA con los nodos ancestrales. La lisina catalítica se señala con rojo. KDPGA: magenta; KDPGalA: gris; N8: rojo; N19: verde; N20: azul; N21: amarillo; N58: naranja; N134: cyan. Figuras realizadas en PyMOL (Schrödinger, 2010).**

## 5. Discusión

### 5.1. Relaciones evolutivas entre los miembros de la subfamilia de enzimas relacionadas a la NAL

La subfamilia de enzimas relacionadas a la NAL presenta miembros implicados en vías relacionadas al metabolismo de carbohidratos y aminoácidos. Estas vías representan una parte vital de cualquier organismo, ya que sin ellas no serían capaces de obtener la energía necesaria para sobrevivir; por lo tanto, la aparición y divergencia de estas enzimas debió haberse dado hace millones de años en organismos procariontes. Si se observa el árbol filogenético de los tres dominios, los organismos que se encuentran en las partes más profundas, es decir, bacterias y arqueas hipertermófilas presentan con mayor ocurrencia la vía ED (Romano & Conway, 1996), por lo que la aldolasa ancestral pudo haber presentado actividad de KDPGA y, posteriormente, sufrir una duplicación génica que dio origen a la actividad de KDPGalA (Figura 26), ya que ambas enzimas realizan la misma reacción que difiere en la estereoselectividad por su sustrato (KDPG o KDPGal).

La estrecha relación que tienen KDPGA y KDPGalA se puede observar en el árbol resultante, ya que comparten un ancestro común más cercano que con las otras aldolasas (Figura 16). Por un lado se encuentra KDPGA perteneciente a organismos quimioorganótrofos sacarolíticos, que pueden ser anaerobios o aerobios y de vida libre, comensales o parásitos de animales, de las clases Clostridia, Bacilli, Flavobacteria, Ignavibacteria, Synergistia, Fusobacteriia, Spirochaetes y Thermotogae (Figura 17a); también en especies de cianobacterias que se caracterizan por poseer un metabolismo fotolitoautótrofo, así como en algunas especies de algas que pertenecen al fitoplancton como las diatomeas (Figura 17c), donde se ha caracterizado la vía ED en la mitocondria y que el nivel de expresión de KDPGA depende del tiempo y la cantidad de exposición del microorganismo a la luz (Fabris et al., 2012). La presencia del grupo de KDPGA de eucariontes que parece tener un ancestro cercano con las secuencias de KDPGA bacterianas puede deberse a un transporte horizontal que se dio de bacteria a algas, o bien, pudieron haber adquirido la enzima por medio de endosimbiosis. Por el otro lado, se encuentran todas las KDPGalA de bacterias pertenecientes a clases del phylum Proteobacteria, que incluye organismos con una amplia gama de metabolismos bacterianos que van desde fotótrofos hasta quimiolitotróficos, y que pueden ser simbioses de plantas como los géneros *Rhizobium* y *Bradyrhizobium* de la clase

Alphaproteobacteria, o bien, patógenos de animales como la mayor parte de las especies de la clase Gammaproteobacteria; y la KDPGalA que pertenece al insecto *Ceratitits capitata* que pudo haber sido producto de un transporte de genes a partir de una especie de parásito bacteriano (Figura 17g).

Un aspecto que resalta en el árbol obtenido es que las secuencias de HOGA y KDPGA bacterianas se encuentran mayormente separadas; incluso, parece ser que hubo una duplicación génica que dio origen, por un lado, a KDPGalA y, por el otro lado, al ancestro de KDPGA y HOGA (Figura 16b); sin embargo, dentro del clado correspondiente a HOGA, se encontraron algunas secuencias pertenecientes a KDPGA (Figura 17a). La explicación que podría darse para esta ubicación aparentemente incorrecta es que, en realidad, la KDPGA y la HOGA de bacterias es la misma enzima, al igual que como se reporta para *E. coli* (Egan et al., 1992), por lo que la distinción en color sólo fue una característica atribuida de acuerdo a los valores de E-value que se obtuvieron al realizar los BLAST de cada enzima; por lo tanto, las secuencias de KDPGA/HOGA fueron separadas en el árbol debido al organismo al cual pertenecen. De esta manera, las secuencias de KDPGA/HOGA pertenecientes a las bacterias del phylum Proteobacteria se parecen más entre ellas que si se comparan contra las secuencias de KDPGA/HOGA de los organismos de las demás clases.

En la literatura no hay reportes que indiquen que la enzima KDPGA también tenga actividad de HOGA en arqueas, a pesar de que las secuencias depositadas en el NCBI tengan el nombre que hace referencia a ambas actividades; sin embargo, en el árbol resultante aparecen dos grupos definidos de secuencias de KDPGA y HOGA de arqueas que se encuentran separados (Figura 17c; 17e). Las secuencias que fueron identificadas con la actividad de KDPGA comparten un ancestro en común con NAL de bacterias y pertenecen a arqueas hipertermófilas de las clases Thermoprotei y Thermoplasmata, donde la glucosa es metabolizada por medio de vías modificadas de ED y EMP (Siebers & Schönheit, 2005) (Figura 17d).

Inicialmente, se había caracterizado la vía modificada de ED no fosforilativa para arqueas hipertermófilas como *S. solfataricus*, *T. acidophilum* y *T. tenax*, por lo que se sabía que la KDPGA que participa en esta vía modificada catalizaba la ruptura reversible de sustratos no fosforilados, sin embargo, en estudios posteriores se propuso una nueva vía modificada semi-fosforilativa para *S. solfataricus* y *T. tenax*, debido a que la KDPGA de ambos organismos es capaz de realizar la ruptura reversible de sustratos fosforilados;

incluso, la enzima de *S. solfataricus* puede realizar la actividad de KDPGalA aceptando sustratos no fosforilados (Ahmed et al., 2005; Buchanan et al., 1999).

El otro grupo de secuencias pertenecientes a arqueas de la clase Halobacteria fue identificado como HOGA de acuerdo al nombre presente en la base de datos del NCBI (Figura 17c), aunque en realidad estudios de resonancia magnética nuclear de carbono-13 y enzimáticos, así como análisis de microarreglos de DNA, revelaron que estos organismos son capaces de degradar la glucosa mediante la vía modificada ED semi-fosforilativa en condiciones aeróbicas (Siebers & Schönheit, 2005); es decir, la enzima que fue catalogada como HOGA, tiene actividad de una enzima KDPGA que acepta sustratos fosforilados. Esta enzima presente en arqueas halófilas pudo haber sido adquirida de bacterias que comparten la misma vía modificada ED por medio de un transporte horizontal de genes (Figura 17c), debido a que la enzima que cataliza la misma reacción para sustratos no fosforilados presente en los grupos de arqueas termófilas no comparte similitud significativa con la KDPGA bacteriana (Buchanan et al., 1999) y que, además, difieren en su estado oligomérico (en arqueas es un homotetrámero, mientras que en bacterias es un trímero).

Cercano al grupo de KDPGA de arqueas, se encuentran las secuencias de NAL (Figura 16b). Esta enzima está presente tanto en bacterias como en vertebrados, involucrada en la ruptura y síntesis de ácido siálico, aunque ambos grupos de organismos poseen enzimas que catalizan la síntesis de éstos  $\alpha$ -cetoácidos, por lo que la participación de NAL no es tan relevante en este aspecto (Angata & Varki, 2002); sin embargo, la ocurrencia de esta enzima es generalmente en bacterias patógenas asociadas a la mucosa de los tractos digestivo y respiratorio, así como en la cavidad oral de vertebrados, debido principalmente a que les permite utilizar el ácido siálico como fuente de carbono y energía. En el subárbol (Figura 17d) se observa que las secuencias de NAL se agrupan en dos clados de bacterias que comparten el hecho de ser comensales o parásitos de animales, siendo, la mayoría, anaerobios facultativos; sin embargo, hay unas cuantas secuencias que le pertenecen a bacterias de vida libre que habitan suelos y aguas, mientras que otras forman relaciones simbióticas con plantas como las especies del género *Sinorhizobium*.

Es probable que la NAL haya sido producto de varios transportes horizontales de genes entre diversos linajes de bacterias debido a que la filogenia de esta enzima no concuerda con la de los organismos donde se encuentra, por ejemplo, a pesar de que la familia

Pasteurellaceae se encuentra relacionada con la familia Enterobacteraceae, sus enzimas aparecen agrupadas en dos clados distintos. De la misma manera, como se observa en el subárbol obtenido, las secuencias de vertebrados comparten un ancestro en común con miembros de la clase Gammaproteobacteria, como *Yersinia pestis* y *Vibrio cholerae*, así como la familia Pasteurellaceae comparte un ancestro en común con la secuencia de *T. vaginalis*. Se ha propuesto que el ancestro de este protozoario parásito del epitelio de animales pudo haber adquirido NAL a partir de un ancestro de la familia Pasteurellaceae por medio de un transporte horizontal de genes, y de esta manera poder adaptarse a utilizar el ácido siálico presente en la superficie celular del epitelio de vertebrados como fuente de carbono y energía, de la misma manera como ocurre en bacterias (de Koning, Brinkman, Jones, & Keeling, 2000). En el caso de la NAL de vertebrados, también pudo haber surgido en este linaje por medio de un transporte horizontal de genes a partir de un ancestro de bacterias patógenas de la clase Gammaproteobacteria.

La enzima ancestral que dio origen a NAL debió haber surgido de una duplicación génica en procariontes termófilos, ya que se ha observado que las NAL de algunos organismos tanto procariontes como *C. perfringens* y *E. coli*, y eucariontes como *S. scrofa*, tienen una temperatura óptima de 65 – 70°, 80° y 70 – 80°, respectivamente, por lo que es considerada una enzima termostable (propiedad característica de proteínas tetraméricas) (Schauer, Sommer, Krüger, van Unen, & Traving, 1999). Debido a la relación estrecha que se observa entre KDPGA de arqueas y la NAL de bacterias, se supondría que el origen se dio en arqueas, donde la enzima que dio origen a ambas tenía una actividad catalítica incipiente para ambas funciones (Figura 26).

Al parecer THDPS y HOGA de eucariontes se encuentran estrechamente relacionadas, esto podría deberse a que ambas están involucradas en el metabolismo de aminoácidos. Se sabe que THDPS está presente en organismos que utilizan la vía derivada del aspartato para sintetizar lisina conocida como la vía del ácido diaminopimélico (DAP): bacterias, protozoarios parecidos a hongos (Oomycota e Hypochytriumycota), plantas vasculares (cloroplastos), y arqueas (Velasco, Leguina, & Lazcano, 2002; Vogel, 1964); mientras que se ha caracterizado a HOGA involucrada en la vía de degradación de la 4-hidroxi-L-prolina, que es uno de los principales componentes del colágeno en animales y hongos (Celerin et al., 1996). Además, se ha propuesto que HOGA sea parte de una ruta de degradación de la L-homoserina (Lane, Shapley, & Dekker, 1971), un intermediario que participa en la biosíntesis de treonina, metionina e isoleucina a partir del aspartato.

Después de la formación de L-ASA (sustrato de THDPS en la vía DAP), otras enzimas son las encargadas de formar L-homoserina para seguir sus propias vías de biosíntesis.

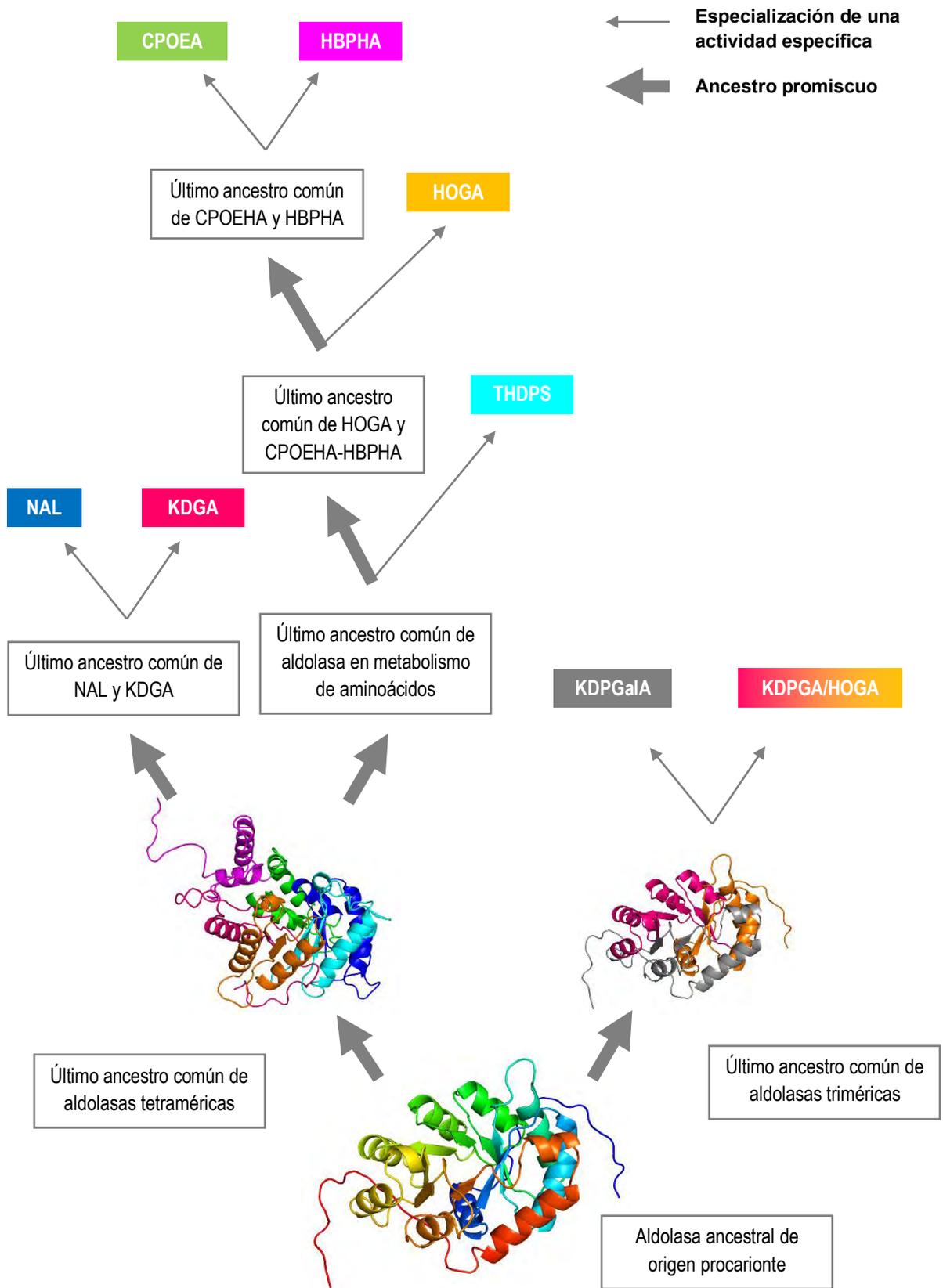
La enzima ancestral que dio lugar a estas dos enzimas pudo haber estado involucrada en primer lugar en la degradación de L-homoserina en un organismo procarionte, para luego sufrir una duplicación que dio origen a la actividad de THDPS, es decir, este organismo era capaz de sintetizar lisina por medio de la vía DAP. Posteriormente, el último ancestro en común de hongos y animales adquirió la capacidad de sintetizar lisina por medio de esta vía por herencia mitocondrial pero después de un tiempo la perdió, por lo que los hongos adaptaron una nueva forma de obtenerla por medio de la vía del ácido  $\alpha$ -aminoadípico (AAA), mientras que los animales comenzaron a obtenerla por medio de la dieta. Esta pérdida de función pudo haber sido una subfuncionalización de la enzima para especializarse únicamente en la degradación de 4-hidroxi-L-prolina en estos dos grupos de eucariontes.

Antes de que el último ancestro de hongos y animales adquiriera HOGA, la enzima debió haberse duplicado en el organismo procarionte para dar origen al ancestro de CPOEA y HBPHA (Figura 26), que debió haber estado involucrada primeramente en la degradación de 4-hidroxi-L-prolina. Actualmente, se sabe que especies del género *Pseudomonas* pueden utilizar la 4-hidroxi-L-prolina como única fuente de carbono; sin embargo, aunque HOGA no es una de las enzimas implicadas en esta vía, otra enzima relacionada a la subfamilia de la NAL que no utiliza piruvato como sustrato (aunque su actividad es inhibida competitivamente por éste), conocida como 1-pirrolina-4-hidroxi-2-carboxilato desaminasa que cataliza la hidrólisis de Pyr4H2C a 2,5-dioxopentanoato liberando amonio, está implicada en la vía de degradación (Watanabe et al., 2012). El ancestro de CPOEA y HBPHA debió haberse duplicado y, por un lado, se especializó en la actividad de desaminasa involucrándose en la vía degradativa de 4-hidro-L-prolina, mientras que la otra copia debió haberse involucrado en las vías de degradación de PAHs (Figura 17f).

En el subárbol de THDPS se encuentran varios clados de eucariontes que agrupan secuencias que se encuentran en los cloroplastos de algas y plantas mayormente relacionadas con bacterias Gram +; secuencias que se encuentran en las mitocondrias de protistas de vida libre como *P. tricornutum* y de parásitos como *Toxoplasma gondii* que pudieron haberla adquirido por herencia del ancestro de bacterias de la clase Betaproteobacteria; y secuencias de oomicetos que forman un grupo hermano con secuencias pertenecientes a bacterias de distintas clases taxonómicas. Las enzimas de la

ruta biosintética de la lisina en oomicetos han sido adquiridas por medio de transporte horizontal y añadidas al genoma de estos protistas (Morris et al., 2009), por lo que es probable que su THDPS sea de origen bacteriano. Sin embargo, hay unas cuantas secuencias pertenecientes animales que no están agrupadas con las demás. Las secuencias del insecto *Aedes aegypti* y el placozoo *Trichoplax adhaerens* comparten un ancestro en común con algunas bacterias Gram + y las secuencias de cloroplastos de algas y plantas; mientras que las secuencias de los ácaros *Ixodes scapularis* y *Metaseiulus occidentalis*, así como la del poliqueto *Capitella teleta* se agrupan dentro del clado correspondiente a la clase Alphaproteobacteria. Probablemente sean los indicios de que alguna vez los animales tuvieron la capacidad de sintetizar lisina, es decir, estos organismos aún conservan el gen que codifica para una THDPS que ya no es funcional debido a que pudo haber perdido una región de regulación para su expresión.

De acuerdo a los resultados obtenidos en el programa Notung, la raíz puede ser colocada en varias ramas del árbol que fueron calificadas con el mismo puntaje y que, por lo tanto, resultaron ser igual de parsimoniosas en cuanto a número de duplicaciones y pérdida de genes. Por este motivo, y debido a la importancia de las rutas metabólicas a las cuales pertenecen la mayoría de las aldolasas de esta subfamilia, no se puede establecer una polaridad absoluta que nos indique cuál de las enzimas surgió primero. De hecho, algunas de estas enzimas pertenecen a vías alternativas relacionadas con la obtención de energía, como KDPGA y KDPGaIA, que se encuentran presentes en organismos principalmente procariontes. Sin embargo, el árbol podría dividirse en dos grupos principales: por un lado se encuentran las enzimas triméricas KDPGA y KDPGaIA de bacterias involucradas en el metabolismo energético, y por el otro lado, las enzimas KDPGA de arqueas, NAL, THDPS, HOGA de eucariontes, CPOEA y HBPHA, que son tetraméricas en su mayoría y que están involucradas en el metabolismo de aminoácidos y sus derivados, principalmente. Debido a que las actividades catalíticas de HOGA y KDPGA se encuentran en ambos lados del árbol filogenético, se podría decir que la enzima ancestral que dio origen a toda esta subfamilia tenía ambas actividades, es decir, era promiscua catalíticamente hablando. Posteriormente, sufriría una duplicación génica que daría lugar al ancestro de las aldolasas triméricas actuales y al ancestro de las aldolasas tetraméricas actuales (Figura 26). La raíz del árbol podría ser colocada entre estos dos grupos debido a que, además de que el puntaje de Notung para esa rama fue de los más altos y que el método de la raíz en la rama más larga coloca la raíz en esta rama, hay una diferencia en el estado oligomérico de las enzimas.



**Figura 26. Propuesta del surgimiento y diversificación de la subfamilia de enzimas relacionadas a la NAL.**

## 5.2. Identificación de residuos de aminoácidos por medio de la reconstrucción de proteínas ancestrales

La interconversión de una actividad enzimática a otra, además del enfoque actual del diseño racional de proteínas, requiere un análisis vertical que tome en cuenta la comparación de las secuencias actuales con las secuencias ancestrales. El objetivo de este análisis es identificar los residuos de aminoácidos implicados en la estabilidad o función de la enzima que no pueden ser identificados mediante el diseño racional. El método para llevar a cabo la identificación de estos residuos es hacer un alineamiento múltiple de las secuencias ancestrales y las secuencias actuales de las cuales se quiere realizar la interconversión de actividades; en este caso se usó THDPS, NAL, KDGP y KDPGaA de *E. coli*. Posteriormente, se identifican los residuos conservados o que han sido sustituidos por residuos similares. Estos residuos son blancos potenciales de mutaciones para comprobar si pueden ser responsables de la interconversión de actividad.

Los ancestros obtenidos a partir de una pequeña cantidad de secuencias del árbol principal conservaron los residuos implicados en la catálisis, unión a sustratos, estereoselectividad (en el caso de las triméricas) y estabilidad de la estructura cuaternaria. Otros residuos que se conservaron no tienen una función aparente indicada en la literatura (Figuras 20 y 24), sin embargo, podrían estar implicados en el mantenimiento y estabilidad del barril debido a que algunos de ellos fueron prolinas y glicinas que aparecen en las vueltas iniciales y finales de las estructuras secundarias (X. Yang, Kathuria, Vadrevu, & Matthews, 2009). Los demás residuos deben ser analizados para identificar su función potencial, y si es que están involucrados indirectamente en la catálisis o en la estereoselectividad de las enzimas actuales. Por otro lado, algunas características resaltaron gracias a la comparación estructural de los nodos ancestrales con las estructuras tridimensionales de las secuencias actuales:

- 1) La posición de las hélices del dominio C-terminal y la hélice H del barril variaron de ancestro a ancestro, pareciéndose más a THDPS en la mayoría de éstos, lo que podría indicar dos cosas: que la enzima ancestral (al menos de THDPS, NAL, HOGA y KDPGA) tenía una estructura muy parecida a THDPS, y que las últimas tres hélices cumplen un papel importante en la catálisis o en el mantenimiento estructural de la enzima.

- 2) Las hélices localizadas en el loop b-B y el loop e-E que se conservaron en todas las estructuras ancestrales contienen algunos residuos conservados importantes para la catálisis de cada enzima.
- 3) Todas las estructuras de los ancestros de las enzimas triméricas conservaron la hélice “extra” que posee KDPGA/HOGA, incluso el nodo ancestral más antiguo, lo que sugiere y apoya la hipótesis de que KDPGA y HOGA fueron las primeras actividades catalíticas de la enzima ancestral de esta subfamilia.
- 4) La hélice localizada en el loop h-H de las enzimas triméricas se conservó en todos los nodos ancestrales, lo que indica que los residuos presentes en ella podrían tener relevancia.

## 6. Conclusiones y perspectivas

En la actualidad, la reconstrucción filogenética ha sido una herramienta útil que nos permite abordar problemas referentes al campo del rediseño de proteínas desde un punto de vista evolutivo, utilizando como referencia los experimentos que ya ha hecho la naturaleza para crear los andamiajes necesarios para las diversas actividades catalíticas que vemos hoy en día; sin embargo, es una herramienta que depende de la metodología que se emplee para llevarla a cabo, por lo que es necesario tener conocimientos previos sobre las proteínas con las que se piensa trabajar, así como establecer las condiciones iniciales a partir de las cuales los programas comenzarán la reconstrucción. En este caso, se eligió el modelo de sustitución LG porque incorpora las tasas de sustitución de aminoácidos de una mayor cantidad de familias de proteínas de las bases de datos, así como una prueba estadística paramétrica, el aLRT, que resultó ser una mejor alternativa al bootstrap debido que a reduce el tiempo de cómputo. Por otro lado, la metodología de la reconstrucción de ancestros empleada para este proyecto fue establecida con la finalidad de escoger las secuencias de una manera objetiva utilizando la distancia evolutiva, es decir, el número de sustituciones por sitio entre cada secuencia.

Se puede concluir que, gracias al análisis del árbol obtenido, los principales mecanismos implicados en la generación de variabilidad de esta subfamilia de aldolasas fueron la duplicación génica que se dio en etapas tempranas de la diversificación de los dominios filogenéticos de la vida, y el transporte horizontal de genes que permitió la adaptación de algunos organismos a nuevos nichos ecológicos. Debido a que estas enzimas están implicadas en vías importantes del metabolismo energético y de biosíntesis de aminoácidos, es decir, rutas que debieron haber aparecido hace millones de años, y que se identificó una gran divergencia incluso entre dominios filogenéticos de cada enzima (el tamaño de sitios en el alineamiento era muy grande y había regiones problemáticas con muchos gaps), es difícil establecer fechas del origen de cada una; sin embargo, se pueden realizar hipótesis que impliquen el posible orden de aparición, e incluso se puede decir que debido a la promiscuidad catalítica de las enzimas ancestrales, la diversificación en la vías metabólicas se dio como consecuencia de un proceso de reclutamiento.

A pesar de que no se identificaron propiamente los residuos de aminoácidos que juegan indirectamente un papel importante en la catálisis, se establecieron las condiciones que podrían permitir en un futuro, utilizando más herramientas bioinformáticas, el reconocimiento de éstos, así como su función.

Las perspectivas de este proyecto son:

- Realizar mutaciones puntuales de los residuos identificados en este trabajo con la finalidad de analizar su función.
- Realizar experimentos de *docking* molecular con la finalidad de observar cuál es la afinidad de las enzimas ancestrales por los ligandos utilizados por las enzimas actuales, y escoger las que potencialmente podrían tener actividad catalítica para después sintetizarlas y probarlas en el laboratorio.
- Escoger una mayor cantidad de nodos ancestrales que permitan ver con más detalle las sustituciones de aminoácidos que se han ido acumulando a lo largo del tiempo, y posteriormente, elegir los posibles residuos que han ido coevolucionando con los residuos ubicados en el sitio activo implicados directamente en la catálisis o en la estabilidad estructural de la enzima.

## Referencias

- Ahmed, H., Ettema, T. J. G., Tjaden, B., Geerling, A. C. M., van der Oost, J., & Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *The Biochemical journal*, 390(Pt 2), 529–40. doi:10.1042/BJ20041711
- Aisaka, K., Igarashi, A., Yamaguchi, K., & Uwajima, T. (1991). Purification, crystallization and characterization of N-acetylneuraminate lyase from *Escherichia coli*. *Biochem J*, 276, 541–546. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1151125&tool=pmcentrez&rendertype=abstract>
- Aisaka, K., & Uwajima, T. (1986). Cloning and constitutive expression of the N-acetylneuraminate lyase gene of *Escherichia coli*. *Appl Environ Microbiol*, 51(3), 562–565. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/3008647>
- Allard, J., Grochulski, P., & Sygusch, J. (2001). Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 98(7), 3679–84. doi:10.1073/pnas.071380898
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403–410. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/2231712>
- Anderson, M., Scholtz, J. M., & Schuster, S. M. (1985). Rat liver 4-hydroxy-2-ketoglutarate aldolase: purification and kinetic characterization. *Archives of biochemistry and biophysics*, 236(1), 82–97. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3966804>
- Angata, T., & Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chemical reviews*, 102(2), 439–69. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11841250>
- Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol*, 55(4), 539–552. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/16785212>
- Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., & Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic acids research*, 40(Web Server issue), W580–4. doi:10.1093/nar/gks498
- Barbosa, J. A., Smith, B. J., DeGori, R., Ooi, H. C., Marcuccio, S. M., Campi, E. M., ... Lawrence, M. C. (2000). Active site modulation in the N-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed *Haemophilus influenzae* enzyme. *J Mol Biol*, 303(3), 405–421. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/11031117>
- Barnsley, E. A. (1983). Phthalate pathway of phenanthrene metabolism: formation of 2'-carboxybenzalpyruvate. *J Bacteriol*, 154(1), 113–117. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/6833175>
- Bell, B. J., Watanabe, L., Rios-Steiner, J. L., Tulinsky, A., Lebioda, L., & Arni, R. K. (2003). Structure of 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase from *Pseudomonas putida*.

- Acta crystallographica. Section D, Biological crystallography*, 59(Pt 8), 1454–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12876349>
- Blickling, S., Renner, C., Laber, B., Pohlentz, H. D., Holak, T. A., & Huber, R. (1997). Reaction mechanism of Escherichia coli dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. *Biochemistry*, 36(1), 24–33. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/8993314>
- Boc, A., Diallo, A. B., & Makarenkov, V. (2012). T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic acids research*, 40(Web Server issue), W573–9. doi:10.1093/nar/gks485
- Brands, S. J. (2013). The Taxonomicon. *Universal Taxonomic Services*. Retrieved from <http://taxonomicon.taxonomy.nl/>
- Bryant, H. N. (2001). Character polarity and the rooting of cladograms. In G. P. Wagner (Ed.), *The Character Concept in Evolutionary Biology* (pp. 319–338). New York: Academic Press.
- Buchanan, C. L., Connaris, H., Danson, M. J., Reeve, C. D., & Hough, D. W. (1999). An extremely thermostable aldolase from Sulfolobus solfataricus with specificity for non-phosphorylated substrates. *Biochem J*, 343 Pt 3, 563–570. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/10527934>
- Burgess, B. R., Dobson, R. C. J., Bailey, M. F., Atkinson, S. C., Griffin, M. D. W., Jameson, G. B., ... Perugini, M. A. (2008). Structure and evolution of a novel dimeric enzyme from a clinically important bacterial pathogen. *The Journal of biological chemistry*, 283(41), 27598–603. doi:10.1074/jbc.M804231200
- Cahyanto, M. N., Kawasaki, H., Nagashio, M., Fujiyama, K., & Seki, T. (2006). Regulation of aspartokinase, aspartate semialdehyde dehydrogenase, dihydrodipicolinate synthase and dihydrodipicolinate reductase in Lactobacillus plantarum. *Microbiology (Reading, England)*, 152(Pt 1), 105–12. doi:10.1099/mic.0.28092-0
- Carr, S. M. (2009). Midpoint rooting, a network of relationships. Retrieved September 03, 2013, from [http://www.mun.ca/biology/scarr/Panda\\_midpoint\\_rooting.html](http://www.mun.ca/biology/scarr/Panda_midpoint_rooting.html)
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*, 17(4), 540–552. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/10742046>
- Celerin, M., Ray, J. M., Schisler, N. J., Day, A. W., Stetler-Stevenson, W. G., & Laudendach, D. E. (1996). Fungal fimbriae are composed of collagen. *The EMBO journal*, 15(17), 4445–53. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452173&tool=pmcentrez&rendertype=abstract>
- Chang, B. S. W., Ugalde, J. A., & Matz, M. V. (2005). Applications of ancestral protein reconstruction in understanding protein function: GFP-like proteins. *Methods Enzymol*, 395, 652–670. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/15865989>
- Chang, & Donoghue. (2000). Recreating ancestral proteins. *Trends Ecol Evol*, 15(3), 109–114. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/10675926>

- Chen, K., Durand, D., & Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of computational biology: a journal of computational molecular cell biology*, 7(3-4), 429-47. doi:10.1089/106652700750050871
- Chen, Z., & Zhao, H. (2006). Protein Design. In S. Lee (Ed.), *Encyclopedia of Chemical Processing* (pp. 2467-2477). New York: CRC Press. doi:10.1081/E-ECHP-120018030
- Choi, K. H., Lai, V., Foster, C. E., Morris, A. J., Tolan, D. R., & Allen, K. N. (2006). New superfamily members identified for Schiff-base enzymes based on verification of catalytically essential residues. *Biochemistry*, 45(28), 8546-55. doi:10.1021/bi060239d
- Conway, T. (1992). The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol Rev*, 9(1), 1-27. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/1389313>
- Cremer, J., Treptow, C., Eggeling, L., & Sahm, H. (1988). Regulation of enzymes of lysine biosynthesis in *Corynebacterium glutamicum*. *Journal of general microbiology*, 134(12), 3221-9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3151991>
- De Koning, A. P., Brinkman, F. S., Jones, S. J., & Keeling, P. J. (2000). Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Molecular biology and evolution*, 17(11), 1769-73. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11070064>
- Dekker, E. E., & Kitson, R. P. (1992). 2-Keto-4-hydroxyglutarate aldolase: purification and characterization of the homogeneous enzyme from bovine kidney. *The Journal of biological chemistry*, 267(15), 10507-14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1587831>
- Delport, W., Poon, A. F. Y., Frost, S. D. W., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics (Oxford, England)*, 26(19), 2455-7. doi:10.1093/bioinformatics/btq429
- Dobson, R. C. J., Griffin, M. D. W., Devenish, S. R. A., Pearce, F. G., Hutton, C. A., Gerrard, J. A., ... Perugini, M. A. (2008). Conserved main-chain peptide distortions: a proposed role for Ile203 in catalysis by dihydrodipicolinate synthase. *Protein science: a publication of the Protein Society*, 17(12), 2080-90. doi:10.1110/ps.037440.108
- Dobson, R. C. J., Griffin, M. D. W., Jameson, G. B., & Gerrard, J. A. (2005). The crystal structures of native and (S)-lysine-bound dihydrodipicolinate synthase from *Escherichia coli* with improved resolution show new features of biological significance. *Acta crystallographica. Section D, Biological crystallography*, 61(Pt 8), 1116-24. doi:10.1107/S0907444905016318
- Dobson, R. C. J., Valegård, K., & Gerrard, J. A. (2004). The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad. *Journal of molecular biology*, 338(2), 329-39. doi:10.1016/j.jmb.2004.02.060
- Dwyer, M. A., Looger, L. L., & Hellinga, H. W. (2004). Computational design of a biologically active enzyme. *Science*, 304(5679), 1967-1971. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/15218149>
- Eaton, R. W. (1994). Organization and evolution of naphthalene catabolic pathways: sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and trans-o-hydroxybenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid. *Journal of bacteriology*, 176(24), 7757-62. Retrieved from

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=197239&tool=pmcentrez&rendertype=abstract>

- Egan, S. E., Fliege, R., Tong, S., Shibata, A., Wolf, R. E. J., & Conway, T. (1992). Molecular characterization of the Entner-Doudoroff pathway in *Escherichia coli*: sequence analysis and localization of promoters for the *edd-eda* operon. *J Bacteriol*, *174*(14), 4638–4646. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/1624451>
- Elshafei, A. M., & Abdel-Fatah, O. M. (1991). Nonphosphorolytic pathway for d-galactonate catabolism in *Aspergillus terreus*. *Enzyme and Microbial Technology*, *13*(11), 930–934. Retrieved from <http://www.sciencedirect.com/science/article/pii/014102299190111M>
- Euzéby, J. P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *International journal of systematic bacteriology*, *47*(2), 590–2. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9103655>
- Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., & Baart, G. J. E. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *The Plant journal: for cell and molecular biology*, *70*(6), 1004–14. doi:10.1111/j.1365-313X.2012.04941.x
- Ferrara, S., Mapelli, E., Sello, G., & Di Gennaro, P. (2011). Characterization of the aldol condensation activity of the trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) cloned from *Pseudomonas fluorescens* N3. *Biochimica et biophysica acta*, *1814*(5), 622–9. doi:10.1016/j.bbapap.2011.03.013
- Fessner, W.-D., & Walter, C. (1997). Enzymatic C-C bond formation in asymmetric synthesis. In F. P. Schmidtchen (Ed.), *Bioorganic Chemistry* (Vol. 184, pp. 97–194). Springer Berlin Heidelberg. Retrieved from [http://dx.doi.org/10.1007/3-540-61388-9\\_63](http://dx.doi.org/10.1007/3-540-61388-9_63)
- Futuyma, D. J. (2005). Evolutionary Biology. In *Evolution* (pp. 2–15). Sunderland: Sinauer Associates.
- García García, M. I., Sola Carvajal, A., García Carmona, F., & Sánchez Ferrer, A. (2012). Characterization of a Novel N-Acetylneuraminase Lyase from *Staphylococcus carnosus* TM300 and Its Application to N-Acetylneuraminic Acid Production. *Journal of agricultural and food chemistry*. doi:10.1021/jf3014102
- García Ortega, A. M., & Ponce Rivas, E. (2003). Metabolismo del carbono en microorganismos de interés biomédico y biotecnológico: Vía de Entner-Doudoroff. *Biotecnología Aplicada*, *20*, 85–94.
- Gerlt, J. A., & Babbitt, P. C. (2009). Enzyme (re)design: lessons from natural evolution and computation. *Curr Opin Chem Biol*, *13*(1), 10–18. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/19237310>
- Glasner, M. E., Gerlt, J. A., & Babbitt, P. C. (2006). Evolution of enzyme superfamilies. *Current opinion in chemical biology*, *10*(5), 492–7. doi:10.1016/j.cbpa.2006.08.012
- Grady, S. R., Wang, J. K., & Dekker, E. E. (1981). Steady-state kinetics and inhibition studies of the aldol condensation reaction catalyzed by bovine liver and *Escherichia coli* 2-keto-4-hydroxyglutarate aldolase. *Biochemistry*, *20*(9), 2497–502. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7016177>

- Griffin, M. D. W., Billakanti, J. M., Wason, A., Keller, S., Mertens, H. D. T., Atkinson, S. C., ... Pearce, F. G. (2012). Characterisation of the first enzymes committed to lysine biosynthesis in *Arabidopsis thaliana*. *PLoS one*, 7(7), e40318. doi:10.1371/journal.pone.0040318
- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, 52(5), 696–704. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/14530136>
- Gunji, Y., Tsujimoto, N., Shimaoka, M., Ogawa-Miyata, Y., Sugimoto, S.-I., & Yasueda, H. (2004). Characterization of the L-lysine biosynthetic pathway in the obligate methylophilic *Methylophilus methylotrophus*. *Bioscience, biotechnology, and biochemistry*, 68(7), 1449–60. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15277749>
- Halling, S. M., & Stahly, D. P. (1976). Dihydrodipicolinic acid synthase of *Bacillus licheniformis*. Quaternary structure, kinetics, and stability in the presence of sodium chloride and substrates. *Biochimica et biophysica acta*, 452(2), 580–96. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1009127>
- Han, M. V., & Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC bioinformatics*, 10, 356. doi:10.1186/1471-2105-10-356
- Harms, M. J., & Thornton, J. W. (2010). Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol*, 20(3), 360–366. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/20413295>
- Hoganson, D. A., & Stahly, D. P. (1975). Regulation of dihydrodipicolinate synthase during growth and sporulation of *Bacillus cereus*. *Journal of bacteriology*, 124(3), 1344–50. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=236046&tool=pmcentrez&rendertype=abstract>
- Hordijk, W., & Gascuel, O. (2005). Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics*, 21(24), 4338–4347. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/16234323>
- Hult, K., & Berglund, P. (2007). Enzyme promiscuity: mechanism and applications. *Trends in biotechnology*, 25(5), 231–8. doi:10.1016/j.tibtech.2007.03.002
- Iwabuchi, T., & Harayama, S. (1998). Biochemical and Genetic Characterization of trans-2'-Carboxybenzalpyruvate Hydratase-Aldolase from a Phenanthrene-Degrading *Nocardioide* Strain. *J Bacteriol*, 180(4), 945–949.
- Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G., & Colman, P. M. (1994). The three-dimensional structure of N-acetylneuraminidase from *Escherichia coli*. *Structure*, 2(5), 361–369. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/8081752>
- Jermann, T. M., Opitz, J. G., Stackhouse, J., & Benner, S. A. (1995). Reconstructing the evolutionary history of the arylglycosyl hydrolase superfamily. *Nature*, 374(6517), 57–59. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/7532788>
- Joerger, A. C., Mayer, S., & Fersht, A. R. (2003). Mimicking natural evolution in vitro: an N-acetylneuraminidase mutant with an increased dihydrodipicolinate synthase activity. *Proc Natl Acad Sci U S A*, 100(10), 5694–5699. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/12711733>

- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature protocols*, 7(8), 1511–22. doi:10.1038/nprot.2012.085
- Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27–30. doi:10.1093/nar/28.1.27
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(Database issue), D109–14. doi:10.1093/nar/gkr988
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, 30(14), 3059–3066. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/12136088>
- Kiyohara, H., & Nagao, K. (1978). The Catabolism of Phenanthrene and Naphthalene by Bacteria. *Journal of General Microbiology*, 105(1), 69–75. doi:10.1099/00221287-105-1-69
- Kiyohara, Hohzoh, Nagao, K., & Nomi, R. (1976). Degradation of Phenanthrene through o-Phthalate by an Aeromonas sp. *Agr Biol Chem*, 40(6), 1075–1082. doi:<http://dx.doi.org/10.1271/bbb1961.40.1075>
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Wägele, J. W., & Misof, B. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front Zool*, 7, 10. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/20356385>
- Kuhm, a E., Knackmuss, H. J., & Stolz, a. (1993). Purification and properties of 2'-hydroxybenzalpyruvate aldolase from a bacterium that degrades naphthalenesulfonates. *The Journal of biological chemistry*, 268(13), 9484–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8486638>
- Kumpaisal, R., Hashimoto, T., & Yamada, Y. (1987). Purification and characterization of dihydrodipicolinate synthase from wheat suspension cultures. *Plant physiology*, 85(1), 145–51. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1054219&tool=pmcentrez&rendertype=abstract>
- Lane, R. S., Shapley, A., & Dekker, E. E. (1971). 2-keto-4-hydroxybutyrate aldolase. Identification as 2-keto-4-hydroxyglutarate aldolase, catalytic properties, and role in the mammalian metabolism of L-homoserine. *Biochemistry*, 10(8), 1353–64. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5580656>
- Lawrence, M. C., Barbosa, J. A., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C., & Marcuccio, S. M. (1997). Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminidase. *J Mol Biol*, 266(2), 381–399. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/9047371>
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol*, 25(7), 1307–1320. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/18367465>

- Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)*, *23*(1), 127–8. doi:10.1093/bioinformatics/btl529
- Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research*, *39*(Web Server issue), W475–8. doi:10.1093/nar/gkr201
- Lilley, G. G., Barbosa, J. A., & Pearce, L. A. (1998). Expression in Escherichia coli of the putative N-acetylneuraminase lyase gene (nanA) from Haemophilus influenzae: overproduction, purification, and crystallization. *Protein expression and purification*, *12*(3), 295–304. doi:10.1006/prev.1997.0841
- Maddison, W. P., & Maddison, D. R. (2011). Mesquite: a modular system for evolutionary analysis. Version 2.75. Retrieved from <http://mesquiteproject.org>
- Madigan, M. T., Martinko, J. M., Stahl, D. A., & Clark, D. P. (2012). Catabolism of Organic Compounds. In *Brock Biology of Microorganisms* (13th ed., pp. 372–410). San Francisco: Benjamin Cummings.
- Mazelis, M., Whatley, F. R., & Whatley, J. (1977). The enzymology of lysine biosynthesis in higher plants. The occurrence, characterization and some regulatory properties of dihydrodipicolinate synthase. *FEBS letters*, *84*(2), 236–40. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/598503>
- Meysick, K. C., Dimock, K., & Garber, G. E. (1996). Molecular characterization and expression of a N-acetylneuraminase lyase gene from Trichomonas vaginalis. *Molecular and biochemical parasitology*, *76*(1-2), 289–92. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8920014>
- Mirwaldt, C., Korndörfer, I., & Huber, R. (1995). The crystal structure of dihydrodipicolinate synthase from Escherichia coli at 2.5 Å resolution. *J Mol Biol*, *246*(1), 227–239. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/7853400>
- Morris, P. F., Schlosser, L. R., Onasch, K. D., Wittenschlaeger, T., Austin, R., & Provart, N. (2009). Multiple horizontal gene transfer events and domain fusions have created novel regulatory and metabolic networks in the oomycete genome. *PloS one*, *4*(7), e6133. doi:10.1371/journal.pone.0006133
- NaturaLista. (2013). NaturaLista: Especies. *Comisión Nacional de Biodiversidad (Conabio)*. Retrieved from <http://conabio.inaturalist.org/taxa>
- Nees, S., Schauer, R., & Mayer, F. (1976). Purification and characterization of N-acetylneuraminase lyase from Clostridium perfringens. *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, *357*(6), 839–53. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/182637>
- O'Brien, P. J., & Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & biology*, *6*(4), R91–R105. doi:10.1016/S1074-5521(99)80033-7
- Patil, R. V., & Dekker, E. E. (1992). Cloning, nucleotide sequence, overexpression, and inactivation of the Escherichia coli 2-keto-4-hydroxyglutarate aldolase gene. *J Bacteriol*, *174*(1), 102–107. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/1339418>

- Patrick, G. L. (2004). Aldehydes and ketones. In *BIOS Instant Notes in Organic Chemistry* (2nd ed., pp. 167–204). New York: Taylor & Francis.
- Pauluhn, A., Ahmed, H., Lorentzen, E., Buchinger, S., Schomburg, D., Siebers, B., & Pohl, E. (2008). Crystal structure and stereochemical studies of KD(P)G aldolase from *Thermoproteus tenax*. *Proteins*, *72*(1), 35–43. doi:10.1002/prot.21890
- Peekhaus, N., & Conway, T. (1998). What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J Bacteriol*, *180*(14), 3495–3502. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/9657988>
- Pérez-Zamorano, B. (2007). *Directed evolution of two enzymes of the N-acetylneuraminate lyase family*. University of Leeds.
- Pond, S. L. K., & Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics (Oxford, England)*, *21*(10), 2531–3. doi:10.1093/bioinformatics/bti320
- Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics (Oxford, England)*, *21*(5), 676–9. doi:10.1093/bioinformatics/bti079
- Pupko, T., Pe, I., Shamir, R., & Graur, D. (2000). A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. *Molecular Biology and Evolution*, *17*(6), 890–896. doi:10.1093/oxfordjournals.molbev.a026369
- Reher, M., Fuhrer, T., Bott, M., & Schönheit, P. (2010). The nonphosphorylative Entner-Doudoroff pathway in the thermoacidophilic euryarchaeon *Picrophilus torridus* involves a novel 2-keto-3-deoxygluconate-specific aldolase. *Journal of bacteriology*, *192*(4), 964–74. doi:10.1128/JB.01281-09
- Ridley, M. (2004a). The Rise of Evolutionary Biology. In *Evolution* (3rd ed., pp. 3–20). Malden: Blackwell Publishing.
- Ridley, M. (2004b). Natural Selection and Random Drift in Molecular Evolution. In *Evolution* (3rd ed., pp. 155–193). Malden: Blackwell Publishing.
- Romano, A. H., & Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Research in Microbiology*, *147*(6), 448–455. Retrieved from <http://www.sciencedirect.com/science/article/pii/0923250896839982>
- Ryu, H. C., Park, S. Y., Lee, S. H., & Kim, J. S. (2010). Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of 2-keto-3-deoxy-6-phosphogluconate aldolase from *Zymomonas mobilis* ZM4. *Acta crystallographica. Section F, Structural biology and crystallization communications*, *66*(Pt 4), 471–3. doi:10.1107/S1744309110007323
- Sánchez-Carrón, G., García-García, M. I., López-Rodríguez, A. B., Jiménez-García, S., Sola-Carvajal, A., García-Carmona, F., & Sánchez-Ferrer, A. (2011). Molecular characterization of a novel N-acetylneuraminate lyase from *Lactobacillus plantarum* WCFS1. *Applied and environmental microbiology*, *77*(7), 2471–8. doi:10.1128/AEM.02927-10
- Schauer, R., Sommer, U., Krüger, D., van Unen, H., & Traving, C. (1999). The terminal enzymes of sialic acid metabolism: acylneuraminate pyruvate-lyases. *Bioscience reports*, *19*(5), 373–83. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10763805>

- Schauer, R., & Wember, M. (1996). Isolation and characterization of sialate lyase from pig kidney. *Biological chemistry Hoppe-Seyler*, 377(5), 293–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8828820>
- Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.5.
- Seo, J.-S., Keum, Y.-S., & Li, Q. X. (2012). Mycobacterium aromativorans JS19b1(T) Degrades Phenanthrene through C-1,2, C-3,4 and C-9,10 Dioxygenation Pathways. *International biodeterioration & biodegradation*, 70, 96–103. doi:10.1016/j.ibiod.2012.02.005
- Siebers, B., & Schönheit, P. (2005). Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Current Opinion in Microbiology*, 8(6), 695–705. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1369527405001700>
- Sirbasku, D. A., & Binkley, S. B. (1970). Purification and properties of N-acetylneuraminase lyase from beef kidney cortex. *Biochimica et biophysica acta*, 206(3), 479–82. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5459552>
- Smith, E. L. (1970). Evolution of Enzymes. In P. D. Boyer (Ed.), *The Enzymes* (3rd ed., Vol. Volume 1, pp. 267–339). New York: Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1874604708601699>
- Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, 28(6), 1102, 1104. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10868275>
- Szumilo, T. (1981). Pathway for D-galactonate catabolism in nonpathogenic mycobacteria. *J Bacteriol*, 148(1), 368–370. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/7287628>
- Taha, T. S., & Deits, T. L. (1994). Purification and characterization of 2-keto-3-deoxy-6-phosphogluconate aldolase from Azotobacter vinelandii: evidence that the enzyme is bifunctional towards 2-keto-4-hydroxy glutarate cleavage. *Biochemical and biophysical research communications*, 200(1), 459–66. doi:10.1006/bbrc.1994.1471
- Takayama, S., McGarvey, G. J., & Wong, C. H. (1997). Microbial aldolases and transketolases: new biocatalytic approaches to simple and complex sugars. *Annual review of microbiology*, 51, 285–310. doi:10.1146/annurev.micro.51.1.285
- Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet*, 5(5), 366–375. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/15143319>
- Vandamme, A.-M. (2009). Basic concepts of molecular evolution. In P. Lemey, M. Salemi, & A.-M. Vandamme (Eds.), *The Phylogenetic Handbook, a practical approach to phylogenetic analysis and hypothesis testing* (2nd ed., pp. 3–29). New York: Cambridge University Press.
- Velasco, A. M., Leguina, J. I., & Lazcano, A. (2002). Molecular evolution of the lysine biosynthetic pathways. *Journal of molecular evolution*, 55(4), 445–59. doi:10.1007/s00239-002-2340-2
- Vimr, E. R., & Troy, F. A. (1985). Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol*, 164(2), 845–853. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/3902799>

- Vogel, H. J. (1964). Distribution of Lysine Pathways Among Fungi : Evolutionary Implications. *American Society of Naturalists*, 98(903), 435–446.
- Walters, M. J., Srikannathasan, V., McEwan, A. R., Naismith, J. H., Fierke, C. A., & Toone, E. J. (2008). Characterization and crystal structure of Escherichia coli KDPGal aldolase. *Bioorg Med Chem*, 16(2), 710–720. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/17981470>
- Watanabe, S., Morimoto, D., Fukumori, F., Shinomiya, H., Nishiwaki, H., Kawano-Kawada, M., ... Watanabe, Y. (2012). Identification and characterization of D-hydroxyproline dehydrogenase and Delta1-pyrroline-4-hydroxy-2-carboxylate deaminase involved in novel L-hydroxyproline metabolism of bacteria: metabolic convergent evolution. *The Journal of biological chemistry*, 287(39), 32674–88. doi:10.1074/jbc.M112.374272
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/19151095>
- Webster, F. H., & Lechowich, R. V. (1970). Partial purification and characterization of dihydrodipicolinic acid synthetase from sporulating Bacillus megaterium. *Journal of bacteriology*, 101(1), 118–26. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=250458&tool=pmcentrez&rendertype=abstract>
- Williams, P. D., Pollock, D. D., Blackburne, B. P., & Goldstein, R. A. (2006). Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol*, 2(6), e69. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/16789817>
- Wong, T. Y., & Yao, X. T. (1994). The DeLey-Doudoroff Pathway of Galactose Metabolism in Azotobacter vinelandii. *Appl Environ Microbiol*, 60(6), 2065–2068. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/16349292>
- Yang, X., Kathuria, S. V., Vadrevu, R., & Matthews, C. R. (2009). Betaalpha-hairpin clamps brace betaalphabetamodules and can make substantive contributions to the stability of TIM barrel proteins. (A. Hofmann, Ed.) *PloS one*, 4(9), e7179. doi:10.1371/journal.pone.0007179
- Yang, Z. (2006a). Phylogeny reconstruction: overview. In *Computational Molecular Evolution* (pp. 73–99). New York: Oxford University Press.
- Yang, Z. (2006b). Maximum likelihood methods. In *Computational Molecular Evolution* (pp. 100–144). New York: Oxford University Press.
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature reviews Genetics*, 13(5), 303–14. doi:10.1038/nrg3186

## Apéndice 1: Clases taxonómicas

<b>Bacteria</b>	<b>Especies</b>
Actinobacteria	20
Alphaproteobacteria	104
Aquificae	7
Bacilli	44
Bacteroidia	5
Betaproteobacteria	46
Caldilineae	1
Chrysiogenetes	1
Clostridia	20
Cyanobacteria	2
Cytophagia	2
Deferribacteres	4
Deltaproteobacteria	19
Epsilonproteobacteria	3
Erysipelotrichia	3
Flavobacteriia	1
Fusobacteriia	8
Gammaproteobacteria	186
Ignavibacteria	1
Lentisphaeria	2
Mollicutes	5
Negativicutes	5
Nitrospira	5
Opitutae	1
Planctomycetacia	3
Spartobacteria	1
Sphingobacteriia	4
Spirochaetes	6
Synergistia	2
Thermodesulfobacteria	2
Thermomicrobia	1
Thermotogae	3
Verruimicrobiae	1
<b>Total general</b>	<b>518</b>

<b>Archaea</b>	<b>Especies</b>
Archaeoglobi	4
Halobacteria	42
Methanobacteria	7
Methanococci	11
Methanomicrobia	23
Methanopyri	1
Thermoplasmata	4
Thermoprotei	11
<b>Total general</b>	<b>103</b>

<b>Eukarya</b>	<b>Especies</b>
Actinopterygii	9
Agaricomycetes	8
Amphibia	3
Anthozoa	1
Arachnida	2
Ascidiacea	1
Aves	5
Bacillariophyceae	2
Bangiophyceae	1
Branchiopoda	1
Bryopsida	1
Chaetothyriomycetes	1
Chlorophyceae	2
Conoidasida	2
Coscinodiscophyceae	1
Cryptophyta	2
Cyanidiophyceae	1
Dacrymycetes	1
Echinoidea	1
Enteropneusta	1
Eurotiomycetes	5
Hydrozoa	1
Imbricatea	1
Insecta	4
Leotiomycetes	4
Leptocardii	1
Liliopsida	7
Lycopodiopsida	1
Magnoliopsida	11
Mammalia	45
Mucormycotina	1
Oomycetes	3
Parabasalia	1
Pelagophyceae	2
Perkinsea	1
Phaeophyceae	1
Pinopsida	1
Polychaeta	2
Prasinophyceae	5
Prymnesiophyceae	1
Reptilia	2
Sordariomycetes	17
Spermatopsida	1
Trebouxiophyceae	2
Tricoplacia	1
Ustilaginomycetes	2
Variosea	1
Wallemiomycetes	1
<b>Total general</b>	<b>171</b>

## Apéndice 2: Estructuras tridimensionales

RaptorX presenta como resultado los mejores PDBs que sirvieron como plantados para realizar el alineamiento con la secuencia problema, así como los valores P-value y uGDT para calificar la calidad de la estructura resultante. El valor uGDT representa el puntaje no normalizado de GDT (Global Distance Test) que se define como  $1*N(1)+0.75*N(2)+0.5*N(4)+0.25*N(8)$ , donde  $N(x)$  es el número de residuos con el RMSD local más pequeño que  $x$ . El valor GDT es calculado dividiendo uGDT entre la longitud del dominio, multiplicado por 100. Graficado los valores de  $-\log(P\text{-value})$  vs. los valores de uGDT, se considera razonablemente bueno un modelo estructural si el puntaje GDT es mayor a 50 (Källberg et al., 2012).

### SECUENCIA RESULTADOS DE RAPTORX

MIEMBROS DE LA SUBFAMILIA SIN ESTRUCTURA TRIDIMENSIONAL													
<b>CPOEA</b> <b>1-332</b>	P-value: 7.99e-09 uGDT(GDT): 221(67) uSeqId(SeqId): 45(14) Score: 329 Templates: 3lbcA 3pudA 3fluA 3qzeA 3lerA												
<b>HBPFA</b> <b>1-334</b>	P-value: 1.05e-08 uGDT(GDT): 219(66) uSeqId(SeqId): 46(14) Score: 309 Templates: 3lbcA 3pudA 3qzeA 3fluA 2r8wA												
NODOS ANCESTRALES DE THDPS Y NAL													
<b>N6</b> <b>64-416</b>	P-value: 5.00e-09 uGDT(GDT): 242(68) uSeqId(SeqId): 137(39) Score: 319 Templates: 2ehhA 2yxxA 1o5kA 2vc6A 3lerA												
<b>N7</b> <b>64-416</b>	P-value: 1.05e-08 uGDT(GDT): 236(67) uSeqId(SeqId): 122(35) Score: 315 Templates: 2yxxA 2ehhA 1o5kA 3si9A 2v8zA												
<b>N8</b> <b>32-357</b>	P-value: 1.87e-08 uGDT(GDT): 233(71) uSeqId(SeqId): 111(34) Score: 324 Templates: 2ehhA 3lclA 1o5kA 2yxxA 3si9A												
<b>N9</b> <b>30-351</b>	P-value: 4.67e-08 uGDT(GDT): 244(76) uSeqId(SeqId): 127(39) Score: 322 Templates: 3lclA 1o5kA												
<b>N104</b> <b>1-300</b>	P-value: 1.62e-07 uGDT(GDT): 268(89) uSeqId(SeqId): 173(58) Score: 300 Templates: 2r91A 1w37A 2nuwA 3lclA 2ehhA												
<b>N109</b>	<table border="0"> <tr> <td><b>1-63</b></td> <td><b>64-400</b></td> </tr> <tr> <td>P-value: 2.62e-02</td> <td>P-value: 5.23e-09</td> </tr> <tr> <td>uGDT(GDT): 7(10)</td> <td>uGDT(GDT): 250(74)</td> </tr> <tr> <td>uSeqId(SeqId): 7(11)</td> <td>uSeqId(SeqId): 134(40)</td> </tr> <tr> <td>Score: 21</td> <td>Score: 332</td> </tr> <tr> <td>Template: 3k2l:A</td> <td>Templates: 3s5nA 3fluA 3pudA 3si9A 2ehhA</td> </tr> </table>	<b>1-63</b>	<b>64-400</b>	P-value: 2.62e-02	P-value: 5.23e-09	uGDT(GDT): 7(10)	uGDT(GDT): 250(74)	uSeqId(SeqId): 7(11)	uSeqId(SeqId): 134(40)	Score: 21	Score: 332	Template: 3k2l:A	Templates: 3s5nA 3fluA 3pudA 3si9A 2ehhA
<b>1-63</b>	<b>64-400</b>												
P-value: 2.62e-02	P-value: 5.23e-09												
uGDT(GDT): 7(10)	uGDT(GDT): 250(74)												
uSeqId(SeqId): 7(11)	uSeqId(SeqId): 134(40)												
Score: 21	Score: 332												
Template: 3k2l:A	Templates: 3s5nA 3fluA 3pudA 3si9A 2ehhA												

\*Valores de GDT menores a 50 resaltados en rojo.

**SECUENCIA      RESULTADOS DE RAPTORX**  
**NODOS ANCESTRALES DE KDPGA/HOGA Y KDPGALA**

<b>N8</b> <b>40-273</b>	P-value: 1.76e-07 uGDT(GDT): 174(74) uSeqId(SeqId): 70(30) Score: 231 Templates: 1vlwA 3vcrA 2v81A
<b>N19</b> <b>1-256</b>	P-value: 1.71e-07 uGDT(GDT): 179(70) uSeqId(SeqId): 80(31) Score: 240 Templates: 4e38A 1vhcA 1mxsA 1wbhA 3vcrA
<b>N20</b> <b>1-256</b>	P-value: 9.59e-08 uGDT(GDT): 177(69) uSeqId(SeqId): 76(30) Score: 242 Templates: 1vhcA 4e38A 1mxsA 1wbhA 3vcrA
<b>N21</b> <b>1-244</b>	P-value: 2.16e-07 uGDT(GDT): 184(75) uSeqId(SeqId): 112(46) Score: 223 Templates: 2v81A 1vlwA 4e38A 3vcrA
<b>N58</b> <b>1-235</b>	P-value: 7.30e-08 uGDT(GDT): 201(85) uSeqId(SeqId): 129(55) Score: 235 Templates: 1euaA 1mxsA 3vcrA 4e38A 1vhcA
<b>N134</b> <b>1-220</b>	P-value: 4.41e-07 uGDT(GDT): 183(83) uSeqId(SeqId): 67(31) Score: 220 Templates: 1fwrA 4e38A 1vhcA 1vlwA 1mxsA