



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**

**FACULTAD DE CIENCIAS**  
**BIOLOGÍA EXPERIMENTAL**

**EVOLUCIÓN DE RUTAS METABÓLICAS Y TEORÍA DE REDES**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**CLAUDIA SIERRA MENCHACA**

**TUTOR PRINCIPAL DE TESIS: DR. ARTURO CARLOS II BECERRA BRACHO**  
**FACULTAD DE CIENCIAS, UNAM**

**COMITÉ TUTOR: DR. PEDRO EDUARDO MIRAMONTES VIDAL**  
**FACULTAD DE CIENCIAS, UNAM**

**DR. PABLO PADILLA LONGORIA**

**INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS, UNAM**

**CD. MX. DICIEMBRE, 2018**



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.





**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**

FACULTAD DE CIENCIAS  
BIOLOGÍA EXPERIMENTAL

**EVOLUCIÓN DE RUTAS METABÓLICAS Y TEORÍA DE REDES**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**CLAUDIA SIERRA MENCHACA**

**TUTOR PRINCIPAL DE TESIS: DR. ARTURO CARLOS II BECERRA BRACHO**

FACULTAD DE CIENCIAS, UNAM

**COMITÉ TUTOR: DR. PEDRO EDUARDO MIRAMONTES VIDAL**

FACULTAD DE CIENCIAS, UNAM

**DR. PABLO PADILLA LONGORIA**

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS, UNAM

**MÉXICO, CD. MX. DICIEMBRE, 2018**



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO  
POSGRADO EN CIENCIAS BIOLÓGICAS  
FACULTAD DE CIENCIAS  
DIVISION ACADÉMICA DE INVESTIGACIÓN Y POSGRADO

OFICIO FCIE/DAIP/1062/2018

ASUNTO: Oficio de Jurado

Lic. Ivonne Ramírez Wence  
Directora General de Administración Escolar, UNAM  
Presente

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 27 de agosto de 2018, se aprobó el siguiente jurado para el examen de grado de MAESTRA EN CIENCIAS BIOLÓGICAS (BIOLOGÍA EXPERIMENTAL) de la alumna SIERRA MENCHACA CLAUDIA con número de cuenta 94503142 con la tesis titulada "Evolución de rutas metabólicas y teoría de redes", realizada bajo la dirección del DR. ARTURO CARLOS II BECERRA BRACHO:

Presidente: DR. GERMAL COCHO GIL  
Vocal: DR. LUIS JOSÉ DELAYE ARREDONDO  
Secretario: DR. PEDRO EDUARDO MIRAMONTES VIDAL  
Suplente: DRA. LORENA DEL CARMEN CABALLERO CORONADO  
Suplente: M. EN C. SARA ERNESTINA ISLAS GRACIANO

Sin otro particular, me es grato enviarle un cordial saludo

ATENTAMENTE  
"POR MI RAZA HABLARA EL ESPIRITU"  
- Ciudad Universitaria, Cd. Mx., a 8 de noviembre de 2018

DR. ADOLFO GERARDO NAVARRO SIGÜENZA  
COORDINADOR DEL PROGRAMA



AGNS/MMVA/ASR/ipp

## **Agradecimientos Institucionales**

Al Posgrado en Ciencias Biológicas, UNAM.

Gracias al apoyo económico otorgado por el CONACYT para realizar mis estudios e investigación de maestría (beca 239898)

A mi Tutor el Dr. Arturo Becerra Bracho y a los miembros del Comité Tutor, los Doctores Pablo Padilla y Pedro Miramontes, por su apoyo y comprensión durante el desarrollo de mi tesis.

## **Agradecimientos personales**

Agradezco a los miembros del laboratorio de Origen de la vida de la Facultad de Ciencias de la UNAM.

Especialmente a los doctores Antonio Lazcano y Arturo Becerra, así como a la M. en C. Sara Islas y al M. en C. José Ervin Silva por su amistad, formación y apoyo.

A la Dra. Daniela Sosa Peredo por su muy valiosa e indispensable colaboración en el desarrollo de la tesis (asesoría especializada y edición en latex), a Daniel Terrones Zapata por su importante trabajo como corrector de estilo, a Norma Loera por su gran apoyo técnico y a Alejandro Carrillo por su apoyo y sus referencias.

Al Dr. José Campillo, y a Carlos González por su ayuda con la presentación.

Agradezco profundamente a mis padres por su apoyo incondicional.

A los doctores Thuluz Mesa Menchaca, María Esther Brandan de Menchaca, Arturo Menchaca Rocha y Alicia González por sus valiosos consejos.

A Marcos Martínez por su oportuna asistencia y a Marcos Ramírez por su ayuda.

A los miembros de mi jurado por revisar mi tesis y sus apoyos.

A Lilia Espinosa, Armando Rodríguez, al Sr. Antonio Suárez y al personal de las ventanillas de posgrado por su comprensión y ayuda.

A los miembros de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, tanto como a los miembros del comité de atención a las personas con discapacidad.

**Dedico esta tesis a Jorge Sierra Calderón y Rosa María Menchaca Rocha,  
mis queridos padres.**



## Contenido

Resumen .....	12
Abstract.....	13
Introducción.....	14
Objetivo .....	15
1 Antecedentes.....	16
1.1.3 Otras teorías.....	18
1.2 La biosíntesis de las purinas como metabolismo antiguo y su evolución .....	20
1.3 Biosíntesis <i>de novo</i> de purinas .....	25
1.3.1 Aspectos importantes de las purinas y su biosíntesis.....	25
1.3.2 La síntesis de IMP .....	27
1.3.3 Formación de ribonucleótidos de Adenina y Guanina .....	29
1.4 Redes .....	31
1.4.1 Redes de similitud de secuencias .....	34
2 Metodología.....	37
2.1 Metodología primera parte: Análisis de similitud de las secuencias de RBNP en una muestra de genomas completos.....	37
2.2 Segunda Metodología: Construcción de redes (SSN) con la similitud de secuencias reportadas en la base de datos KEGG (paralogs de la RBNP de <i>e.coli</i> ) .....	42
3.1 Resultados de redes chicas .....	46
3.1.1 Ribose-phosphate pyrophosphokinase (PrsA) .....	47
3.1.2 Amidophosphoribosyltransferase(PurF) .....	49
3.1.3 Phosphoribosylamine–glycine ligase(PurD) .....	53
3.1.4 Phosphoribosylglycinamide formyltransferase 1 (PurN) .....	56
3.1.5 Phosphoribosylformylglycinamide synthase (purL).....	59
3.1.6 Phosphoribosylformylglycinamide cyclo-ligase (purM).....	63
3.1.7 5-(carboxyamino)imidazole ribonucleotide synthase(purK).....	65
3.1.8 Phosphoribosylaminoimidazole-succinocarboxamide synthase (purC).....	68
3.1.9 Adenylosuccinate lyase (purB).....	71
3.1.10 phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase (purH).....	73
3.2 Resultados de las redes grandes .....	77
3.2.1 Ribose-phosphate pyrophosphokinase (PrsA) .....	79
3.2.2 Amidophosphoribosyltransferase(PurF) .....	84

3.2.3 Phosphoribosylamine–glycine ligase(PurD) .....	88
3.2.4 Phosphoribosylglycinamide formyltransferase 1 (PurN).....	93
3.2.5 Phosphoribosylformylglycinamide synthase (purL) .....	97
3.2.6 Phosphoribosylformylglycinamide cyclo-ligase (purM).....	101
3.2.7 5-(carboxyamino)imidazole ribonucleotide synthase(purK).....	105
3.2.9 Adenylosuccinate lyase (purB) .....	108
4. Discusión .....	114
4.1 Discusión general de la metodología .....	114
4.2 Discusión de los aportes de la primera metodología.....	114
4.3 Discusión de los aportes de la segunda metodología .....	115
4.4 Discusión de la organización de la estructura de las redes de similitud de secuencias (módulos y relaciones de similitud entre un módulo y otro).....	116
4.5 Discusión sobre las implicaciones evolutivas de los resultados del primer análisis realizado (Redes Chicas).....	118
4.5.1 Discusión sobre las diferencias del grado de conservación evolutiva y distribución filogenética entre las enzimas de la ruta de biosíntesis de purinas .....	119
4.5.2 Enzimas fusionadas (importancia e implicaciones evolutivas de su distribución filogenética) .....	122
4.5.3 Duplicaciones .....	126
Conclusiones.....	129
5.1 Conclusiones del análisis de las redes chicas.....	129
5.2 Conclusiones del análisis de las redes grandes .....	129
5.3 Conclusiones finales .....	130

# Índice de Imágenes

## 1. Lista de Figuras

Figura 1.1 Modelo de Horowitz	17
Figura 2 Comparación de los modelos de Patchwork y Horowitz	19
Figura 3 Fusiones enzimáticas de PurM, PurN y PurD y Fusión génica y de <i>PurL</i> grande y <i>PurL</i> chica.	22
Figura 4: Fusiones entre las enzimas PurK, PurC y PurE. [42]	23
Figura 5: Estructura de las purinas	25
Figura 6: Biosíntesis <i>de novo</i> de purinas desde el compuesto PRPP hasta IMP	26
Figura 7: Precursores de IMP	30
Figura 8: Síntesis de GTP y ATP a partir de IMP	31
Figura 9: Redes libres de escala y redes aleatorias.	33
Figura 10: Clusters	35
Figura 11: Diagrama de flujo para la primera parte de la metodología	39
Figura 12: Características para construir los esquemas de las relaciones entre los homólogos de cada enzima de la RBNP (Metodología 1)	40
Figura 13 Diagrama de Flujo para la segunda parte de la metodología	44
Figura 14: Características para construir los esquemas de las relaciones entre esquemas de los homólogos de cada enzima de la RBNP (Metodología 2)	45

## 2 Lista de gráficas de redes

Red 1 chica de enzimas homólogas de <i>prsA</i>	48
Red 2 chica de enzimas homólogas de <i>purF</i>	¡Error! Marcador no definido.
Red 3 chica de enzimas homólogas de <i>purD</i>	54
Red 4 chica de enzimas homólogas de <i>purN</i>	57
Red 5 chica de enzimas homólogas de <i>purL</i>	60
Red 6 chica de enzimas homólogas de <i>purM</i>	64
Red 7 chica de enzimas homólogas de <i>purK</i>	66
Red 8 chica de enzimas homólogas de <i>purC</i>	69
Red 9 chica de enzimas homólogas de <i>purB</i>	72
Red 10 chica de enzimas homólogas de <i>PurH</i>	75
Red 11 grande de enzimas homólogas de <i>prsA</i>	81
Red 12 grande de enzimas homólogas de <i>purF</i>	85
Red 13 grande de enzimas homólogas de <i>purD</i>	90
Red 14 grande de enzimas homólogas de <i>PurN</i>	94
Red 15 grande de enzimas homólogas de <i>purL</i>	98
Red 16 grande de enzimas homólogas de <i>purM</i>	102
Red 17 grande de enzimas homólogas de <i>purK</i>	105
Red 18 grande de enzimas homólogas de <i>PurB</i>	108
Red 19 grande de enzimas homólogas de <i>purH</i>	111

### 3 Tablas

Tabla 3. 1: Lista de organismos candidatos usados en la primera parte de la metodología.....	41
Tabla 3. 2.2 Enzimas de la Ruta <i>de novo</i> de biosíntesis de purinas de <i>Escheri- chia coli</i> O6:K2:H1 CFT073 (UPEC).....	43
Tabla 3. 3 Resultados de la red chica de las enzimas homólogas a PrsA de diferentes organismos...	49
Tabla 3. 4 Resultados de la red chica de las enzimas homólogas a PurF de diferentes organismos. ....	52
Tabla 3. 5: Resultados de la red chica de enzimas homólogas a PurD de diferentes organismos ....	55
Tabla 3. 6 Resultados de la red chica de las enzimas homólogas a PurN de diferentes organismos. ....	58
Tabla 3. 7 Resultados de la red chica de las enzimas homólogas a PurL de diferentes organismos. ....	61
Tabla 3. 8 Resultados de la red chica de las enzimas homólogas a PurM de diferentes organismos	64
Tabla 3. 9 Resultados de la red chica de las enzimas homólogas a PurK de diferentes organismos.	67
Tabla 3. 10 Resultados de la red chica de las enzimas homólogas a PurC de diferentes organismos ....	70
Tabla 3. 11 Tabla de resultados de la red chica de las enzimas homólogas a PurB de diferentes organismos. ....	73
Tabla 3. 12 Tabla de resultados de la red chica de las enzimas homólogas a PurH de diferentes organismos ....	76
Tabla 3. 13 Principales nodos de la red grande directamente relacionados con prsA en el genoma de E.Coli.....	83
Tabla 3. 14 Principales nodos de la red grande directamente relacionados con PurF en el genoma de E.coli.....	87
Tabla 3. 15 Principales nodos de la red grande directamente relacionadas a PurD en el genoma de E.coli.....	92
Tabla 3. 16 Principales nodos de la red grande directamente relacionadas a PurN en el genoma de E.coli.....	96
Tabla 3. 17 Principales nodos de la red grande directamente relacionados con PurL en el genoma de E.coli .....	100
Tabla 3. 18 Resultados de la red de enzimas homologas directamente relacionadas con Pur M dentro del genoma de <i>E.coli</i> . ....	104
Tabla 3. 19 Principales nodos de la red grande directamente relacionados con PurK en el genoma de <i>E.coli</i> .....	107
Tabla 3. 20 Principales nodos de la red grande directamente relacionadas con purB en el genoma de <i>E.coli</i> .....	110
Tabla 3. 21 Principales nodos de la red grande directamente relacionados con PurH en el genóma de E.Coli .....	113

## Resumen

La evolución de las enzimas y de las rutas metabólicas se ha venido estudiando desde mediados del siglo XX a través de los análisis de comparación de secuencias moleculares, lo que nos ha permitido principalmente reconstruir la evolución de diferentes enzimas de manera individual mediante la construcción de árboles filogenéticos. Sin embargo, en este trabajo abordé el problema bajo una perspectiva de redes de similitud de secuencias, lo que me permitió tener una visión entre diferentes grupos de enzimas de diferentes rutas metabólicas. Se consideró como referencia de estudio las enzimas de una ruta metabólica muy conservada: la de la biosíntesis *de novo* de purinas. Con cada una de estas enzimas se realizaron dos tipos diferentes de redes de similitud de secuencias, cada una basada en diferentes valores de restricción en la comparación de secuencias (80% vs. 30% de similitud de secuencias de aminoácidos). Entre los principales resultados de esta investigación se encontró que las redes de similitud de secuencias, hechas con valores muy estrictos de restricción tuvieron muy pocos nodos, lo que es un indicio del alto grado de conservación de dicha ruta. En cambio, con el segundo tipo de redes de similitud, pudimos obtener el material necesario para la elaboración de redes con un número suficiente de nodos, cuya estructura formó módulos de enzimas conectadas entre sí que compartieron uno o más dominios, así como la misma actividad enzimática general del código de enzimas (EC). Resultado que me permitió inferir que podrían ser más comunes las duplicaciones a menor escala que las de una enzima completa, como las que se dan a nivel de dominios o incluso menores.

## **Abstract**

The evolution of metabolic pathways has been studied during the last five decades, using molecular sequence comparison analysis, which has given us an individualized overview of enzyme evolution by the construction of their phylogenetic trees. However by exploring a sequence similarity network approach (SSN) allowed us to visualize the sequence similarity relatedness of different groups of enzymes at same time, and at different scales. The Purine biosynthetic pathway was used, being an example of an early and conserved metabolic pathway. Based on amino acid sequence similarities two cut off values were selected in order to construct two different types of (SSN) networks (80% and 30% of similitude) for each enzyme from this pathway. The main results of this research show that the sequence similarity networks, made with very high restriction values, had scarce nodes (an indication of the conservation of the purine metabolic pathway). On the other hand, with the second type of similarity networks, it was possible to obtain enough data for the construction of real networks, having more nodes. The structure of these networks showed different clusters of enzymes shearing one or more protein domains, as well as the same general catalytic function (first enzyme's code number, EC). The meaning of these results is that duplications that occur on a smaller scale, such as domain duplications, could be more frequent than expected, more than those of complete enzymes

## Introducción

El tema de la evolución temprana de las enzimas y de las rutas metabólicas se ha venido explorando desde hace varias décadas, sin embargo de los años noventa al presente, esta área se ha podido desarrollar aún mejor gracias a la gran cantidad de genomas que se han secuenciado, a las bases de datos que desde hace tiempo guardan dicha información (Kegg, NCBI) y a la creación de herramientas bioinformáticas como las tipo blast que hasta hoy nos han permitido realizar comparaciones muy específicas del conjunto de enzimas pertenecientes a cada ruta metabólica, a través de sus secuencias.

Estos métodos tradicionales de comparación de secuencias han sido eficaces para conocer la historia evolutiva de las enzimas, especialmente las más conservadas, y para hacer inferencias basadas en la homología de las enzimas sobre el origen y evolución de las rutas metabólicas.

Gracias a dichos métodos se han logrado avances, como por ejemplo, la reconstrucción de algunas rutas metabólicas como la de la histidina [7]; el descubrimiento de muchas de las enzimas más conservadas que existen, y que pudieron estar presentes en el último ancestro común, además del descubrimiento de evidencias trascendentales sobre la importancia que pudieron haber tenido los modelos de Patchwork y Horowitz en el origen y evolución de las rutas metabólicas [7]. Sin embargo los métodos mencionados, de comparación de secuencias, por sí solos resultan estar limitados para resolver ciertos problemas evolutivos, como los que han surgido con los últimos hallazgos hechos en esta área, expuestos a continuación:

- a) Las enzimas pueden presentar diferentes grados de similitud, completa o a nivel de dominios.
- b) Los dominios podrían constituir unidades evolutivas independientes, como se ve en el proceso evolutivo conocido como “domain shuffling” y en las enzimas multidominio.

Estos hallazgos nos muestran un esquema más complejo sobre la evolución de las enzimas y de las rutas metabólicas del que se había dilucidado en años anteriores.

Por otra parte, las metodologías tradicionales, se encuentran limitadas en cuanto a que:

- 1) No pueden explicar una visión integrada sobre la evolución del metabolismo como tal [11]. Hasta ahora se entiende, por ejemplo, que las rutas metabólicas no están aisladas unas de otras dentro de los organismos, sino que forman parte de una red muy compleja de reacciones metabólicas interconectadas entre sí.
- 2) No se ha logrado comprender de manera conjunta y en su totalidad la dimensión real de la evolución por duplicación a diferentes escalas.
- 3) No ha sido posible visualizar de una manera integrada las relaciones evolutivas entre las secuencias de diferentes enzimas lejanamente relacionadas.

Por ello se ha hecho necesaria la creación de nuevas herramientas que puedan dar un nuevo enfoque a este tipo de problemas, como las que se han propuesto mediante la construcción de grafos, como las redes de redes de similitud de secuencias, un método reciente y muy poco explorado que integra los métodos de comparación de secuencias con la construcción de redes basadas en el parecido de las secuencias de las enzimas.

En conclusión si deseamos, entender mejor el origen y la evolución del metabolismo como tal, así como sus restricciones, patrones de organización etc, además de lograr tener una visión integrada sobre la evolución de las secuencias de las enzimas visualizando en conjunto el parecido que tienen entre sí, unas con otras, es necesario hacer más estudios como este y tener en cuenta otras investigaciones que se han hecho en los últimos años, que vinculen la visión que plantea la evolución del metabolismo de manera integral [11] [24] [2][26][41].

## **Objetivo**

El objetivo fundamental de este trabajo es estudiar los posibles procesos que participaron en la evolución de la biosíntesis *de novo* de purinas (RBNP) y otras enzimas relacionadas con dos metodologías distintas, una de comparación de secuencias tradicional con valores de comparación muy estrictos donde se incluyeron varios organismos, y otra de comparación de secuencias con parámetros más laxos complementada con la construcción de redes de similitud de secuencias (SSN), donde solo se utiliza *E.coli* como organismo modelo.



## **1 Antecedentes**

### **1.1 Origen y evolución de rutas metabólicas y del metabolismo en general**

A lo largo del tiempo, el estudio de la evolución del metabolismo se ha centrado en el origen de las rutas metabólicas, por ello, se han propuesto varias teorías que explican el surgimiento y la evolución de estos procesos, como la Hipótesis de Patchwork [18] [39] [19]; aunque la mayoría de estas teorías han tomado como base teórica común la duplicación y la divergencia de las enzimas, consideradas en conjunto como la fuerza evolutiva más importante a este nivel.

#### **1.1.1 Hipótesis de Horowitz**

Norman Horowitz, genetista norteamericano, fue él el primer autor en dar una explicación acerca del origen y evolución temprana de las rutas metabólicas [18]. Horowitz sugirió que dicha evolución fue el resultado de la adquisición y acumulación secuencial de enzimas en orden inverso al que estas actualmente poseen dentro de una vía dada. Basándose en las ideas de Oparin, esta propuesta establece un vínculo entre el ambiente primitivo y el surgimiento del metabolismo biológico, planteando que los intermediarios bioquímicos de las rutas metabólicas ya estaban presentes en el medio prebiótico. Años después, con el descubrimiento de los operones, Horowitz pudo desarrollar mejor su propuesta, conocida en ese momento como “La Hipótesis Retrógrada”. Afirmó que el agrupamiento de algunos genes biosintéticos es el resultado de una sucesión de duplicaciones genéticas en tandems. Sin embargo, años después, fueron encontrados una serie de problemas al planteamiento de su hipótesis, con lo cual surgieron otras propuestas [7].

## CAPÍTULO 1. ANTECEDENTES

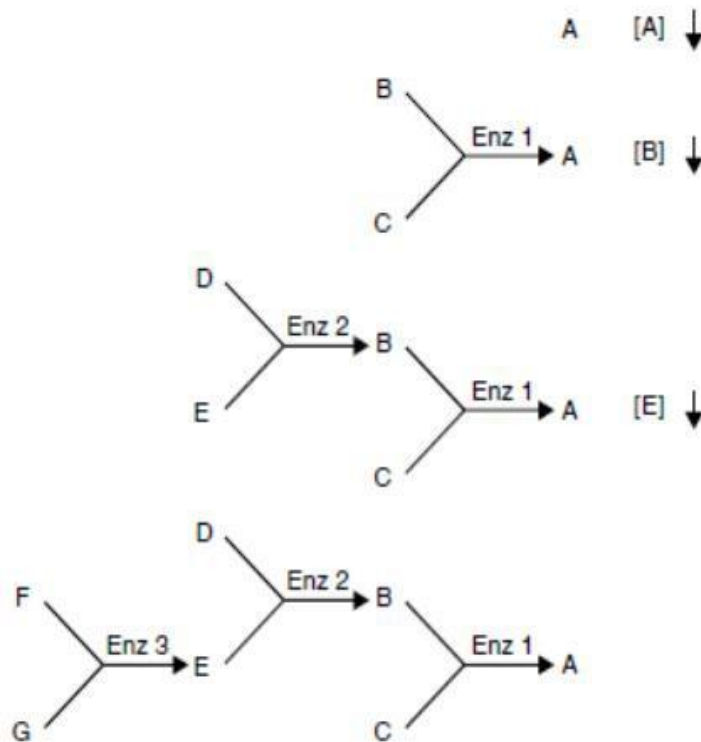


Figura 1.1 Modelo de Horowitz

La segunda teoría que se propuso sobre la evolución de las rutas metabólicas fue la teoría del patchwork, cuyos autores fueron Waley Ycas y Jensen[39][19]. Esta teoría supone que las rutas biosintéticas fueron ensambladas por un mecanismo tipo Patchwork (bricolaje) en el que participaron enzimas primitivas sin una afinidad específica por el sustrato. Dichas rutas primitivas, por lo tanto, estaban compuestas por enzimas que podían utilizar diferentes sustratos parecidos (Figura 1.2) De este modo las secuencias de dichas enzimas evolucionaron mediante la divergencia y la duplicación génica, mecanismos que ayudaron a incrementar el tamaño de los genes y la especificidad de las enzimas por el sustrato. Una de las huellas moleculares que probaron que el mecanismo de patchwork jugó un papel importante en el ensamblaje de las rutas metabólicas primitivas, se encontró en la ruta de la biosíntesis de la histidina, donde se vio que la ATP fosforibosiltransferasa codificada por el gen *hisG* es homóloga a otras fosforibosiltransferasas de diferentes rutas (anabolismo de nucleótidos) demostrando que dicha ruta se ensambló a partir del reclutamiento de genes que producían enzimas poco especializadas

pertenecientes a grandes familias y probando que dicho mecanismo está muy generalizado, observándose igualmente en el metabolismo de purinnucleótidos, donde la duplicación génica ha jugado un papel importante en su evolución temprana[7] [13].

### 1.1.3 Otras teorías

La evolución de las rutas metabólicas ha sido explicada por otras teorías. Una de ellas es la de Granick, quien propone que la formación de rutas metabólicas se da de principio a fin asumiendo que la síntesis de compuestos más simples precede la síntesis de compuestos más complejos. Esto quiere decir que las enzimas que están al inicio de la ruta metabólica evolucionaron primero. Como puede verse, esta propuesta es ya muy diferente a la de Horowitz. Además, Granick no considera que los compuestos prebióticos hayan jugado un papel importante en la evolución de las rutas metabólicas. Un ejemplo que cumple con este modelo es la ruta para la formación de la clorofila [15]. Por su parte, Roy Sidharta considera que hubo una evolución retrograda por saltos que se dio a partir de enzimas multifuncionales. Él asume que las primeras rutas metabólicas no evolucionaron a partir de enzimas generadas *de novo*, porque sus intermediarios eran muy inestables, por lo que más bien, para que estos no fueran degradados, se requería reclutar rápidamente enzimas multifuncionales ya existentes que pudieran procesar inmediatamente los productos generados por las primeras enzimas. Igualmente, propone que a partir de la duplicación de una sola enzima multifuncional que podía catalizar varias reacciones consecutivas, pudieron originarse todos los pasos de una sola ruta metabólica. Así mismo, a partir de una sola enzima multifuncional se pudieron generar enzimas precursoras para distintas rutas metabólicas [30]. Lazcano y Miller proponen que el origen de los compuestos biológicos, ahora sintetizados por enzimas, pudieron haberse dado a partir de reacciones tanto semienzimáticas como no enzimáticas. Esta propuesta se da como resultado de un experimento en el que se logró el crecimiento prototrófico de

## CAPÍTULO 1. ANTECEDENTES

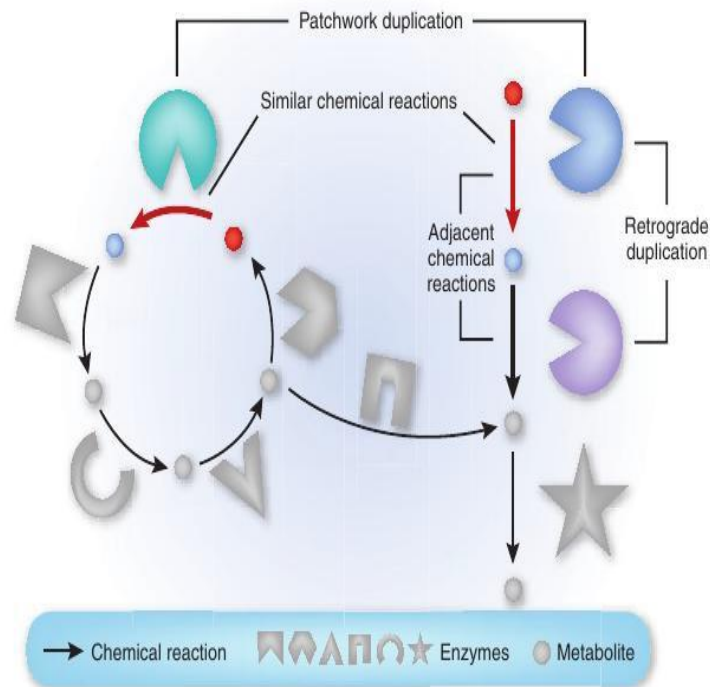


Figura 2 Comparación de los modelos de Patchwork y Horowitz

Este esquema detalla las dos teorías prevalentes acerca de cómo la duplicación de genes puede explicar la formación de una red metabólica. Las enzimas duplicadas que participan en reacciones adyacentes se dice que han sido objeto de “evolución retrgrada”, mientras que las enzimas duplicadas provenientes de aquellas enzimas que están involucradas en reacciones similares se dice que han sido objeto de “evolución patchwork”. En el esquema las proteínas homólogas se indican mediante la misma figura iluminada con diferentes colores y los sustratos metabólicos, productos y reacciones similares están iluminados con el mismo color [24].

una sepa de *Klebsiella pneumoniae*, a la que habiéndole mutado el gen *HisH* se le sometió a una concentración elevada de amonio ( $\text{NH}_3$ ), concluyendo que la primera reacción de la biosíntesis de la histidina se pudo haber llevado a cabo solo con la presencia del  $\text{NH}_3$  antes del surgimiento de *HisH* [23]. Además de las teorías antes mencionadas, los autores S. Sunyaev y S. Schmidt [31]. sugieren que hay otros factores que han jugado un papel importante en la evolución del metabolismo. Uno de ellos es la existencia de una sobre representación de ciertos metabolitos esenciales en la bioquímica de los seres vivos (ATP, NAD, NADH, etc.), conocidos como

metabolitos *Hub* porque conectan una gran cantidad de reacciones bioquímicas en una sola red metabólica que incluye a las rutas más importantes del metabolismo central. Dicho fenómeno implica que habría una mayor probabilidad de que una nueva ruta metabólica se originase a partir de enzimas capaces de utilizar estos intermediarios, a diferencia de enzimas (poco abundantes) únicamente aptas para procesar metabolitos poco comunes. Evidencias de esta propuesta se han encontrado en rutas metabólicas con una evolución muy reciente como la de la degradación del pentaclorofenol (PCP). Es por esto que dichos autores no descartan que la evolución temprana de las rutas metabólicas pudo haber estado influenciada desde un principio por un sesgo de este tipo, en donde los intermediarios más abundantes del metabolismo primitivo y algunas restricciones de tipo estructural y bioquímica, impuestas por los mismos, limitaron desde etapas muy tempranas el espacio de posibilidades evolutivas de la formación de rutas metabólicas [31].

La evolución y el reclutamiento de las enzimas que participan en las rutas metabólicas resultan fenómenos tan complejos que se ha sugerido que las hipótesis mencionadas anteriormente no son excluyentes, sino que pudieron haber sido complementarias e incluso cada una pudo haber sido predominante en diferentes etapas de la evolución temprana del metabolismo [17][12].

## **1.2 La biosíntesis de las purinas como metabolismo antiguo y su evolución**

La ruta *de novo* de ribonucleótidos purínicos además de resultar indispensable para la gran mayoría de los seres vivos [9]; es una ruta de las más conservadas y de mayor distribución a nivel filogenético [16]; lo cual nos sugiere que pudo haber estado presente en el cenestro [7]; aunque sus mismos productos (adenina, guanina, xantina e hipoxantina) y algunos de sus precursores (AICA y AICAR) pudieron ser mucho más antiguos, habiéndose sintetizado prebióticamente [9]. Por otro lado, el grado de conservación de la ruta *de novo* de ribonucleótidos purínicos resulta ser mayor con respecto a sus intermediarios que en cuanto a sus enzimas, habiendo como consecuencia una mayor variedad evolutiva en estas últimas[7].

## CAPÍTULO 1. ANTECEDENTES

La distribución filogenética y el nivel de conservación de cada enzima resulta ser variable [41]. Por ejemplo, mientras que las enzimas *PRSA*, *PurF*, *PurD*, *PurL*, *PurM*, *PurC* y *PurB* están distribuidas en todos los organismos, hay otras que tienen una distribución particular como son *PurN* o *PurT*, *PurK/PurE-I* o *PurE-II*, *PurH* o *PurP* y *PurJ* o *PurO*, siendo estas últimas probablemente las menos conservadas.

Otro aspecto evolutivo que se ha considerado de dicha ruta es la relación evolutiva que existe entre sus enzimas tomando en cuenta la similitud entre sus dominios, con lo cual se les ha clasificado en 2 grandes superfamilias principales:

- 1) ATPgrasp: Incluye a las enzimas *PurD*, *PurT*, *PurK* y *PurP*. Todos sus miembros tienen los dominios A, B y C y poseen el fold ATP-grasp, único de este grupo.
- 2) PurM: Contiene a las enzimas *PurL* y *PurM*, las cuales tienen un ple- gamiento similar y un motivo muy conservado de unión a ATP.

Se cree que los miembros de cada una de estas superfamilias evolucionaron a partir de una enzima ancestral generalista diferente para cada superfamilia [20].

Recapitulando lo anterior, es claro que la evolución de cada una de las enzimas de la presente ruta, además de distinguirse por no ser homogénea en cuanto a su nivel de conservación filogenético y su relación evolutiva a nivel superfamilia, tampoco resulta homogénea en cuanto a sus mecanismos evolutivos, siendo estos algo diversos; entre los cuales se pueden identificar:

- 1) Convergencias Funcionales: un ejemplo son las convergencias que existen entre las amidotransferasas *PurF* y *PurL*, entre las enzimas dependientes de tetrahydrofolato *PurN* y *PurH* y entre las ciclohidrolasas *PurJ* *PurO*.
- 2) Duplicaciones de tres tipos:
  - a) Completas, como las que se presentan en las enzimas *PurF* y *PurF'*, *PurE I* y *PurE II*, *PurM* y *purM''*, hallado en drosófila, y *PurE* y *PurE'* encontrado en bacterias metanógenas.
  - b) De tipo múltiple, como las duplicaciones del gen *PurP* que en algunas arqueas se

encuentra presente en múltiples copias (ortólogas).

c) Parciales, como son las duplicación que encontramos en los genes de *PurL* y *PurM*. Sin embargo no es completa ya que la estructura de *PurM* es un homodímero y cada uno de sus monómeros se parece a la mitad de la estructura de *PurL* [35].

3) Fusiones génicas y fusiones enzimáticas: son re arreglos en la organización postranscripcional de genes. Estos tipos de fusiones forman proteínas multifuncionales como las que se mencionan a continuación:

a) *PurL corta* y *PurL larga*, mientras que la primera es un complejo enzimático formado por el ensamblaje de los productos de los genes *PurL*, *PurQ* y *PurS*, la segunda enzima es el resultado de la fusión completa pero a nivel génico entre dichos genes (*PurL*, *PurQ* y *PurS*). (Figura 1.3)

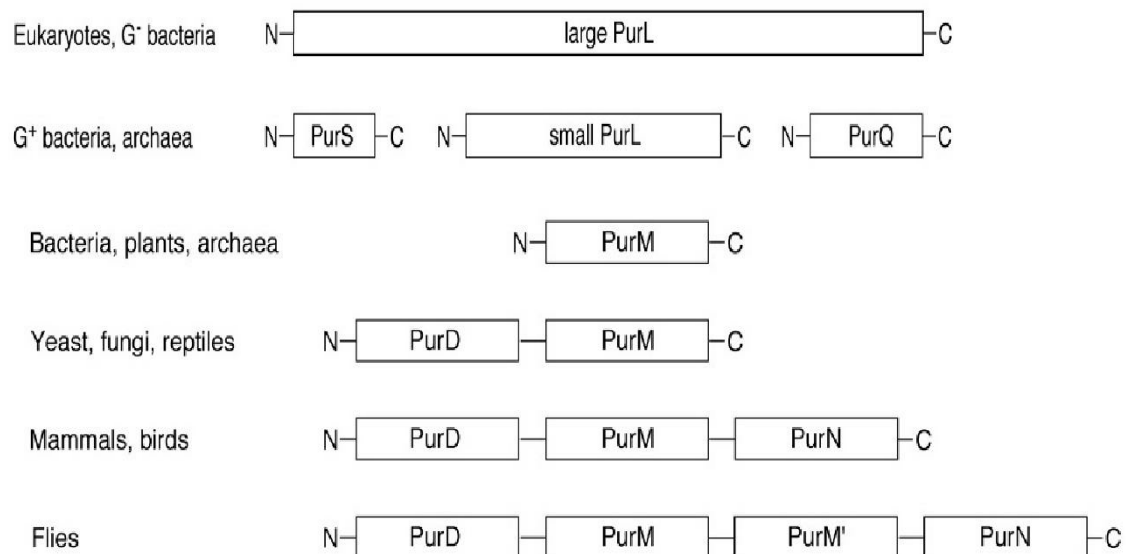


Figura 3 fusiones enzimáticas de PurM, PurN y PurD y Fusión génica y enzimática de *PurL* grande y *PurL* chica [42].

b) La enzima bifuncional *PurHJ* que es una enzima bifuncional, resultado de una fusión entre los genes *PurH* y *PurJ* [42].

c) El rearrreglo de los genes *PurM*, *PurN* y *PurD* en diferentes organismos. Mientras que en las bacterias, en las plantas y en las arqueas la enzima *PurM* es monofuncional, en las levaduras, los hongos y los reptiles se encuentra fusionada con *PurD* y en los mamíferos con *PurD* y *PurN*.

Por otro lado, en las bacterias pueden coexistir fusionadas *PurN* y *PurT*.

d) Dos ejemplos más de fusiones de enzimas son el de *PurE* clase I y *PurK*; el de *PurE* clase II y *PurC*. En el primer caso *PurE* I y *PurK* forman una enzima bifuncional parcialmente fusionada por una especie de puente molecular. Esta enzima se encuentra presente en las levaduras y las plantas mientras que en la mayoría de las bacterias *purE* I y *PurK* se expresan separadamente [42]. En el segundo caso *PurC* se encuentra fusionada con *PurE* clase II en los vertebrados mientras que en los organismos menos complejos *PurC* es una enzima independiente. Figura 1.5

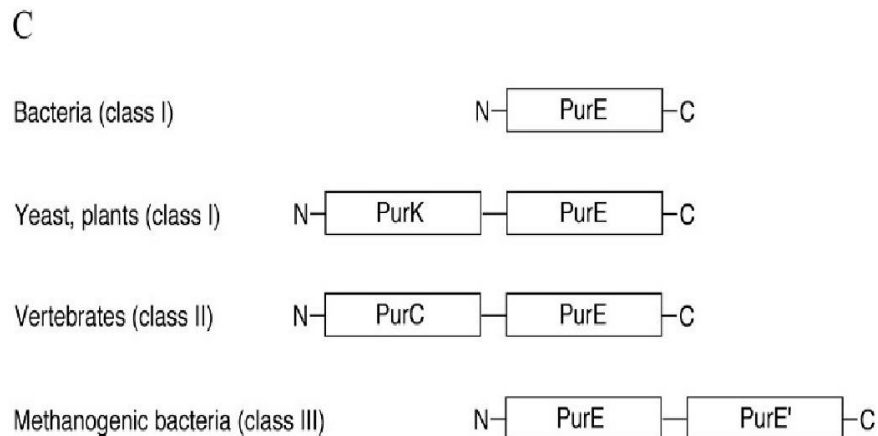


Figura 4: Fusiones entre las enzimas *PurK*, *PurC* y *PurE*[42].



Dichas fusiones son el resultado de adaptaciones que han adquirido los organismos a lo largo del tiempo y, al parecer, no cumplen necesariamente con un patrón filogenético. A través de ellas los organismos han logrado evitar la constante degradación de los metabolitos inestables, como el PRA y el NCAIR evitando su exposición al ambiente al ser transferidos directamente de una enzima a otra a través de una especie de puentes estructurales. Este fenómeno se ve con mayor frecuencia en los eucariontes pero también se puede llegar a observar en bacterias y arqueas.

## CAPÍTULO 1. ANTECEDENTES

### 1.3 Biosíntesis *de novo* de purinas

#### 1.3.1 Aspectos importantes de las purinas y su biosíntesis.

Las purinas son compuestos heterocíclicos orgánicos, constituidos por un anillo purínico y un anillo imidazol, que son esenciales para los seres vivos porque forman parte de sus ácidos nucleicos (DNA y RNA) como la adenina y la guanina, que también son parte de las alarmonas, compuestos muy antiguos que sirven como moléculas energéticas, cofactores biológicos y neurotransmisores [40]. Además existen otras purinas menos importantes, como la hipoxantina y la xantina que funcionan como intermediarios importantes en la síntesis de las mismas purinas.

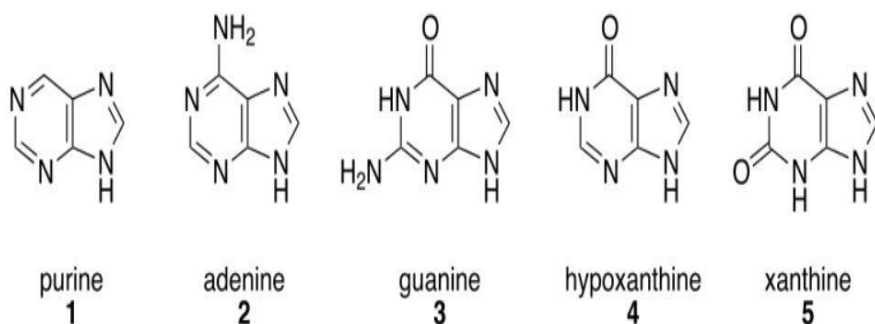


Figura 5: Estructura de las purinas

La característica principal de la formación de las purinas es que éstas se sintetizan como nucleósidos y nunca como bases libres. Existen 2 maneras de sintetizar estas moléculas bióticamente. La primera es de manera gradual a través de la síntesis *de novo* de purinas (Figura 1.7), mientras que la segunda se da mediante atajos metabólicos conocidos como rutas metabólicas de salvamento, donde se reciclan moléculas que son productos de degradación metabólica [8]; como la Adenina, la Guanina y la hipoxantina, la PRPP, la IMP, etc. [43]; transformándolos rápidamente en nucleósidos purínicos, principalmente mediante la acción de enzimas conocidas como PRTasas [44]. Estas pequeñas rutas resultan esenciales para

los organismos que han perdido, de forma total o parcial, la síntesis *de novo* de purinas como algunos parásitos obligados, incluyendo por ejemplo a *Micoplasma genitalium*, *Giardia lamblia*, *Trichomonas vaginalis*, etc. [8][34][36].

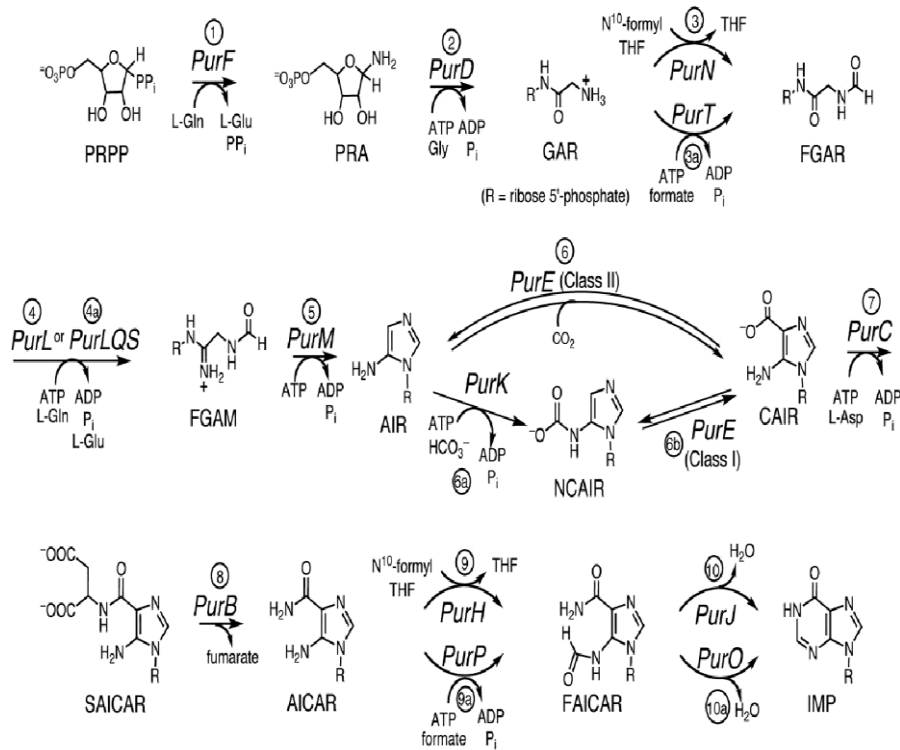


Figura 6: Biosíntesis *de novo* de purinas desde el compuesto PRPP hasta IMP

Tanto la síntesis *de novo* de purinas, como las rutas metabólicas de salvamento suelen ser complementarias, pues además de que se retroalimentan a través de sus intermediarios metabólicos, se regulan entre sí. Al haber mayor concentración de nucleosidos, o bases libres disminuye la actividad de la síntesis *de novo*, mientras que hay una mayor activación de las rutas de salvamento y viceversa. De este modo ambas rutas juegan un papel esencial en la regulación intracelular de la concentración de nucleótidos [7].

La ruta de la biosíntesis *de novo* de purinas es muy importante pues, aparte de ser esencial para la mayor parte de los organismos, tiene características bioquímicas especiales.

En primer lugar, está relacionada con otras rutas metabólicas de igual importancia como:

1.- La biosíntesis de tiamina, con la que comparte el sustrato AIR (ribonucleótido

aminoimidazol).

2.-La biosíntesis de la histidina, con la que comparte el ribonucleótido AICAR (aminoimidazolcarboxamida) y el metabolito PRPP [9]; que conecta la síntesis *de novo* con otras rutas en las que también participa como la biosíntesis del triptófano, las rutas de salvamento de purinas y la síntesis de pirimidinas [38].

En segundo lugar esta ruta se distingue también por su regulación, donde hay 2 retroalimentaciones muy importantes: en la primera parte de la ruta los productos finales AMP y GMP, además de sus derivados, son principalmente inhibidores competitivos de la formación de IMP durante su propia síntesis. En la segunda parte de la ruta, curiosamente el GTP, activa la síntesis de IMP a AMP y contrariamente el ATP activa la síntesis de IMP a GMP. Otra característica de esta ruta es que sus precursores principales son el aspartato, el fumarato, glutamina, glycina y el H<sub>2</sub>O como se muestra en la figura 1.8, y que se divide en dos partes sintetizándose primero el *nucleótido inosina-monofosfato* (IMP) cuya formación se da en 11 pasos y posteriormente se sintetizan separadamente. El AMP y el GMP [38].

### 1.3.2 La síntesis de IMP

Lo que se sabe sobre la síntesis de IMP es que, además de reunir 7 enzimas dependientes de ATP (PurD, PurL, PurM, PurC, PurT, PurK y PurP) integra otras enzimas con funciones que son similares.

La primera reacción que se da para llevar a cabo la síntesis de IMP es la activación de la ribosa-5-fosfato, un producto proveniente de la ruta de las pentosas. La enzima que realiza este paso es la pyrofosfosinasa, que a partir del ATP forma el producto 5-fosforibosylpyrofosfato (PRPP).

En el segundo paso de la biosíntesis de purinas se adquiere el átomo N9 del anillo purínico. La enzima amidofosforilosiltransferasa (Figura PurF) cataliza el PRPP quitándole el grupo pirofosfato para convertirlo en beta-5- fosforibosilamina (PRA). Esta reacción se lleva a cabo en dos partes: en la primera se utiliza glutamina y agua dando como resultado ácido glutamínico + NH<sub>3</sub>; en la segunda reacción el NH<sub>3</sub> se combina con el PRPP dando como

resultado PRA+fosfato (PPi).

En el tercer paso, se adquieren los átomos C4, C5 y N7 de la purina. A partir de PRA y un grupo carboxilo de la glicina, se sintetiza la glicinamida ribonucleotido (GAR) a partir de la GAR sintetasa (PurD), finalmente el grupo carboxilo de la glicina forma una amida con el grupo amino de PRA dando como resultado el compuesto GAR.

En el cuarto paso se adquiere el átomo C8 de la purina. Aquí, la GAR transformilasa (PurN) forma formylglicinamida ribonucleótido (FGAR). En esta reacción el N10-formyltetrahydrofolato dona un grupo formyl el cual es un cofactor que transfiere los átomos de C1 de moléculas como serina, glicina y formato a diferentes tipos de aceptores. En algunos organismos (bacterias y arqueas) PurT es una enzima alternativa a purN.

En el quinto paso se adquiere el átomo N3 de la purina. Comienza al sintetizarse el formylglycinamidina ribonucleótido (FGAM) a partir de la transferencia de un grupo amida que se une al anillo púinico. Esta reacción la lleva a cabo la enzima FGAM sintetaza (PurL).

En el sexto paso se cierra el anillo imidazol mediante una reacción de condensación que requiere ATP dando como resultado la síntesis del ribonucleótido 5-aminoimidazol (AIR). Esta reacción es llevada a cabo por la AIR sintetasa (PurM).

En el séptimo paso se adquiere el átomo C6. En algunos eucariontes se lleva a cabo por la enzima AIR carboxilasa (PurE clase II) la cual utiliza una molécula de CO<sub>2</sub> para formar el grupo carboxilo (HCO<sub>3</sub>) que se une a AIR para obtener finalmente la molécula ribonucleótido caboxyaminoimidazol (CAIR), mientras que en otros organismos como las bacterias, esta reacción se lleva a cabo en dos pasos que están mediados por enzimas independientes: la PurK y la PurE Clase I [29]. La PurK cataliza la carboxilación ATP dependiente del grupo amino N(5) de AIR para dar N5-CAIR, y posteriormente la PurE Clase I hace re arreglos de esta molécula para dar como resultado CAIR.

En el octavo paso se adquiere el N1 de la purina a través de una reacción de condensación en la que participa un aspartato, dando como resultado el ribonucleótido 5-aminoidazol-4-(N-succinilocaroxamida) (SAICAR) que es catalizado por SAICAR sintetasa (PurC). En esta

reacción hay una hidrólisis de ATP similar a la reacción del paso 3 [38].

En el noveno paso se elimina el fumarato. El rompimiento de SAICAR libera al fumarato, dando como resultado el ribonucleótido 5-aminimidazol-4- carboxamida (AICAR) en una reacción catalizada por adenylosuccinato lyasa (PurB).

Los pasos 8 y 9 son similares a las reacciones del ciclo de la urea donde la citrulina es aminada para formar arginina. En ambas rutas metabólicas el grupo amino del aspartato es transferido a un aceptor mediante una reacción de acoplamiento ATP-dependiente, seguida de la eliminación del fumarato. En las plantas y microorganismos el AICAR se forma también en la biosíntesis de histidina.

En el onceavo paso se da la ciclización del anillo purínico para dar como resultado el IMP. Este paso consiste en la eliminación de agua y la catálisis de la IMP ciclohidrolasa (PurJ)[25]. En otros organismos como algunas arqueas los últimos pasos los realizan las enzimas PurO y PurP respectivamente.

### **1.3.3 Formación de ribonucleótidos de Adenina y Guanina**

El IMP tiene la característica de que no se acumula en las células, sino de que se convierte rápidamente a AMP y GMP.

El AMP se sintetiza en dos pasos: en la primera reacción el grupo amino del aspartato se liga al IMP en una reacción en donde se da la hidrólisis del

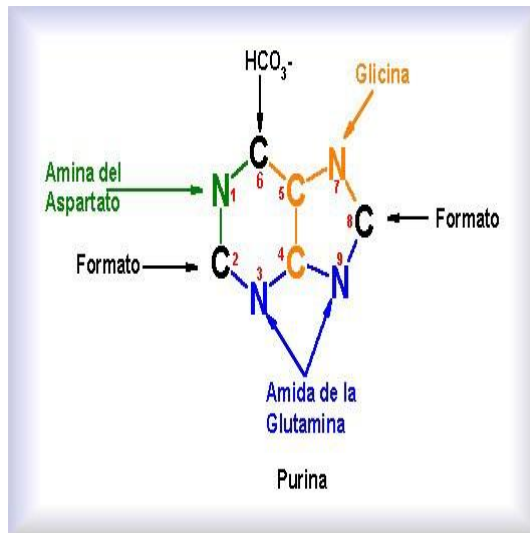


Figura 7: Precusores de IMP

GTP a GDP + P para dar como resultado Adenilosuccinato. En la segunda reacción la *Adenosuccinato liasa* elimina fumarato del Adenilosuccinato para formar AMP. La *Adenilosuccinato liasa* también cataliza el paso 9 en la formación del IMP. A partir del IMP también se sintetiza GMP en dos reacciones: en la primer reacción la *IMP dehidrogenasa* cataliza la oxidación de IMP dependiente de NAD<sup>+</sup> para formar *Xantosina monofosfato* (XMP, el ribonucleotido de la base Xantina); posteriormente el XMP se convierte a GMP a través del reemplazo de su grupo 2-Keto por un nitrógeno de la amida de una amina.

Todo lo anterior sucede en una reacción donde se da la hidrólisis de ATP a AMP + PP. Para ser integrados a las moléculas de ácidos nucleicos, los nucleósidos monofosfatos deben ser primero convertidos a sus correspondientes nucleósidos trifosfatados. Primero, los nucleósidos monofosfatados se fosforilan convirtiéndose en nucleósidos difosfatados y esta reacción es sintetizada por las monofosfato cinasas, (adenilato cinasa y guanilato cinasa). Posteriormente los nucleósidos difosfatos se convierten en sus correspondientes trifosfatos (ATP y GTP) mediante la reacción catalizada por las correspondientes nucleósido fosfato quinasas.

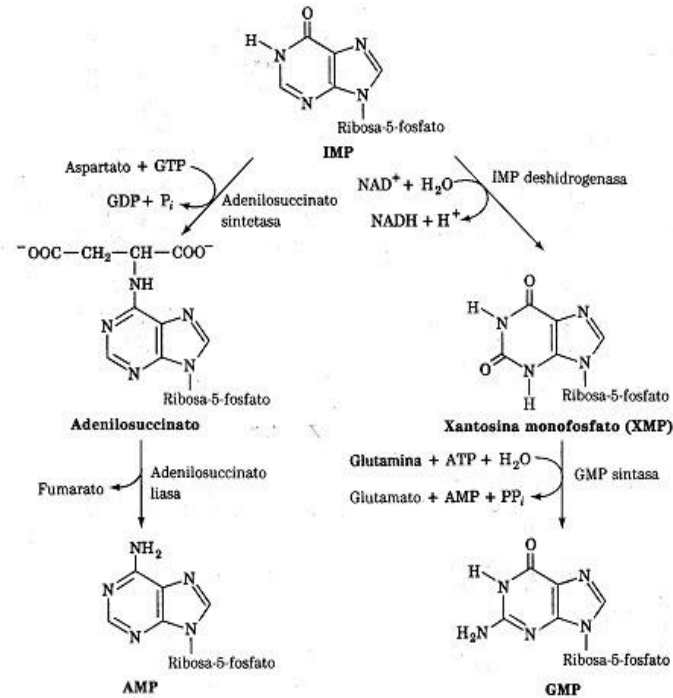


Figura 8: Síntesis de GTP y ATP a partir de IMP

## 1.4 Redes

En los últimos años se ha logrado un gran avance tecnológico científico proveniente del éxito que ha tenido el enfoque reduccionista respecto a la resolución de diferentes problemas científicos. Sin embargo, hasta ahora resulta difícil predecir y entender el comportamiento de sistemas tan complejos como son los sistemas biológicos, analizados exclusivamente como la suma de sus partes más elementales y sus interacciones individuales, por lo que se ha hecho necesario crear nuevos métodos menos reduccionistas que nos permitan estudiarlos mejor y nos sirvan para entender más a fondo aspectos sobre sus propiedades y su dinámica.

Recientemente, la teoría de redes ha resultado ser en este sentido una gran promesa, especialmente aplicada al estudio de los sistemas biológicos, desde las cadenas alimenticias a nivel ecológico hasta las interacciones moleculares que se dan entre los genes, metabolitos y enzimas al interior de una célula [1].

En sí la importancia del enfoque de las redes reside en que, identificando las partes esenciales de un sistema biológico y sus interacciones más importantes, es posible representarlo de manera abstracta como un grafo (mapa gráfico del conjunto de dichas



interacciones) y estudiarlo como tal a través de sus propiedades específicas, permitiéndonos dilucidar el comportamiento de todo el sistema [33].

Un ejemplo de las propiedades que pueden presentar las redes son: las propiedades del mundo pequeño, las propiedades de comunidades como la modularidad y las distribuciones de grado, entre otras. Las redes biológicas pueden tener dichas características, pero particularmente las más importantes de ellas son: el predominio de una distribución libre de escala en cuanto a los nodos más próximos entre sí, y la presencia de una estructura jerárquica en cuanto a los mismos elementos [14]. Por otro lado, es importante mencionar que el origen de todas estas propiedades puede ser explicado mediante tres modelos esenciales de crecimiento de redes entre los cuales están el Modelo Libre de Escala, el Modelo Jerárquico y el Modelo Aleatorio.

La propiedad más importante para hacer el análisis general de una red es el *grado*, definido por el número de conexiones que tiene cada nodo de la red con otros nodos y está representado por el índice  $K$  [11]. Si graficamos el  $K$  de todos los nodos de una red, y obtenemos una distribución normal donde la mayor parte de los nodos de la red tiene un promedio determinado de conexiones que coincide con el punto medio de la curva, podemos inferir que la red con la que estamos trabajando es una red aleatoria. A diferencia de las redes aleatorias, en las redes libres de escala existe un sesgo en la cantidad de conexiones entre los nodos del sistema (distribución de Poisson) que responde a una ley de potencias [3]; donde hay pocos nodos muy conectados y muchos nodos con pocas conexiones. Figura 1.10

Esta cualidad resulta de gran importancia pues nos muestra que los sistemas biológicos son resistentes a perturbaciones aleatorias externas, mientras que son altamente vulnerables a las perturbaciones dirigidas a nodos muy conectados [33].

Las propiedades del mundo pequeño indican que tan interconectadas están las partes del sistema [14]; en donde los nodos que no son vecinos entre sí pueden ser alcanzados desde cualquier nodo origen a través de un número relativamente corto de saltos entre ellos [3]. En el caso de las redes biológicas, con esta propiedad se puede ver qué tanta información comparte toda la red [33].

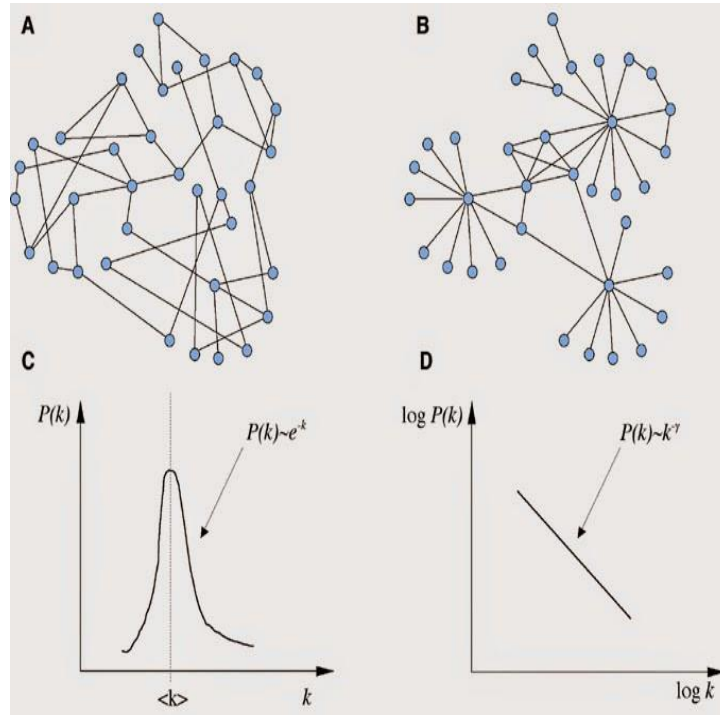


Figura 9: Redes libres de escala y redes aleatorias. Esta figura muestra la diferencia entre una red aleatoria *A* y una red libre de escala *B* con las respectivas gráficas *C* y *D* que indican el número total de conexiones de cada nodo  $p(k)$  dentro de la red.

Otra propiedad importante de algunas redes biológicas es la modularidad, cuya característica principal es la división del sistema en subunidades conocidas como módulos. Una red modular se compone de distintas subredes cuyos elementos comparten más conexiones entre sí, al interior del mismo módulo, que con elementos de otras subredes. Así estas estructuras definen subsistemas que actúan como módulos funcionales, estructurales u organizacionales, más o menos independientes, que se comunican entre sí, y además pueden sufrir cambios internos sin afectar gravemente a los otros módulos del sistema [11]; por ello esta propiedad puede ser fundamental para favorecer la independencia y especificidad de procesos de adaptación, especialización y evolución en los sistemas biológicos [11].

Las redes biológicas más comunes que se estudian a nivel molecular son:

1. Redes de regulación (GRN). Son un conjunto de segmentos de DNA de una célula que interactúan entre sí indirectamente a través de sus RNAs o de sus productos de transcripción (factores de transcripción) y con otras sustancias celulares.

2. Interacciones proteína-proteína. Este tipo de redes se refieren a la asociación de moléculas de proteínas y al estudio de estas asociaciones desde el punto de vista de la bioquímica y el de transducción de señales. Las interacciones físicas entre las proteínas son muy importantes para varias funciones biológicas. Por ejemplo, las señales que vienen desde el exterior de una célula son controladas y transportadas hacia el interior de ésta a través de varias interacciones tipo proteína -proteína con la molécula que viene del exterior actuando como señal. Este proceso también se conoce como transducción de señales. Al parecer en este tipo de redes, los nodos más conectados son moléculas más primitivas probablemente incorporadas al metabolismo en etapas iniciales del origen y evolución de los sistemas biológicos [5]. De este modo el metaboloma podría estar influido por su historia previa.

3. Redes metabólicas. Comprenden el conjunto total de procesos metabólicos y físicos que determinan las propiedades bioquímicas y fisiológicas de una célula. De este modo estas redes comprenden las reacciones químicas del metabolismo, así como las interacciones que guían dichas reacciones. Se ha observado que estas redes presentan propiedades del mundo pequeño, como el ejemplo mencionado anteriormente de los metabolitos *hu* [31].

4.- Redes de duplicación y divergencia. Uno de los pocos autores que ha explorado redes como las de este trabajo, Ricard Solé, hizo un modelo teórico de crecimiento de redes a partir de la duplicación y divergencia de genes que al compararlo con redes reales observó que éste presentaba propiedades del mundo pequeño con el mismo patrón de las redes reales. Sin necesidad de imponer presión selectiva hacia redes menos frágiles, el sistema evolucionó espontáneamente hacia un sistema robusto. También al igual que las redes reales presentaban modularidad.

### **1.4.1 Redes de similitud de secuencias**

Las redes de similitud de secuencias (SSNs) son una herramienta que nos permite visualizar las relaciones de semejanza entre varias secuencias de proteínas diferentes de una manera sencilla, por ejemplo, construyendo una red donde cada secuencia es representada como un nodo y donde el parecido entre una secuencia y otra (o la relación de similitud) se representa con una línea (arista o "edge") [50].

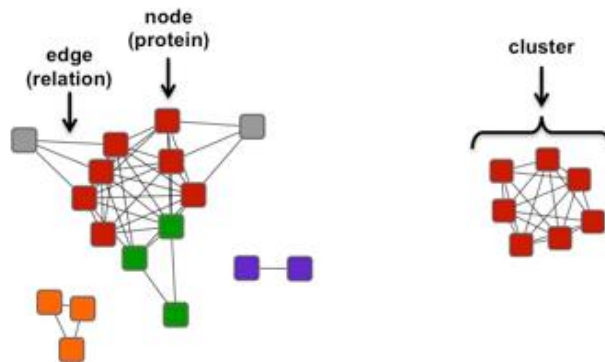


Figura 10: Clusters

Representación de una red de similitud de secuencias en donde los grupos de nodos del mismo color representan cada uno un cluster de proteínas con diferente función.

El uso de estas redes empezó en el 2009 y desde entonces ha sido útil para varios fines, como la visualización y el estudio de la relaciones entre las superfamilias de las enzimas. Estas redes, por ejemplo, fueron utilizadas para crear la clasificación del pfam, en la que las enzimas son clasificadas primero dentro de familias y las familias dentro de clanes. (Medini 10) Esto con el fin de dar un nuevo contexto al análisis de las enzimas individuales.

Las redes de similitud de secuencias también nos han ayudado a hacer predicciones de funciones de proteínas poco conocidas, e identificar homologías lejanas entre diferentes enzimas, pues se ha encontrado que estas redes permiten la percepción de tendencias con respecto a la información ortogonal de las mismas, pudiéndose obtener información relacionada con su función, mapeada en el contexto de similitud de secuencias. Los trabajos de Atkinson y Babbitt por ejemplo, muestran una correlación entre las relaciones de similitud y las de funcionalidad a un nivel muy general de las enzimas. Atkinson (2009) Babbitt.

Con la secuenciación de genomas completos, hoy es posible hacer una reconstrucción de la red de las reacciones bioquímicas que tienen diversos organismos, desde las bacterias

hasta los humanos. Algunas de estas redes como la de *E.coli* y *Bacillus subtilis* actualmente están disponibles en la web mediante servidores como el EcoCyc [22] [4].

En esencia, las redes son herramientas muy útiles que nos pueden servir para estudiar mejor la evolución temprana de las rutas metabólicas desde un punto de vista global, pues las técnicas convencionales de biología molecular, por sí solas, son limitadas y no nos proporcionan gran información acerca de los procesos evolutivos en etapas muy tempranas de la vida.

## 2 Metodología

### 2.1 Metodología primera parte: Análisis de similitud de las secuencias

Para estudiar la evolución de la ruta metabólica de la biosíntesis de purinas y otras enzimas relacionadas, este trabajo se realizó en dos partes y se utilizó una metodología distinta para cada una. La primera parte consistió en hacer un análisis de homología tipo *Blast* (con valor de restricción de 80 % de similitud) de cada una de las enzimas de la ruta *de novo* de la biosíntesis de purinas (RBNP) considerando la secuencias de las enzimas de la sepa de *Escherichia coli* O6:K2:H1 CFT073 (UPEC) que se muestran en la tabla 3.13, como modelo para compararlas contra los genomas de diferentes organismos, mostrados en la tabla 2.1. de RBNP en una muestra de genomas completos.

Estos genomas representan una muestra de aproximadamente 30 organismos de cada dominio, los cuales se eligieron considerando principalmente que no fueran extremófilos, que no tuvieran genomas muy pequeños y no fueran ni parásitos obligados ni endosimbiontes. Sin embargo, hubo algunas excepciones dada la naturaleza de algunos organismos, y el sesgo en la disposición de datos, ya que la mayoría de las arqueas son extremófilas y hay pocas secuenciadas, por otro lado el amplio conocimiento que se tiene de algunas bacterias cuyas características se salían de las restricciones de elección como *B. subtilis*, hizo que las incluyéramos en esta muestra. Posteriormente los resultados del análisis anterior se clasificaron en tablas de Excel de acuerdo a los parámetros descritos en la figura 2.1, y se representaron en esquemas tipo redes (figura 2.2) uno por cada enzima de la RBNP, y se hace referencia a ellas con el nombre de redes chicas.

Finalmente, para cada red chica también se hicieron unas tablas que indican solamente los nombres de las enzimas homólogas más importantes de cada enzima de la RBNP, junto con su Ec, Ko y el tipo de organismos que las presentan; así como las rutas metabólicas a las que pertenecen. La especificación de los organismos varía en cada tabla

por considerarse más relevante, solo en algunos casos, enfatizar las especies y el número de copias de enzimas homólogas que estos presentan.

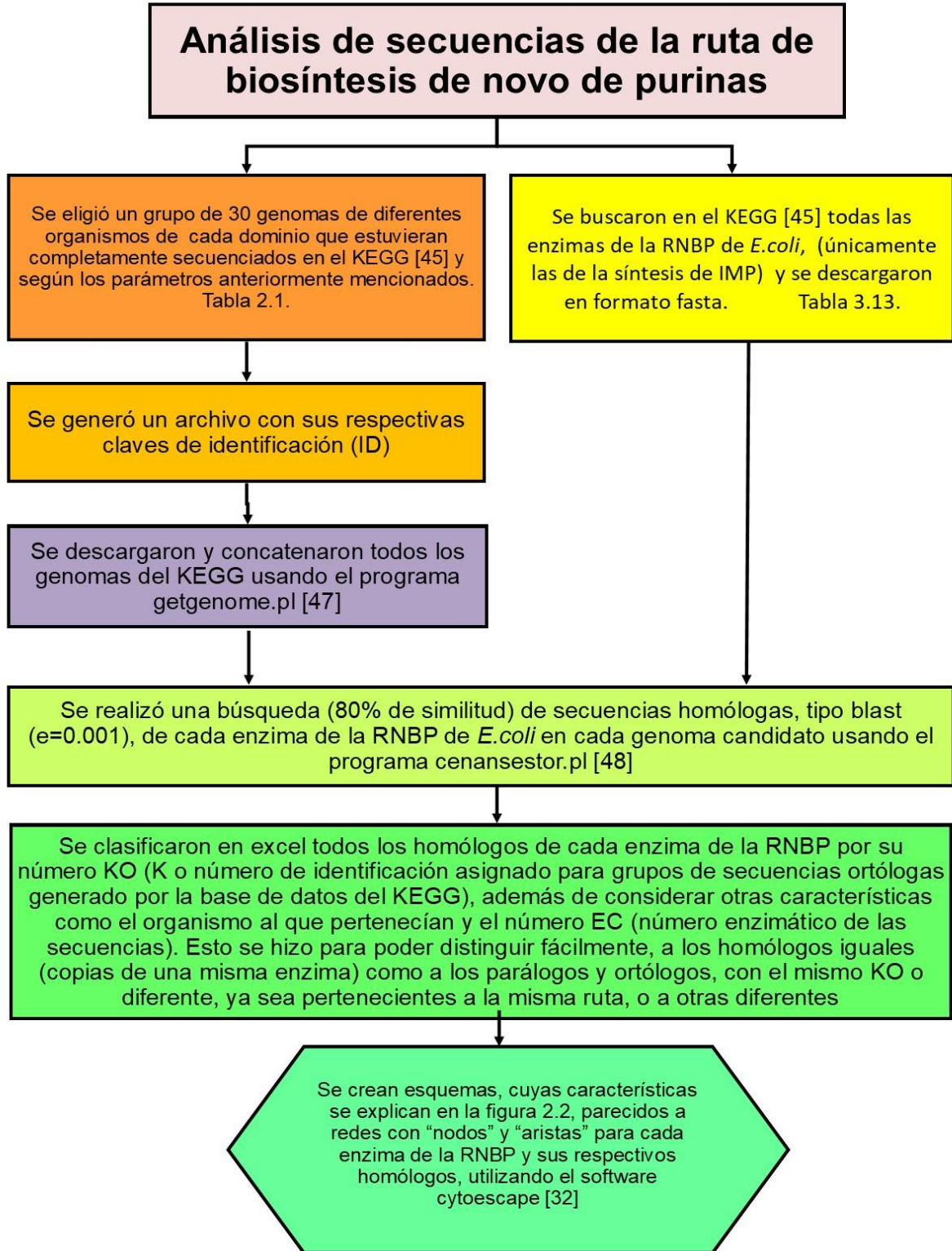


Figura 11: Diagrama de flujo para la primera parte de la metodología



## CAPÍTULO 2. METODOLOGÍA

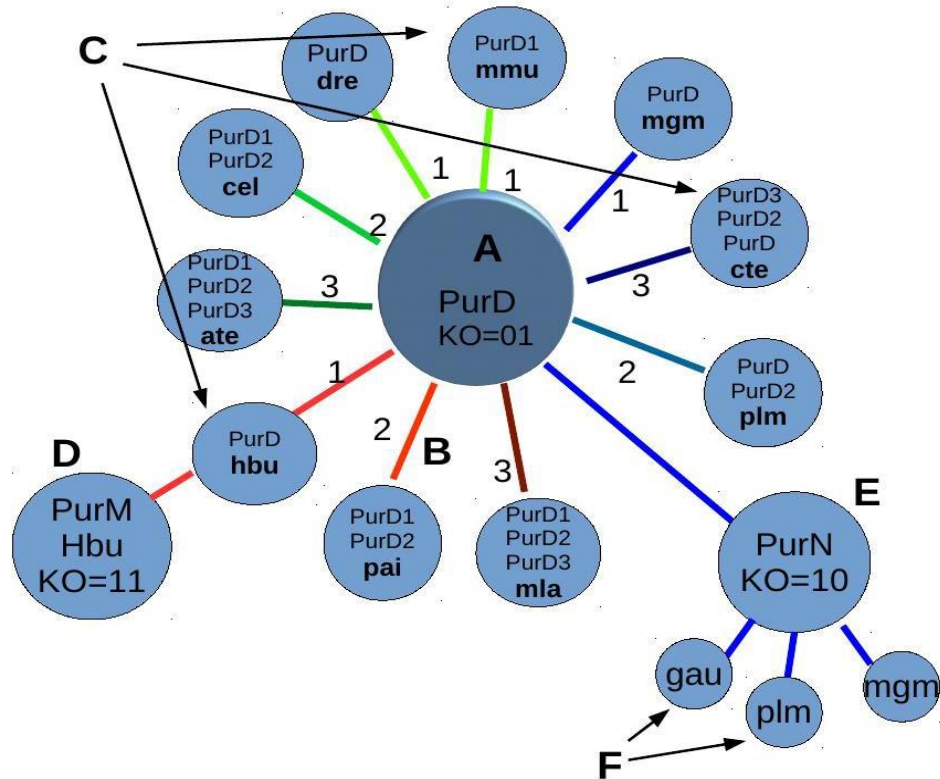


Figura 12: Características para construir los esquemas de las relaciones entre los homólogos de cada enzima de la RBNP (Metodología 1)

Esta figura representa cada una de las características que se consideraron para crear los esquemas de las relaciones de cada una de las enzimas de la RBNP con sus respectivos homólogos. Las enzimas de la RBNP se representan como “nodos” centrales de tamaño grande en cada red (A). Los nodos pequeños (C) representan a los ortólogos individuales y a los grupos de homólogos encontrados en cada organismo que tienen el mismo KO que la enzima principal, y que están conectados a ella (A) por aristas de colores: el rojo representa a las arqueas, el azul a las bacterias y el verde a los eucariontes; la intensidad del color aumenta con el número de copias de homólogos presentes en cada organismo, el cual se ubica sobre cada arista (B). Por otro lado, todos los homólogos de cada enzima de la RBNP que presentaron un KO diferente, tanto los que se encontraron en un solo organismo, como los homólogos que se encontraron en varios, están representados de manera muy particular. Mientras que los primeros están conectados al nodo del organismo al que pertenecen y se representan con un nodo de tamaño mediano (D), los segundos se encuentran directamente conectados al nodo de la enzima principal de cada red y también se representan con un nodo tamaño mediano, etiquetado con su respectivo nombre (E), al que a su vez se unen los diferentes organismos que presentaron ortólogos de esta misma enzima u homólogos

prácticamente iguales (F).

Eucariontes	clave	Arqueas	Clav	Bacterias	clave
<i>Thalassiosira pseudonana</i>	Tps	<i>Nanoarchaeum equitans</i>	Neq	<i>Aquifex aeolicus</i>	aae
<i>Phaeodactylum tricornutum</i>	Pti	<i>Nitrosopumilus maritimus</i>	Nmr	<i>Meiothermus ruber</i>	mrb
<i>Trichomonas vaginalis</i>	Tva	<i>Acidilobus saccharovorans</i>	Asc	<i>Herpetosiphon aurantiacus</i>	hau
<i>Paramecium tetraurelia</i>	Ptm	<i>Thermofilum pendens</i>	Tpe	<i>Chlorobium limicola</i>	cli
<i>Tetrahymena thermophila</i>	Tet	<i>Pyrobaculum neutrophilum</i>	Tne	<i>Chlorobium tepidum</i>	cte
<i>Dictyostelium discoideum</i>	Ddi	<i>Caldivirga maquilingensis</i>	Cma	<i>Anabaena sp</i>	anb
<i>Naegleria gruberi</i>	Ngr	<i>Pyrobaculum aerophilum</i>	Pai	<i>Synechocystis sp. PCC 6803</i>	syn
<i>Monosiga brevicollis</i>	Mbr	<i>Metallosphaera sedula</i>	Mse	<i>Aminobacterium colombiense</i>	aco
<i>Aspergillus oryzae</i>	Aor	<i>Sulfolobus solfataricus</i>	Sso	<i>Elusimicrobium minutum</i>	emi
<i>Saccharomyces cerevisiae</i>	See	<i>Hyperthermus butylicus</i>	Hbu	<i>Planctopirus limnophila</i>	plm
<i>Cyanidioschyzon merolae</i>	Cme	<i>Thermosphaera aggregans</i>	Tag	<i>Gemmatimonas aurantiaca</i>	gau
<i>Ostreococcus lucimarinus</i>	Olu	<i>Desulfurococcus kamchatkensis</i>	Dka	<i>Coraliomargarita akajimensis</i>	caa
<i>Chlamydomonas reinhardtii</i>	Cre	<i>Staphylothermus marinus</i>	Smr	<i>Fusobacterium nucleatum subsp. nucleatum ATCC 25586</i>	fnu
<i>Physcomitrella patens patens</i>	Ppp	<i>Aeropyrum pernix</i>	Ape	<i>Fibrobacter succinogenes</i>	fsu
<i>Zea mays</i>	Zma	<i>Thermococcus kodakarensis</i>	Tko	<i>Cytophaga hutchinsonii</i>	chu
<i>Oryza sativa</i>	Osa	<i>Methanothermobacter thermautotrophicus</i>	Mth	<i>Porphyromonas gingivalis W83</i>	pgi
<i>Populus trichocarpa</i>	Pop	<i>Methanocella paludicola</i>	Mpd	<i>Candidatus Solibacter usitatus</i>	sus
<i>Arabidopsis thaliana</i>	Ath	<i>Candidatus Methanosphaerula palustris</i>	Mpl	<i>Spirochaeta smaragdinae</i>	ssm
<i>Trichoplax adhaerens</i>	Tad	<i>Methanoregula boonei</i>	Mbn	<i>Mycobacterium marinum</i>	mmi
<i>Nematostella vectensis</i>	Nve	<i>Methanocorpusculum labreanum</i>	Mla	<i>Mycobacterium smegmatis MC2155</i>	msm
<i>Schistosoma mansoni</i>	Smm	<i>Methanosaeta thermophila</i>	Mtp	<i>Clostridium acetobutylicum ATCC 824</i>	cac
<i>Brugia malayi</i>	Bmy	<i>Methanohalobium evestigatum</i>	Mev	<i>Bacillus subtilis subsp. subtilis 168</i>	bsu
<i>Caenorhabditis elegans</i>	Cel	<i>Methanohalophilus mahii</i>	Mma	<i>Caulobacter crescentus CB15</i>	ccr
<i>Acyrtosiphon pisum</i>	Api	<i>Methanosarcina acetivorans</i>	Mac	<i>Geobacter sp. M21</i>	gem
<i>Drosophila melanogaster</i>	Dme	<i>Methanococcus aeolicus</i>	Mae	<i>Wolinella succinogenes</i>	wsu
<i>Strongylocentrotus purpuratus</i>	Spu	<i>Methanococcus maripaludis</i>	Mmp	<i>Chromobacterium violaceum</i>	cvi
<i>Ciona intestinalis</i>	Cin	<i>Methanocaldococcus jannaschii</i>	Mja	<i>Marinomonas sp. MWYL1</i>	mmw
<i>Branchiostoma floridae</i>	Bfo	<i>Methanoculleus marisnigri</i>	Mem	<i>Magnetococcus marinus</i>	mgm
<i>Danio rerio</i>	Dre	<i>Methanococcoides burtonii</i>	Mbu	<i>Escherichia coli K-12 MG1655</i>	eco
<i>Mus musculus</i>	Mmu				

Tabla 3. 1: Lista de organismos candidatos usados en la primera parte de la metodología

## CAPÍTULO 2. METODOLOGÍA

### **2.2 Segunda Metodología: Construcción de redes (SSN) con la similitud de secuencias reportadas en la base de datos KEGG (parálogos de la RBNP de *E.coli*)**

Esta segunda metodología propuesta se realizó con el fin de llevar a cabo la idea inicial de la presente investigación, que consistió en explorar las redes de similitud de secuencias como herramienta para el estudio de la evolución de las enzimas de la RBNP, lo que no fue posible con la metodología anterior, cuyos resultados fueron escasos para la construcción de redes verdaderas, por lo que se tuvieron que procesar de diferente manera que los obtenidos con esta última metodología.

Con esta segunda metodología se buscó encontrar una amplia muestra de “parálogos”, que nos permitieran por un lado visualizar la diversidad de estos homólogos en cuanto a las rutas a las que pertenecen, y su cercanía con la RBNP, y por otro que nos permitieran construir, a través de un método muy sencillo, diferentes redes de similitud de secuencias utilizando el genoma de un solo organismo modelo: *Escherichia coli* (sepa O6:K2:H1 CFT073 (UPEC)) y las enzimas de la RBNP de este organismo como referencia de comparación (Ver tabla 3.13)

El procedimiento que se realizó con la segunda metodología se describe a continuación (figura 13). Para poder construir una red de similitud de secuencias para cada una de las enzimas de la RBNP primero se buscaron los parálogos con mayor similitud (best hits 30% de similitud aprox.) de cada una de las enzimas de la RBNP de *E.coli* en la base de datos del KEGG (versión 2011) a los que se les identificó como un 1er nivel de parálogos, posteriormente, en la misma base de datos, se buscaron las enzimas parálogas de cada enzima encontrada en este primer nivel de homología para formar el segundo nivel de homólogos, que a su vez fueron utilizadas para realizar una última búsqueda de homólogos, formando así un tercer nivel. Para organizar esta información, primero se construyeron listas en Word de cada enzima de la ruta con sus respectivos parálogos de primero, segundo y tercer nivel, se pasó la información a tablas de Excel, y se definieron los siguientes criterios para construir cada red (figura 14) con el programa cytoscape:

- a. Cada enzima de la RBNP es el nodo principal de su respectivo grafo
- b. Las aristas representan: similitud entre secuencias
- c. La red es direccional

Posteriormente, se hizo una búsqueda en la base de datos pfam [46]; para identificar los dominios principales de las enzimas presentes en los grafos construidos, marcando cada uno de ellos con distintos símbolos. (Figura 14) y consultando su clasificación.

Finalmente, los resultados de esta sección (redes grandes) se complementaron haciendo diferentes tablas para mostrar con mayor claridad a las enzimas más relacionadas con cada enzima de la RBNP dentro de cada red.

Abreviatura	Nombre de la enzima de la RBNP de <i>E. coli</i>
PrsA	ribose-phosphate pyrophosphokinase
PurF	Amidophosphoribosyltransferase
PurD	GAR sintasa (phosphoribosylamine-glycine ligase)
PurN	FGAR transformilasa (phosphoribosylglycinamide formyltransferase)
PurL	FGAM sintasa (phosphoribosylformylglycinamide synthase)
PurM	AIR sintasa (phosphoribosylformylglycinamide cyclo-ligase)
PurK	CAIR, PAICS AIR carboxilasa phosphoribosylaminoimidazole carboxylase
PurC	SAICAR sintasa (phosphoribosylaminoimidazole-Succinocarboxamide synthase)
PurB	adenylosuccinate lyase
PurH	Ciclohídroclase (phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase)

Tabla 3. 2.2 Enzimas de la Ruta *de novo* de biosíntesis de purinas de *Escherchia coli* O6:K2:H1 CFT073 (UPEC)

## CAPÍTULO 2. METODOLOGÍA

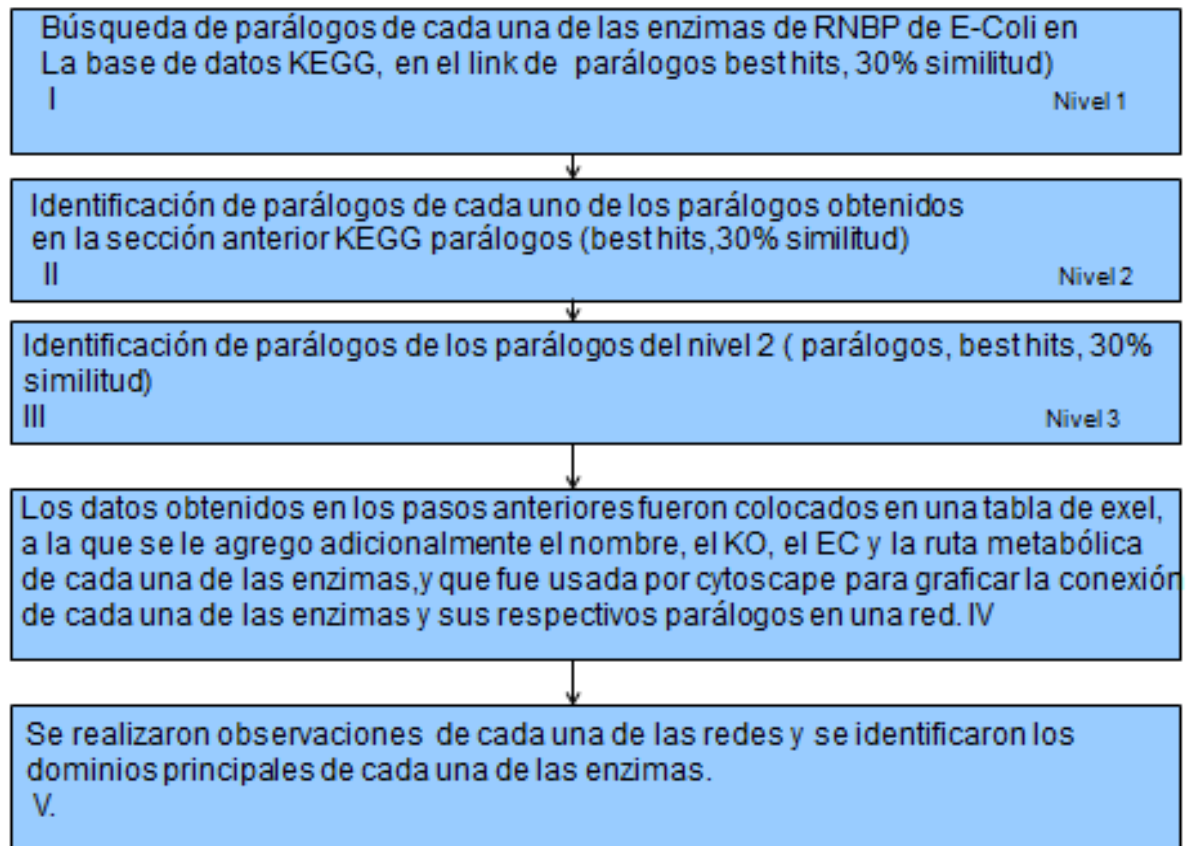


Figura 13 Diagrama de Flujo para la segunda parte de la metodología

## RED DE SIMILITUD DE SECUENCIAS: CARACTERÍSTICAS

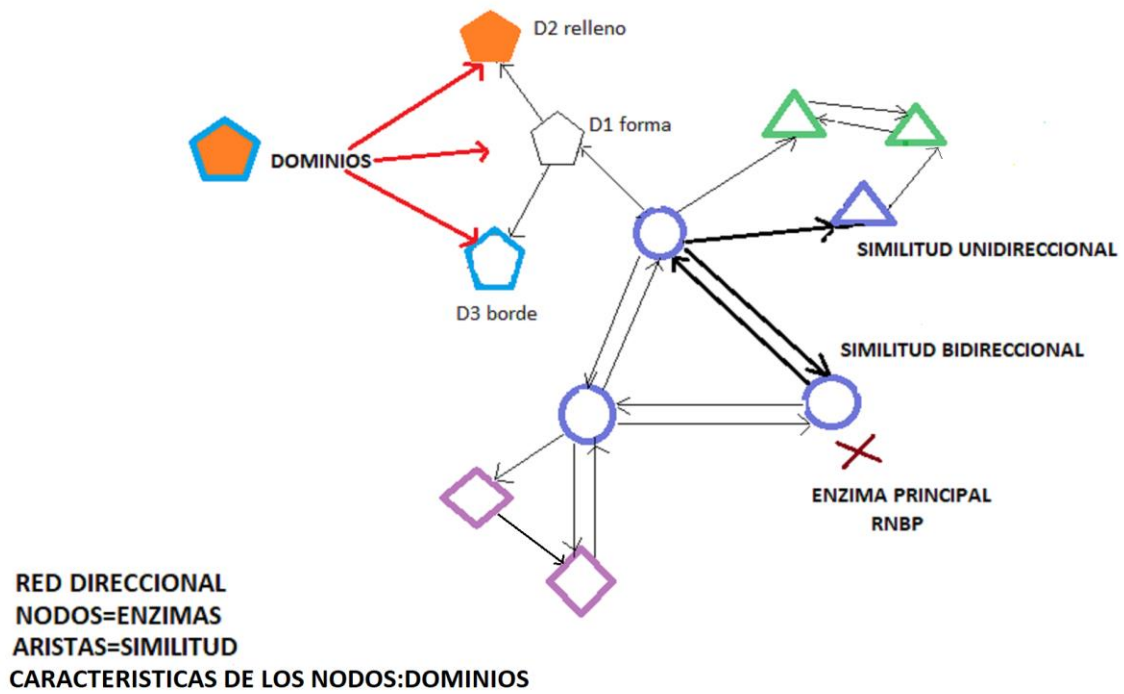


Figura 14: Características para construir los esquemas de las relaciones entre esquemas de los homólogos de cada enzima de la RBNP (Metodología 2)

### 3. Resultados

En este capítulo se presentan los resultados obtenidos para cada enzima involucrada en la RBNP de las dos fases de la investigación descritas previamente en la metodología.

#### 3.1 Resultados de redes chicas

Los resultados principales de esta sección fueron los siguientes:

1.-En general se encontraron muy pocas enzimas homólogas (diferente KO)

de las enzimas de la RBNP, mismas que se dividieron en los siguientes tres grupos:

A) Enzimas sin ningún homólogo: prsA, purH, purB, purC

B) Enzimas con menos de 4 secuencias homólogas: purF, purK, purM, purD, purL

C) Enzimas con más de 4 secuencias homólogas: purN

Y aunque la cantidad de estos homólogos fue variable, en cada caso, se encontraron en un solo organismo o en varios.

2-Se vio que muchos organismos de los considerados en este análisis, en especial los eucariontes, presentaron un número variable de "copias" (Homólogos con el mismo KO) de cada una de las enzimas de la RBNP con los siguientes rangos: En eucariontes de 0 a 5 (con mayor frecuencia dos) En bacterias y arqueobacterias de 0 a 2 (con mayor frecuencia 1 sobretodo en arqueas) siendo PrsA un buen ejemplo de las enzimas en las que se vieron este tipo de duplicaciones.

3.- Se encontraron varias enzimas hipotéticas homólogas a cada una de las enzimas de la RBNP, en diferentes organismos que se tomaron en cuenta. Ejemplo PurB.

4.- Enzimas Fusionadas

Se encontró, como ya se había reportado anteriormente, que algunas de las enzimas de la biosíntesis de purinas se encuentran fusionadas en los eucariontes purD, purM, purN, purC, purK y purL.

### 3.1.1 Ribose-phosphate pyrophosphokinase (PrsA)

La representación de las secuencias homólogas de PrsA en el esquema elaborado con cytoescape se caracteriza por tener un solo macro nodo donde están representadas la PrsA y las copias de PrsA de cada uno de los organismos considerados. En dicho esquema no se muestra ningún ortólogo o parálogo con un KO y/o EC diferente de esta enzima. En las bacterias hay de una a dos de este tipo de copias por organismo, en eucariontes hay de 2 a 7 y en arqueas solo hay un representante de PrsA para cada organismo excepto en *N.equitans*, lo que resulta esperado por ser un organismo endoparásito. Los eucariontes también se distinguen por tener varias enzimas hipotéticas de este tipo, inclusive hay organismos en los que todos los parálogos son proteínas hipotéticas (12 casos).





## CAPÍTULO 3. RESULTADOS

Nombre	KO	EC	Organismos	Ruta metabólica
PrsA	K00948	(EC:2.7.6.1)	Bacterias (1 y 2 ortólogos de PrsA), Arqueas (0 y 1 ortólogos de PrsA), Eucariontes (4 o más ortólogos de PrsA)20 Eucariontes con ortólogos hipotéticos de PrsA (1, 2 y 3)	Purines, Pentose phosphate path, Metabolic paths, Biosynthesis secondary metabolites, microbial metabolism

Tabla 3. 3 Resultados de la red chica de las enzimas homólogas a PrsA de diferentes organismos.

### 3.1.2 Amidophosphoribosyltransferase(PurF)

En cuanto a los resultados de PurF se puede observar que el homólogo más importante de purF es de la *glucosamina fructuosa-6-fosfato aminotransferasa (glms1)*. Existe una homología entre las enzimas PurF y glms que se puede ver más claramente en las bacterias. En el primer nodo correspondiente a PurF se encontró que: Las arqueas tienen de 0 a 2 copias de purF. De 29 organismos 7 arqueas no tienen dicha enzima, 7 arqueas tienen dos copias de purF y 15 arqueas tienen una sola purF. En cuanto a las bacterias podemos ver que todas tienen una purF excepto dos organismos que tienen 2 copias. La mayor parte de los eucariontes tiene una sola purF, aunque hay 3 excepciones, una de ellas tiene 3 copias de purF y las otras dos, 2 copias. Por otro lado hay alrededor de seis o siete eucariontes que no presentan purF o tienen proteínas hipotéticas.

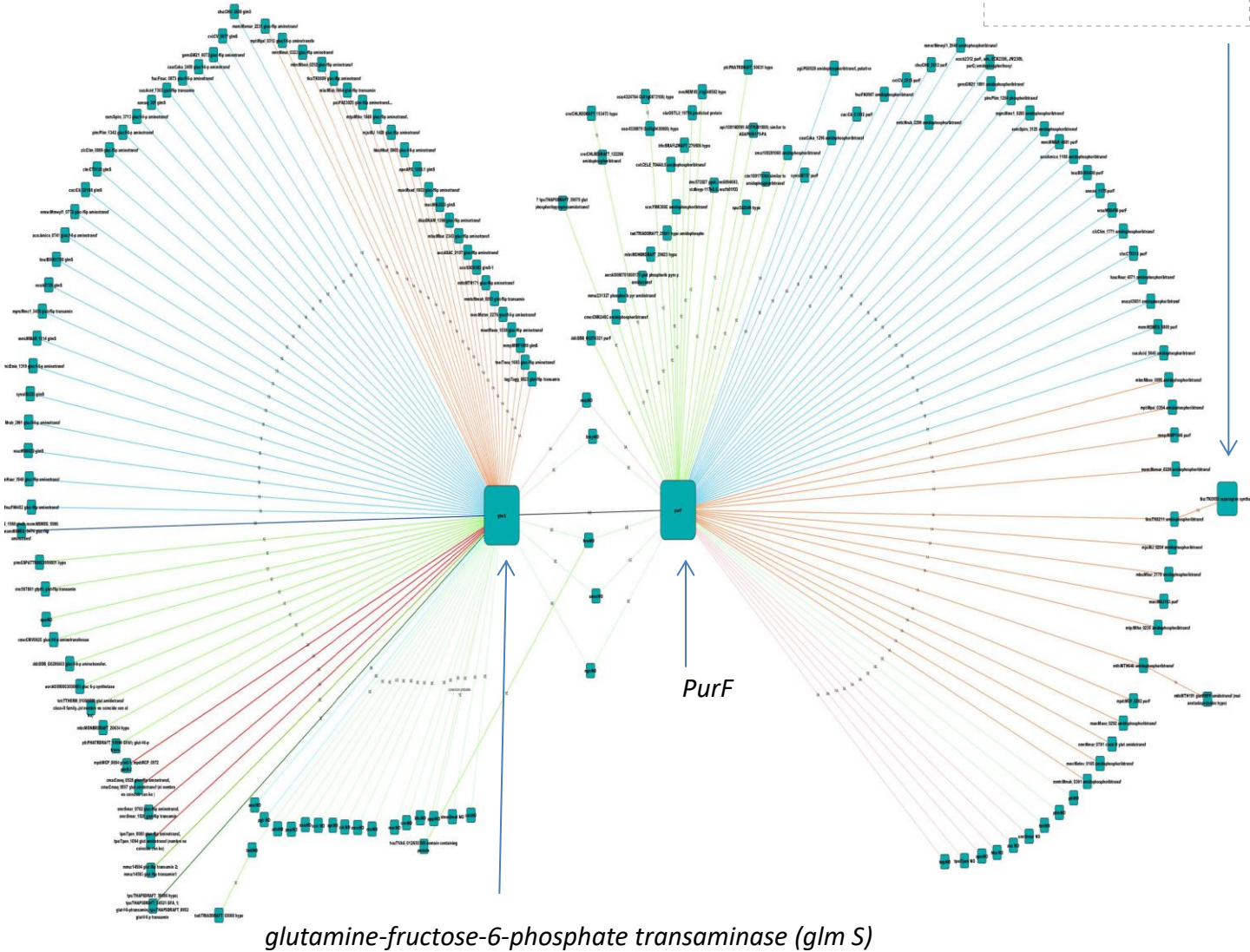
El segundo nodo importante de la red chica de purinas corresponde a la enzima *glms*. La mayoría de las arqueas tuvieron una copia de esta enzima, habiendo solo 4 con 2 y faltando únicamente en *N.equitans*. En cuanto al número de copias de *glas* en las bacterias, se vio que todas tuvieron una, excepto 3 organismos, uno de los cuales tuvo tres copias y los otros dos ninguna.

### *CAPÍTULO 3. RESULTADOS*

En cuanto a la presencia de glms en los eucariontes, se observó que la mayoría no la tuvieron, con excepción de 7 casos que presentaron una sola copia. Por último, las arqueas fueron las únicas en las que en ningún caso carecen de la enzima glms, a diferencia de las bacterias en las que purF es la que no está ausente. Por último, en varios eucariontes, 16 casos, falta solo la glms (la mayoría en plantas y animales, y además en un hongo y un parabasálido) aunque no la purF, que está ausente sólo en 3 protistas quienes sí presentan cada uno una copia de glms.

El resto de las enzimas homólogas a purF son enzimas que no comparten todos los organismos, sino que están presentes en uno solo. La mayoría de ellas son proteínas hipotéticas excepto una, que es la asparagine synthase (AsnB) Esta enzima ya se había reportado anteriormente como homóloga de purF

Asparagine synthase



Red 2 chica de enzimas homólogas de purF

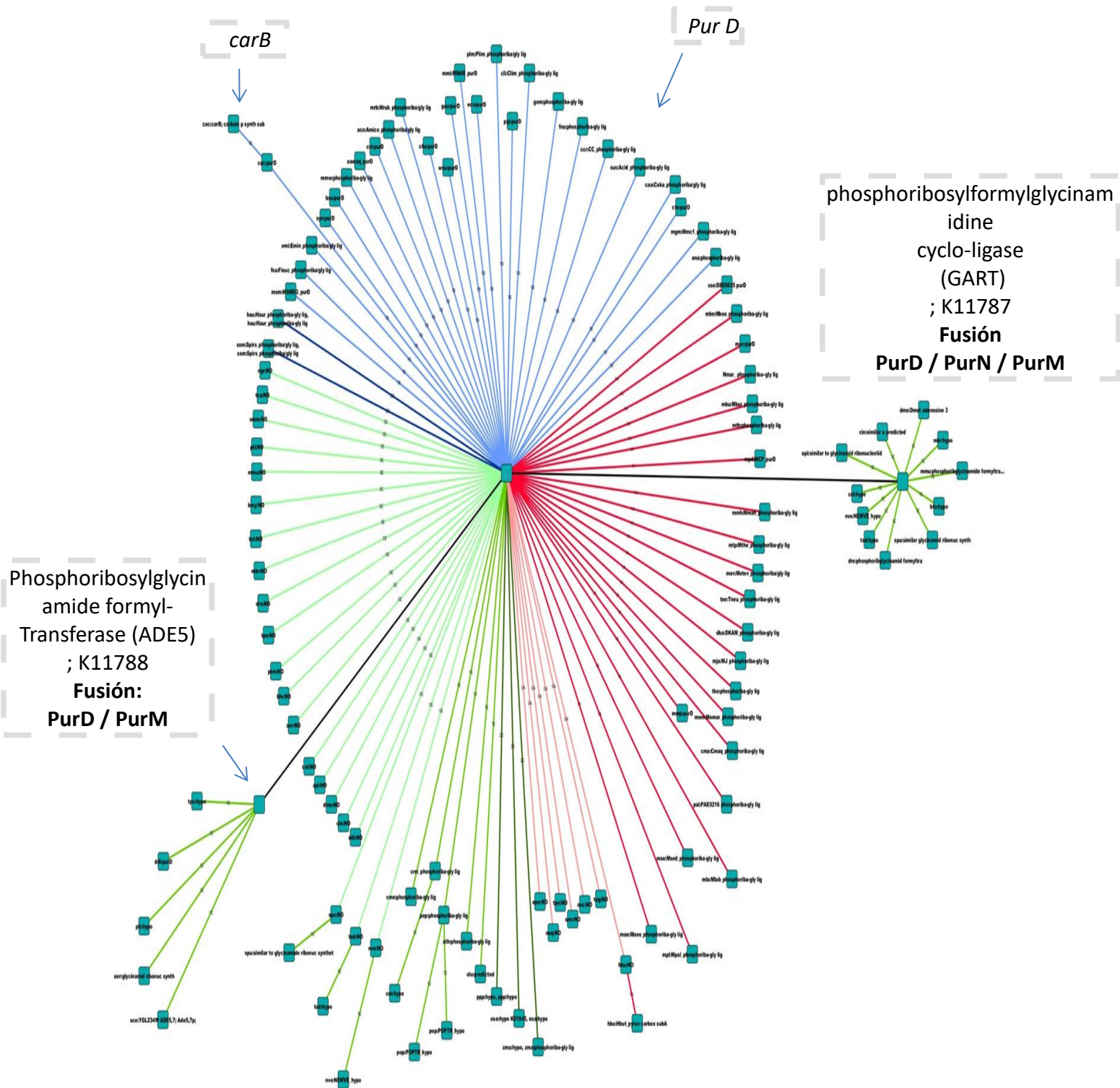
Nombre	KO	EC	Organismos	Ruta metabólica
ecc:c2856 purF; amidophosphoribtransf	K00764	(EC:2.4.2.14)	Bacterias (1 y 2 c), Arqueas (1 c), Crenarchaeota (0 y 2 c), Eucariontes (0, 1, 2 y 3 c)	purine, alanine, aspartate and glutamate metabolism, Metabolic path, biosynthesis secondary metabolites
glmS-1; glucosamine-fructose-6-phosphate aminotranstransferase	K00820	EC:2.6.1.16)	Bacterias (0, 1 y 3 c), Arqueas (1 y 2 c), Eucariontes (0, 1, 2 y 3 c)	Amino sugar and nucleotide sugar metabolism, alanine, aspartate and glutamate metabolism, metabolic path, biosynthesis secondary metabolites
tko:TK0950 asparagine synthase (glutamine-hydrolyzing)	K0195	3[EC:6.3.5.4]	Euryarchaeota	Alanine, aspartate and glutamate metabolism, Metabolic path, Biosynthesis secondary metabolites
Hipotéticas	NO	NO	7 Eucariontes	

Tabla 3. 4 Resultados de la red chica de las enzimas homólogas a PurF de diferentes organismos.

### 3.1.3 Phosphoribosylamine–glycine ligase (PurD)

Esta red está compuesta principalmente de 3 macronodos: el que representa a purD (K01945) y otros 2; además de un cuarto nodo individual. Estos nodos son: 1.- La fósforibosylglycinamida formiltransferasa (ADE5) K11788, EC:[6.3.4.13,6.3.3.1]; (purD fusionada con purM) 2.- La fósforibosylformylglycinamidin ciclo-ligasa (GART) K11787, EC:[6.3.4.13, 2.1.2.2, 6.3.3.1] (purD fusionada con purN y purM) 3.- La carB, carbamoilfosfatosintasa subunidad larga, K1955, EC: 6.3.5.5 En esta red hay otros nodos independientes que corresponden a proteínas hipotéticas o sin ko, que pertenecen a organismos eucariontes, en su mayoría animales y plantas, en los que no está presente la purD normal. Observando cada macronodo correspondiente a los eucariontes de modo individual se ve que el de la phsphoribosylformylglycinamide cyclo-ligase [EC:6.3.4.13 2.1.2.2 6.3.3.1]; K11787 (purD, purN y purM fusionadas) solo la presentan animales y un protista (11 organismos en total). El segundo macronodo de eucariontes que es el de la phosphoribosylglycinamideformiltransferase [EC:6.3.4.13, 6.3.3.1] , K1788, (purD y purM fusionadas) solo tiene 5 componentes de los cuales uno es incierto. En este grupo sólo se observan hongos y protistas.

En cuanto a la enzima purD individual (K01945), se observó en todas las bacterias, en la mayor parte de las arqueas, faltando solo en 6 crenarqueotas y una nanoarqueota. En los eucariontes, sólo la presentan algunas plantas (8 organismos): algas y dicotiledoneas en una copia aunque tres de estos organismos tienen además proteínas hipotéticas y monocotiledoneas y musgos, en 2 copias. Por otro lado, hay un pequeño grupo de 3 animales con enzimas hipotéticas, de las cuales en ortólogos una tiene el K01945 y 2 se buscaron en el listado del Kegg y aparentemente son cercanas al K11787. Por último se vio que hay 6 organismos eucariontes, que además de no presentar purD tampoco tuvieron algún homólogo de ésta y entre ellos, podemos mencionar los siguinetes: *Schistosoma mansoni*, Flat worms; *Brugia malay*, etc.



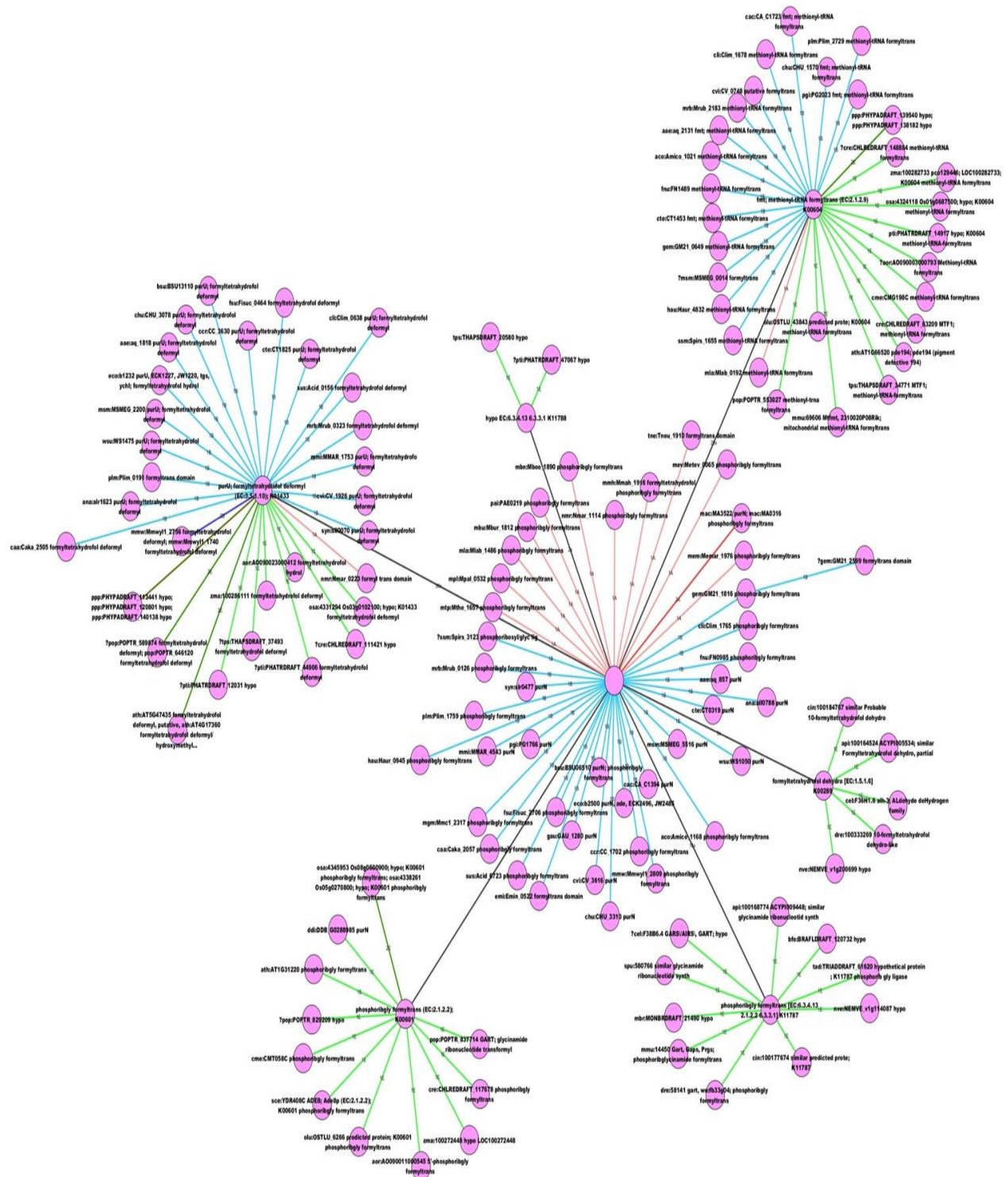
Nombre	KO	EC	Organismos	Ruta metabólica
ecc:c4963 purD; phosphoribosylamine-glycine ligase	K01945	EC:6.3.4.13	Bacterias (1 y 2 c), Arqueas (0 y 1 c),Eucariontes:	
phosphoribosylglycinamide formyl-transferase	K11788	[EC:6.3.4.13 6.3.3.1]	2 Fungi, 3 Protists	
phosphoribosylformylglycinamide cyclo-ligase	K11787	[EC:6.3.4.13 2.1.2.2 6.3.3.1]	10 Animals, 1 Protists	
cac:CA C2644 carB; carbamoyl phosphate synthase large subunit	K01955	[EC:6.3.5.5]	Firmicutes	
hbu:Hbut 1157 pyruvate carboxylase subunit A	NO en kegg ort:k0196 en kegg para- log es k01955 o k01905	(EC:6.3.4.14)	Crenarchaeota	
Hipotéticas	NO	NO	7Eucariontes	

Tabla 3. 5: Resultados de la red chica de enzimas homólogas a PurD de diferentes organismos



### **3.1.4 Phosphoribosylglycinamide formyltransferase 1 (PurN)**

La enzima PurN se encontró presente en todas las bacterias consideradas, (cada una con un solo ortólogo) mientras que no se observó en ningún eucariote. En las arqueas purN se encontró ausente en varias, 17 de estas, especialmente en las crenarqueotas (10 de 12), encontrándose únicamente en 9 de 15 Euryarchaeotas y en una tamoarqueota. Al igual que en las bacterias, las arqueas tuvieron una sola copia de purN excepto una que tuvo 2. Esta enzima, presentó varias enzimas homólogas (Tabla 3.6) algunas de las cuales fueron duplicaciones, como las enzimas purN K00601, fmt, purU; y otras, fusiones de purN con otras enzimas como la enzima GART.



Red 3 chica de enzimas homólogas de purN

Nombre	KO	EC	Organismos	Ruta metabólica
cc:c3018 purN; phosphoribosylglycinamide formyltransferase	K11175	EC:2.1.2.2	Bacterias (0 y 1c [28 org]), Arqueas: (0[17 org],1 [13 org] y 2 c), Eucariontes (0 c [30 org])	Purines One carbon pool by folate, Metabolic path, Biosynth sec
purU; formyltetrahydrofolate deformylase	K01433	EC:3.5.1.10	19 Bacterias (1 y 2 c), 1 Arqueas: Thaumarchaeota (1 c), Eucariontes (1[plantas, protist,	Glyoxylate and dicarboxylate metabolism, One carbon pool by folate
Fmt;methionyl-tRNA formyl transferase	K00604	EC:2.1.2.9	15 Bacterias (1 c), 1 Arqueas: Euryarchaeota (1 c), 13 Eucariontes (1 y 2 c [9 plantas y	One carbon pool by folate
Aminoacyl-tRNA biosynthesis phosphoribosylglycinamide formyltransferase	K00601	EC:2.1.2.2	11 Eucariontes (1 c [mayoria plantas])	Purines, One carbon pool by folate, Metabolic
Phosphoribosylglycinamide Formyltransferase (GART)	K11787	EC:6.3.4.13 2.1.2.2 6.3.3.1	10 Eucariones (9[animales], 1[protists] c)	Purines, One carbon pool by folate, Metabolic
formyltetrahydrofolate dehydrogenase	K00289	EC:1.5.1.6	5 Eucariontes (1 c [animales])	One carbon pool by folate
M211816 phosphoribosylglycimideformyl tranferase	K11175 Ort	EC:2.1.2.2	1 Bacteria: Deltaproteobacteria (1 c)	Purines, One carbon pool by folate, Metabolic
M212599 formyl transferase domain protein	NO, pero en kegg ort es K00607	NO, pero en kegg ort es (2.1.2.-)	1 Bacteria: Deltaproteobacteria (1 c)	No Pathway
Hypothetical protein	K11788	EC:6.3.4.13 6.3.3.1	2 Eucariontes:protistas (1 c)	No Pathway
Proteínas Hipotéticas			9 Eucariontes, 3 bacterias	
Organismos que no presentan PurN ni homólogos			7 Eucariontes ( <i>Brugia malayi</i> , <i>Drosophila melanogaster</i> (fruit fly), <i>Naegleria gruberi</i> , <i>Paramecium tetraurelia</i> , <i>Schistosoma mansoni</i> , <i>Tetrahymena thermophila</i> , <i>Trichomonas vaginalis</i> )	

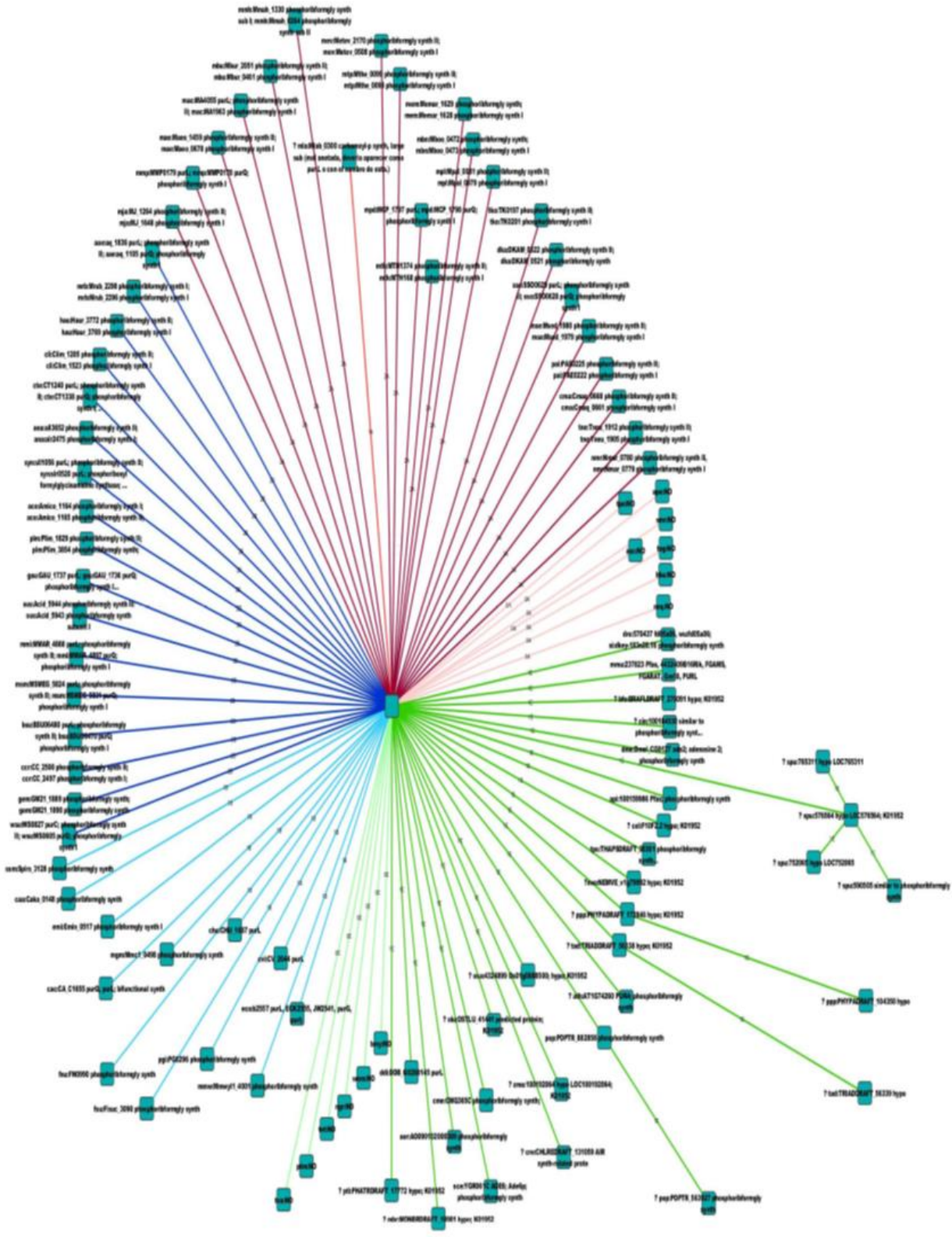
Tabla 3. 6 Resultados de la red chica de las enzimas homólogas a PurN de diferentes organismos. En cada nodo están incluidas las enzimas que no tuvieron KO pero que les asigne el KO que había en la lista de ortólogos, algo importante de notar es que purT no aparece en esta red

### 3.1.5 Phosphoribosylformylglycinamide synthase (purL)

La red chica de la enzima purL (ecc:c3080); phosphoribosylformylglycinamide synthase, (EC:6.3.5.3 , K01952) está compuesta de un macronodo (purL) además de un número muy reducido de otros nodos (6) que corresponden a proteínas hipotéticas individuales, 3 que son de un solo organismo y otras 3 de diferentes.

En esta red se observó que purL es una enzima que está distribuida en los 3 dominios, solo que mientras que está presente en todas las bacterias, falta en algunas arqueas y la mayoría de las proteínas que tienen los eucariontes no son purL sino enzimas hipotéticas, con el mismo KO y EC que los de purL. Entre los eucariontes hubo organismos que, además de tener PurL o alguna enzima equivalente con el mismo KO y EC, presentaron además enzimas hipotéticas que no tienen ni KO ni EC y estos organismos fueron: *Strongylocentrotus purpuratus* (purple sea urchin), que tuvo 3 de estas proteínas hipotéticas; *Trichoplax adhaerens* (Placozoario) que tuvo una; *Populus trichocarpa* (black cottonwood) que tuvo una y *Physcomitrella patens subsp. Patens* ( musgo) que también tuvo una de estas enzimas.

La enzima purL también se caracterizó por que en varios organismos, tanto bacterias como Arqueas, se encontró junto con la enzima purQ. En el grupo de las Arqueas se vieron 7 organismos que no presentaron ninguna de estas enzimas, 6 de estos fueron Crenarchaeotas y 1 Nanoarchaeota, habiendo una sola Arquea (*Methanococcus labreanum*) en la que se encontró purL sin encontrarse la enzima purQ, mientras que el resto de las Arqueas (6 Crenarchaeotas, 1 Thaumarchaeota y 14 Euryarchaeotas), presentaron ambas enzimas. En el grupo de las Bacterias, hubo, en primer lugar, 12 organismos en los que únicamente se observó la enzima purL, y en segundo lugar, se vieron otros 17 organismos en los que purL se encontró junto con purQ. Estos dos conjuntos de bacterias estuvieron compuestos de organismos de diferentes phyla con excepción de dos, que pertenecieron al mismo phylum (Firmicutes) Con respecto a los eucariontes, hubo 6 organismos (4 protistas y 2 animales) en los que no se observó la enzima PurL, ni la de algún otro homólogo de esta, y 22 en los que se encontró únicamente la presencia de esta misma enzima. Entre estos últimos organismos se observaron animales, hongos, plantas y protistas, siendo la mayoría animales y plantas.



Red 4 chica de enzimas homólogas de purL

Nombre	KO	EC	Organismos	Ruta metabólica
c3080 purL; phospho-ribosylformylglycinamide synthase	K01952	(EC:6.3.5.3)	Arqueas (0 [7 organismos: 6 Crenarchaeota, 1 Nanoarchaeota], 1 y 2[21org]), bacterias (1[12] y 2 c [17]) Eucariontes (1 c [22] y 0 c [6 organismos: 4 protistas y 2 animales])	
POPTR 563927 phosphoribosyl-formylglycinamide synthase	NO, en kegg ort 1953	NO	Eucarionte: <i>Populus trichocarpa</i> (black cottonwood)	NO PATHWAY
proteínas hipotéticas			6 Eucariontes (3 <i>Strongylocentrotus purpuratus</i> (purple sea urchin))	

Tabla 3. 7 Resultados de la red chica de las enzimas homólogas a PurL de diferentes organismos. NOTA: En la base de datos del KEGG del laboratorio la proteína hipotética de *C. elegans* aparece con KO de la purL, pero en la base actual del KEGG 2013 aparece sin KO.

### 3.1.6 Phosphoribosylformylglycinamide cyclo-ligase (purM)

Esta red está compuesta por un nodo grande, dos nodos pequeños y tres nodos independientes.

El nodo grande representa a la enzima PurM normal, con un K01933 y un (EC:6.3.3.1), donde se observa que los que presentan más esta enzima son, en primer lugar las bacterias y en segundo las arqueas. Todas las bacterias tienen una sola copia de esta enzima, excepto dos organismos: *Porphyromonas gingivalis* W83 y *Coralimargarita akajimensis*. Mientras que las arqueas presentan una sola copia excepto 7 organismos: 6 Crenarchaeotas, que son: *Aeropyrum pernix*, *Acidilobus saccharovorans*, *Hyperthermus butylicus*, *Staphylothermus marinus*, *Thermosphaera aggregans*, *Thermofilum pendens* y 1 Nanoarchaeota: *Nanoarchaeum equitans*.

Por último, los eucariontes casi no presentan la enzima purM en su forma más simple, es decir sin tener alguna fusión con otra enzima, a excepción de 7 organismos fotosintéticos: 3 algas (*Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*), 1 musgo (*Physcomitrella patens* subsp. *Patens*), 2 dicotiledoneas (*Populus trichocarpa* (black cottonwood), *Arabidopsis thaliana* (thale cress) y 1 monocotiledonea (*Oryza sativa japonica* (Japanese rice). 4 de estos eucariontes tienen 1 copia y 2 tienen 2 copias.

En cuanto al primer nodo pequeño hay 10 organismos con una enzima de nombre phosphoribosylglycinamide formyltransferase (GART) o similar, con un K11787 y con un número (EC:6.3.4.13 2.1.2.2 6.3.3.1) (fusión de purD, purN y purM) de los cuales 9 son animales y 1 protista.

El nodo más pequeño que tiene un K11788 y un (EC:6.3.4.13 6.3.3.1) (fusión de purD y purM) con una enzima de nombre phosphoribosylamine-glycine ligase (ADE5) o similar, tiene solamente 4 organismos que son: 2 hongos y 2 protistas.

Hay 2 eucariontes: 1 protista (*Phaeodactylum tricornutum*) y un animal (*Trichoplax adhaerens*) que no tienen la purM pero tienen una enzima hipotética en vez de esta. Por último existe un nodo independiente que corresponde a la enzima Tagg 0621 hydrogenase expression formation protein (HypE), con un K04655 en ortólogos y que únicamente lo presenta una Crenarchaeota (*aggregans*) que no tiene purM.





## Red 5 chica de enzimas homólogas de purM

	KO	EC	Organismos	Ruta metabólica
phosphoribosylaminoimidazole methetase	K01933	(EC:6.3.3.1)	Arqueas (0 [7 org: 6 Crenarchaeota, 1 Nanoarchaeota] y 1 c), bacterias (0 [2 org: Verrucomicrobia, Bacteroidetes], 1 c), Eucariontes (0, 1[4 org: plantas], 2 [2 org: plantas] y 3 [1 org: plantas]c)	
Phosphoribosylglycinamide formyltransferase o similar	K11787	(EC:6.3.4.13 2.1.2.2.6.3.3.	10 Eucariontes (9 animales, 1 protista)	Purines, One carbon pool by folate, Metabolic path
phosphoribosylamine-glycine ligase o similar	K11788	(EC:6.3.4.13 6.3.3.1)	4 Eucariontes (2 hongos, 2 protistas)	Purines, Metabolic path, Biosynthesis metabolites
hydrogenase expression/formation protein HypE;	K04655	NO	1 Crenarchaeota: <i>Thermosphaera aggregans</i>	NO PATHWAY
Proteínas Hipotéticas			2 Eucariontes: <i>Trichoplax adhaerens</i> , <i>Phaeodactylum tricornutum</i>	NO PATHWAY

Tabla 3. 8 Resultados de la red chica de las enzimas homólogas a PurM de diferentes organismos

## CAPÍTULO 3. RESULTADOS

### 3.1.7 5-(carboxyamino) imidazole ribonucleotide synthase (purK)

Esta red está compuesta de 3 nodos principales. El primer nodo corresponde al K01589 donde se encuentran representados la purK (EC:6.3.4.18) y probablemente a la purE clase II (EC:4.1.1.21) u otra donde hay una fusión entre purk y purE que únicamente está presente en los procariontes.

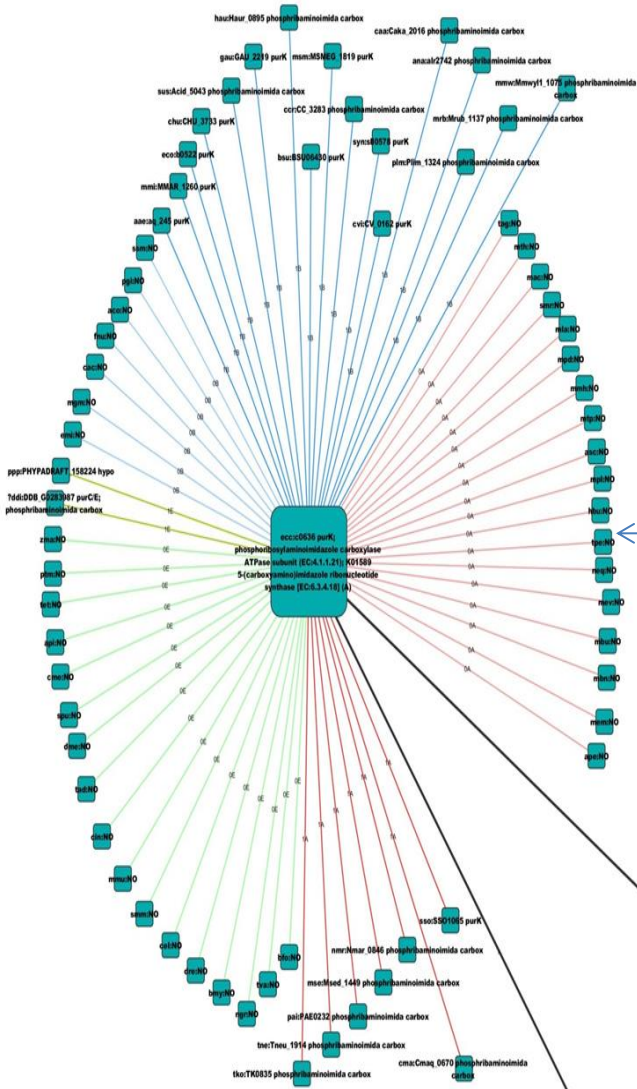
De la purK se encontraron: 2 Arqueas con 1 copia y curiosamente solo se encontraron 4 bacterias con 1 copia. De eucariontes no se encontraron ninguno.

De la segunda enzima: Curiosamente se encontraron 12 bacterias que presentaron 1 copia y 5 arqueas con 1 copia. No se encontró ningún eucarionte. En ambos casos no se encontró ningún patrón con respecto al tipo de organismos.

El segundo nodo corresponde a la enzima purT que tiene el K08289 y un EC:2.1.2.2, en el cual hay 12 bacterias y 6 arqueas, todos con una copia. No se observa ningún patrón en cuanto al tipo de organismos que contienen esta enzima.

El tercer nodo corresponde a la enzima que tiene el K11808 y el EC:4.1.1.21 que probablemente es la enzima PurK-PurE1. Este nodo tiene en total 11 organismos, todos ellos eucariontes, entre los cuales solo hay plantas, hongos y protistas. Todas tienen 1 copia con excepción de una que tiene dos copias y que es una planta (*Populus trichocarpa* (black cottonwood)).

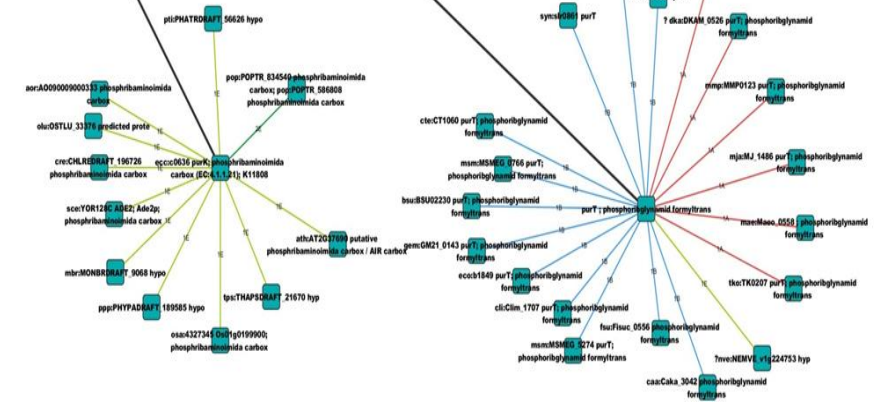
Finalmente solo se encontraron 3 enzimas hipotéticas sin KO pertenecientes a los siguientes eucariontes: una planta: musgo (*Physcomitrella patens* subsp. *Patens*), un protista: Amoebozoa (*Dictyostelium discoideum* (cellular slime mold)) y un animal: Cnidarians (*Nematostella vectensis* (seaanemone)).



PurK

phosphoribosylglycinamide formyltransferase 2  
purT

PurE; clase II  
phosphoribosylaminoimidazolecarboxylase



Red 6 chica de enzimas homólogas de purK

Nombre	KO	EC	Organismos	Ruta metabólica
ecc:c0636 purK; phosphoribosylaminoimidazole carboxylase ATPase subunit (EC:4.1.1.21); 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18] (A)	K01589		En General no más de 1 copia. Bacterias: 17 con, 7 sin. Arqueas: 7 con, 18 sin. Eucariontes: 2 hypo, 16 sin.	Purines, Metabolic path, Biosynth sec Metabolites
		EC:4.1.1.21	Arqueas: 4 Crenarchaeota, 1 Euryarchaeota (la única: <i>Thermococcus kodakaraensis</i> ). Bacterias: 13	
		EC:6.3.4.18	Arqueas: 1 Crenarchaeota y 1 Thaumarchaeota. Bacterias: 1 Actinobacteria, 1 Alphaproteobacteria, 1 Chloroflexi, 1 Verrucomicrobia. 1 Gammaproteobacteria (en ortólogos).	
purT phosphoribosylglycinamide formyltransferase 2	K08289	(EC:2.1.2.2)	Arqueas 6 [5 tipo coccus"] (1 c), Bacterias 12 (1 c), Eucariontes 1(hypo), Cnidarians: <i>Nematostella vectensis</i> (sea anemone) (1c)	Purines, Metabolic path, One carbon pool by folate, Biosynth sec metabolites
ecc:c0636 purK; phosphoribosylaminoimidazole carboxylase ATPase subunit (EC:4.1.1.21);	K11808	(EC:4.1.1.21)	Eucariontes 11 plantas, hongos y protistas (1c, 2c [1 org] )	Purines, Metabolic path, Biosynth sec metabolites
Proteínas Hipotéticas sin KO			3 Eucariontes: 1 Planta; musgo: <i>Physcomitrella patens</i> subsp. <i>Patens</i> (ortólogos ko1589), 1 Protists; Amoebozoa: <i>Dictyostelium discoideum</i> (cellular slime mold) y 1 Animal; Cnidarians: <i>Nematostella vectensis</i> (sea anemone).	

Tabla 3. 9 Resultados de la red chica de las enzimas homólogas a PurK de diferentes organismos.

## CAPÍTULO 3. RESULTADOS

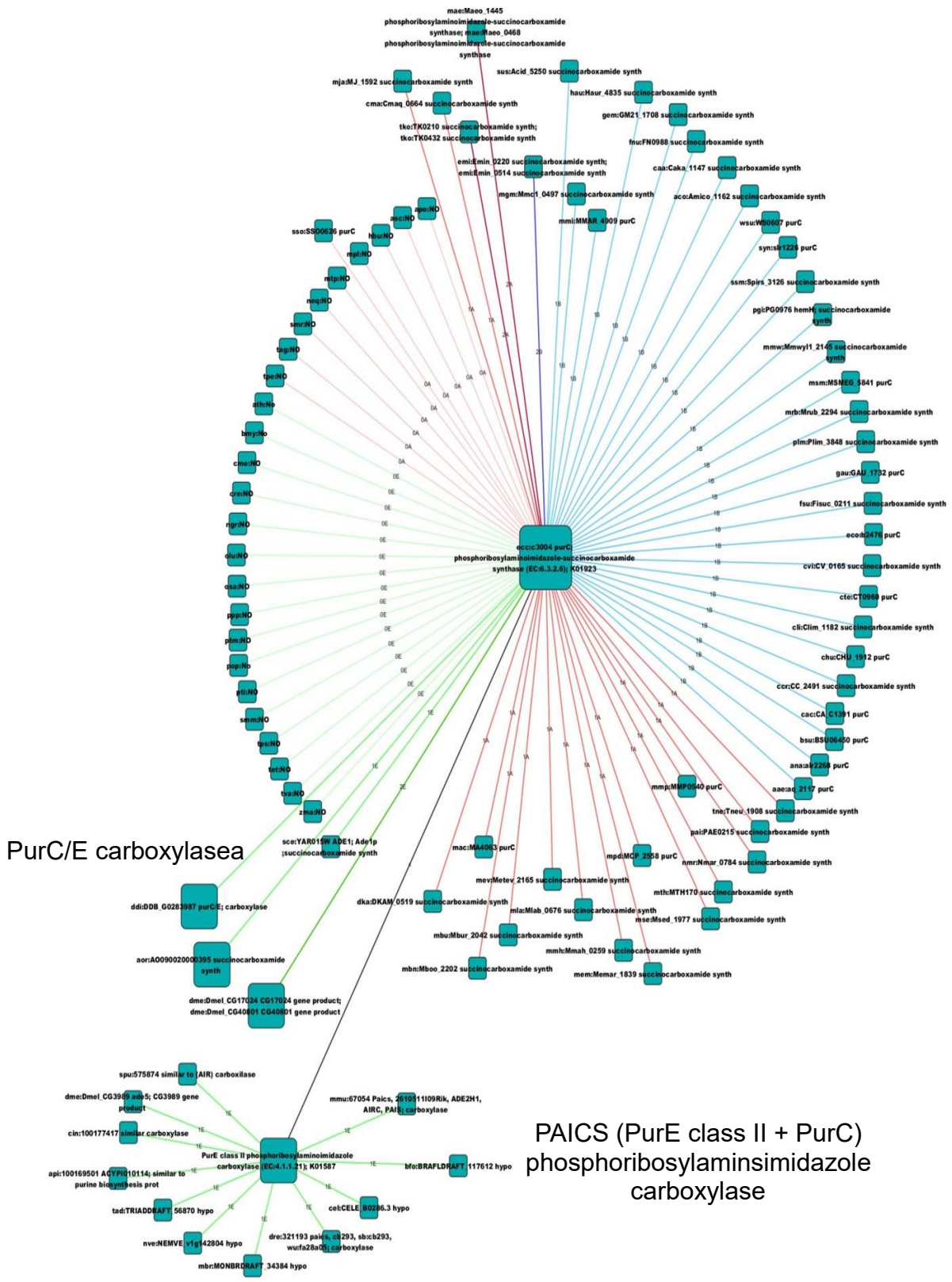
### 3.1.8 Phosphoribosylaminoimidazole-succinocarboxamide synthase (purC)

La red chica de purC consta de un nodo muy grande que corresponde a la misma enzima, un nodo chico y tres independientes. En el nodo grande correspondiente a la phosphoribosylaminoimidazole-succinocarboxamide synthase con un EC:6.3.2.6 y un K01923, se observa que esta enzima se encuentra en primer lugar más representada en bacterias, las cuales presentan todas una copia de dicha enzima con excepción de *Elusimicrobium minutum* (Elusimicrobia) que presenta 2 copias. En segundo lugar se encuentra representada por las arqueas donde 17 organismos presentan 1 copia, la mayoría son Euryarchaeotas (11 organismos), 5 Crenarchaeota y 1 Thaumarchaeota. Dos Euryarchaeotas presentan 2 copias (*Thermococcus kodakaraensis* y *Methanococcus aeolicus*). Diez organismos presentan 0 copias, siete Crenarchaeotas, dos Euryarchaeotas y una Nanoarchaeota.

Por último en los eucariontes casi no se encuentra representada esta enzima, como se esperaba, ya que purC es una enzima que no se presenta en este grupo. A excepción de un organismo, la *Saccharomyces cerevisiae* (budding yeast), donde se encontró como sce:YAR015W ADE1; Ade1p; phosphoribosylaminoimidazole-succinocarboxamide synthetase.

El nodo pequeño corresponde a la enzima PurE class II phosphoribosylaminoimidazole carboxylase, con un K01587, un EC:4.1.1.21 y todos tienen 1 copia. Contiene 11 organismos eucariontes en donde todos son animales menos uno que es un protista: Choanoflagellates (*Monosiga brevicollis*).

Además hay 3 nodos independientes que son eucariontes y no tienen KO por lo que probablemente sean enzimas hipotéticas, aunque presentan el nombre de la enzima y son los siguientes organismos: *Aspergillus oryzae*, *Dictyostelium discoideum* (cellular slime mold) y *Drosophila melanogaster* (fruitfly).



PurC/E carboxylase

PAICS (PurE class II + PurC) phosphoribosylaminimidazole carboxylase

Red 7 chica de enzimas homólogas de purC

Nombre	KO	EC	Organismos	Ruta metabólica
phosphoribosylaminoimidazole-succinocarboxamide synthase	K01923	(EC:6.3.2.6)	Arqueas (0[7:Crenarchaeota, 2: Euryarchaeota, 1: Nanoarchaeota]; 1[5: Crenarchaeota, 1: Thaumarchaeota, 11: Euryarchaeota] y 2[Methanococcus a, Thermococcus k] c), bacterias (1[28] y 2 [1: e.minutum] c), Eucariontes 16(0 [8:plantas, 5:protistas, 2:animals 1: parasitids] c), (1[hongo: <i>Saccharomyces cerevisiae</i> ] c)	Purines, Metabolic path, Biosynthesis metabolites
PurE class II phosphoribosylaminoimidazole carboxylase	K01587	(EC:4.1.1.21)	Eucariontes (1[animales y 1 protista] c)	Purines, Metabolic path
DDB G0283987 purC/E; phosphoribosylaminoimidazole carboxylase	NO, en orto K01589	NO	Eucariontes (1: protists[ <i>Dictyostelium</i> d] c)	NO PATHWAY
AO090020000395 phosphoribosyl-aminoimidazole-succinocarboxamide synthase	NO, en orto K06885 (identidad 0.344)	NO	Eucariontes (1 [Fungi: Ascomycetes] c)	NO PATHWAY
Dmel CG17024 CG17024 gene product from transcript CG17024-R; Dmel CG40801 CG40801 gene product from transcript CG40801-RA	NO (no encontrado en Kegg); NO, K01587 En ortólogos	(EC:4.1.1.21 6.3.2.6), NO	Eucariontes (2: Animals: Arthropods] c)	NO PATHWAY

Tabla 3. 10 Resultados de la red chica de las enzimas homólogas a PurC de diferentes organismos

### 3.1.9 Adenylosuccinate lyase (purB)

La red chica de purB consta de un solo nodo grande además de algunos nodos independientes que corresponden a proteínas hipotéticas de las cuales algunas presentan EC y otras no. En el nodo grande se observa que todos los organismos contienen entre 0 y 2 copias de purB, siendo 1 copia la cantidad más frecuente. Las bacterias es el grupo que la presentan con mayor frecuencia, habiendo solamente 6 organismos con 0 copias, 2 micobacterias, 1 fibrobacter, 1 firmicutes, 1 fusobacteria y 1 verrucomicrobia. En las arqueas solamente hay 7 organismos que no presentan purB y de los cuales 6 son Crenarchaeota y 1 Nanoarchaeota. En los eucariontes es en donde menos se presenta la purB (15 organismos contra 15 que si lo presenta). Dentro de los que no la presentan hay un grupo con proteínas hipotéticas (entre 1 y 3 copias) similares a purB, además está constituido principalmente por plantas y protistas. Entre los eucariontes que sí presentan dicha enzima únicamente hay 3 organismos que presentan 2 copias. Mientras que el resto (12 organismos) sin ser de un tipo en específico, solo presentan 1 copia. Proteínas hipotéticas son 4 organismos eucariontes que no tienen KO, uno de los cuales presenta 3 copias (*Zea mays* (maíz)).





### CAPÍTULO 3. RESULTADOS

Nombre	KO	EC	Organismos	Ruta metabólica
ecc:c1510 purB; adenylosuccina- te lyase	Ko1756	(EC:4.3.2.2)	Arqueas (o[6 Crenarchaeota, 1 Nanoarchaeota] y 1 c), bacterias (o[no hay ninguna protobacteria, hay actinobacterias, Verrucomicrobia, firmicutes, fusobacteria, fibrobacteres ] y 1 c), Eucariontes (o, 1 y 2 c)	Purines, Ala- nine, aspartate and glutamate metabolism, Metabolic path, Biosynth sec metabolites
Hipotetical Pro- Tein	Ko1679	(EC:4.2.1.2)	Solo la tiene un Eu- carionte: <i>Paramecium tetraurelia</i>	Citrate cycle (TCA cycle), Metabolic path, Biosynth sec metabolites, Carbon metabo- lism
Proteínas Hipo- Téticas			4 Eucariontes ( <i>Phys- comitrella patens</i> , <i>Chlamydomonas rein- hardtii</i> , <i>Zea mays</i> (maíz) y <i>Monosiga brevicollis</i> ).	Purines, Ala- nine, aspartate and glutamate metabolism, Metabolic path

Tabla 3. 11 Tabla de resultados de la red chica de las enzimas homólogas a PurB de diferentes organismos.

#### 3.2.10 Phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase (purH)

La red chica de PurH consistió en un solo macronodo, la enzima PurH se encontró en todas las bacterias y cada una de estas tienen 1 copia de dicha enzima, menos 2 organismos. En el segundo grupo, en donde existe mayormente la PurH es en los eucariontes faltando solo en 7 de ellos, correspondientes a diferentes tipos, en los que se pudo haber perdido esta enzima. El único caso de los eucariontes que no presentó la enzima purH sino una similar, fue el ratón (*mus. musculus*). Por otro lado, aunque hay algunos eucariontes que tienen 0 copias, la mayoría tiene 1 y nada más 3 organismos tienen 2 copias, mas no se encontró ningún patrón específico de la presencia de PurH en este sentido. Es en el grupo de las arqueas donde purH se encontró en menos organismos,

únicamente en 9 de estos, todos Euryarqueotas metanógenas y todas sin excepción tienen solamente 1 copia.



Nombre	KO	EC	Organismos	Ruta metabólica
ecc:c4964 purH; Bifunctional phosphoribosyl- laminoimidazo- lecarboxamide formyltrans- ferase/IMP cyclohydrolase (EC:3.5.4.10 2.1.2.3);	K00602	(EC:3.5.4.10 2.1.2.3)	Arqueas(0, 1[9 org: Euryarchaeo- ta]), Bacterias (1 y 2[2 org.] c), Eucariontes (0[6 org], 1 y 2[3 org]	Purines, One carbon pool by folate, Me- tabolic path, Biosynth sec Metabolites

Tabla 3. 12 Tabla de resultados de la red chica de las enzimas homólogas a PurH de diferentes organismos

### 3.2 Resultados de las redes grandes

Los resultados comparativos más importantes de las redes grandes fueron, en primer lugar, que en todas se observaron diferentes módulos de enzimas conectadas entre sí (coincidiendo con lo referido por Patricia Babbitt) [50]; a través de aristas bidireccionales y en diferentes grados. Únicamente las enzimas de los módulos correspondientes al último nivel de similitud, estuvieron interconectadas exclusivamente por aristas unidireccionales. Otra característica en común de todas las redes grandes es que las enzimas dentro de cada uno de sus módulos compartieron en general los mismos dominios principales, además de haber tenido en la mayoría de los casos, la misma cifra correspondiente a la primera posición del número enzimático EC, relacionada con el tipo acción catalítica general que realiza cada enzima. Algunas veces, fue la segunda cifra de este mismo número la que fue compartida entre los componentes de un mismo módulo, que tiene que ver con el grupo funcional (como aldehídos, ésteres, etc.,) sobre el que actúan las enzimas.

Clasificación de la Acción catalítica general de las enzimas (6 grandes grupos o clases): Clase 1: OXIDORREDUCTASAS Clase 2: TRANSFERASAS Clase 3: HIDROLASAS Clase 4: LIASAS Clase 5: ISOMERASAS Clase 6: LIGASAS
--

Otras observaciones que se hicieron con respecto a los módulos y las relaciones de similitud entre un módulo y otro fueron:

En primer lugar, que las enzimas dentro de cada uno de estos grupos al presentar comúnmente los mismos dominios principales, pueden compartir desde uno solo hasta 4 o 5 diferentes. Asimismo se encontraron algunas enzimas que comparten más dominios entre sí, que con el resto del mismo grupo. Sin embargo dichos módulos se distinguieron en general por compartir todos una mínima cantidad de información entre sus componentes, que en promedio fue de dos dominios.

Las enzimas dentro de cada cluster, en general presentaron dominios específicos que no comparten con las enzimas de otros módulos, sin embargo hubo ciertos dominios como los AAA, etc. que se encontraron distribuidos ampliamente en diferentes tipos de enzimas que no pertenecen a los mismos módulos, por lo que estos dominios también se ven sobrerrepresentados en casi todas las redes.

Otro dato encontrado en los módulos de cada red, fue que los diferentes dominios que compartieron entre sí las enzimas de cada uno de estos grupos pertenecen a la misma familia o super familia, aunque dichas enzimas pertenezcan a diferentes rutas metabólicas [50][35][51]. Esta misma relación se identificó entre los dominios de cada enzima vista de manera individual, lo cual coincide con los resultados del artículo “*Domain Combinations in Archaeal, Eubacterial and Eukaryotic Proteomes*” de Sarah Teichmann (Teichmann 2001).

Se observó además, que los módulos mencionados presentaron generalmente enzimas de diferentes rutas metabólicas, que a veces están relacionadas, como la glucólisis y la gluconeogénesis, dos rutas de la biosíntesis de carbohidratos; y la biosíntesis de purinas y el metabolismo de pentosas. Asimismo si puede haber algunos módulos en los que haya más de una enzima perteneciente a la misma ruta metabólica, aunque también algunos de ellos pueden presentar enzimas que sean solamente del mismo tipo, por ejemplo de membrana.

Al ser revisados únicamente los módulos más importantes de cada red, resultó que las enzimas conectadas directamente con la enzima principal (de la RBNP) de cada cluster, pertenecieron en su mayor parte, a rutas metabólicas de especial importancia para los seres vivos, diferentes a la de la biosíntesis de purinas, como las de biosíntesis de aminoácidos y nucleótidos y las pertenecientes al metabolismo central.

Por otra parte, en cuanto a las conexiones que se encontraron entre los módulos de las redes de similitud de secuencias, se observó que mientras unos módulos están conectados unidireccionalmente entre sí a través de una sola arista, otros se encuentran conectados bidireccionalmente mediante dos aristas.

Es probable que las conexiones unidireccionales en general no sean importantes o informativas, ya que en la mayor parte de los casos no revelaron alguna relación de similitud evidente entre las enzimas de los módulos conectados entre sí de esta forma, a diferencia de los módulos que se encuentran relacionados entre sí por conexiones bidireccionales ya que estos suelen compartir dominios que son del mismo clan o la misma súper familia. Sin embargo, también se encontraron algunos de estos módulos que no compartieron dichas características, lo que pudo deberse a la baja similitud que hubo entre algunos pares de enzimas conectadas bidireccionalmente. En este caso se vio que los módulos conectados de esta forma también pueden estar relacionados entre sí a través de interacciones funcionales que ocurren entre la(s) familia(s) de dominios de las enzimas de un módulo y la(s) de las enzimas de otro, como las que se muestran en la base de datos del Pfam [46].

Por último, uno de los resultados posiblemente significativo de las redes construidas fue que las enzimas del tercer nivel de similitud no mostraron un parecido con las del primero, pues no se hallaron conexiones directas entre las enzimas de ambos niveles, a diferencia de las enzimas pertenecientes al segundo nivel que todavía mantienen una similitud con el primero. Y a medida de que se van agregando niveles de similitud dentro de una red, por ejemplo, si se hace un cuarto nivel de parálogos a partir de algunas enzimas del tercero, tampoco aparece alguna similitud entre las secuencias de las enzimas del último nivel con las enzimas de los dos primeros. Finalmente, las enzimas resultantes del cuarto nivel de parálogos tuvieron funciones más especializadas, pero del mismo tipo que las de sus parálogos del tercer nivel.

### **3.2.1 Ribose-phosphate pyrophosphokinase (PrsA)**

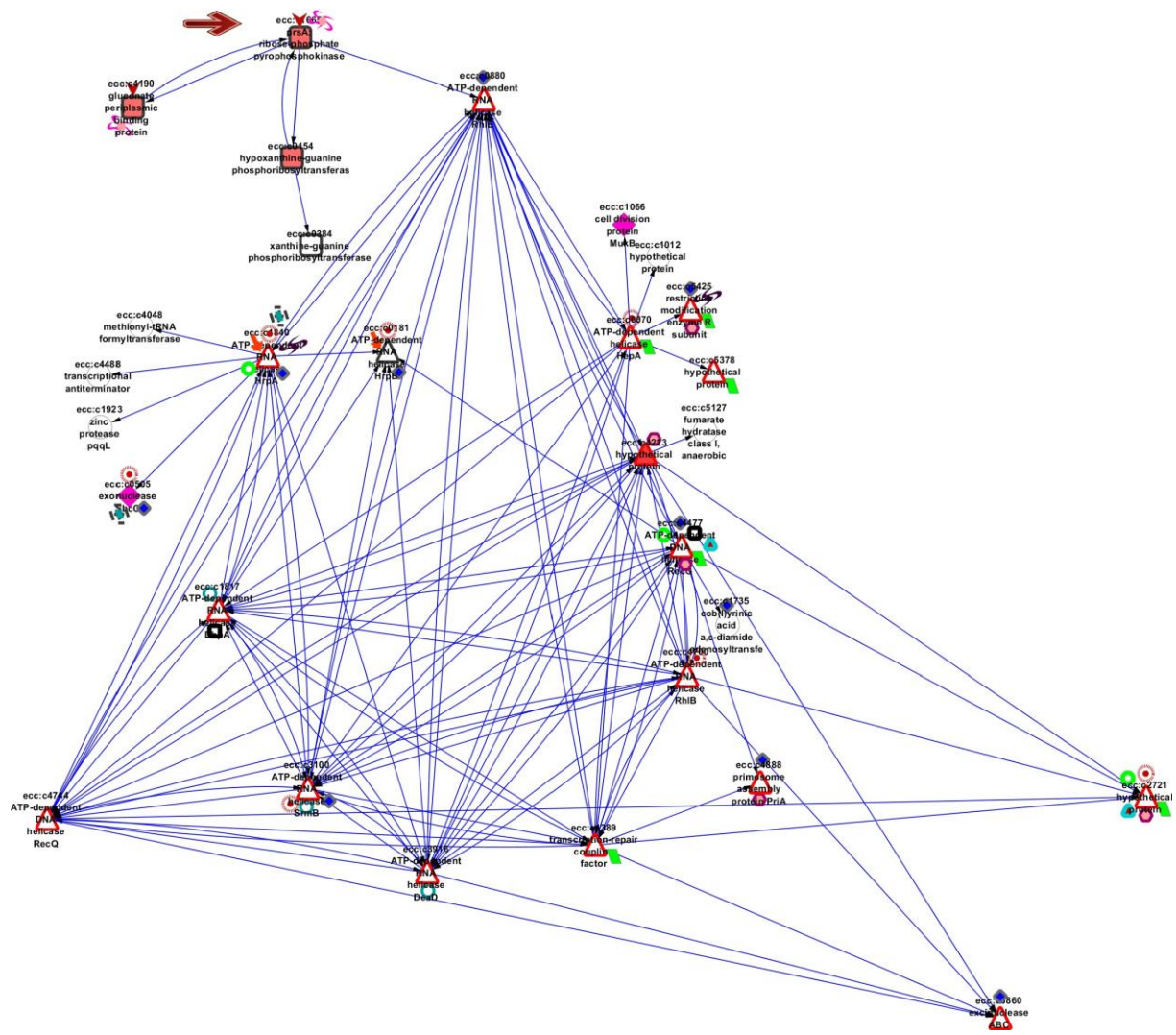
En esta red, de acuerdo con la aplicación de cytoscape (cluster one) para definir clusters de nodos, y viendo que grupos de estos comparten los mismos dominios en esta red, se definen dos módulos principales (Ver red No. 11) de los cuales el más importante es el que incluye a la enzima prsA y las enzimas que comparten mayor información a nivel de estructura primaria con esta. El segundo módulo está formado por la ATP-dependiente RNA helicase RhIE y sus “parálogos”. En la lista de dominios de prsA (Pág.82) se ven los dominios que comparten cada grupo de enzimas.



Como se observa en el primer módulo de esta red y de acuerdo con los valores del KEGG de identidad, solapamiento, tamaño, etc. (de las secuencias) la enzima que comparte mayor información de este tipo con PrsA, es la hypoxanthine-guanine phosphoribosyltransferase y con las que menos lo hace son con la gluconate periplasmic binding protein y con la xanthine-guanine phosphoribosyltransferase, que comparte un solo dominio con PrsA, de modo que no hay una conexión directa entre ambas enzimas.




















El segundo módulo prácticamente no comparte información con el primero y consiste en la ATP-dependiente RNA helicase RhlE y sus “parálogos”.

Estos resultados concuerdan con los reportados por Becerra et al [7]; al respecto de que las enzimas hypoxanthine-guanine, phosphoribosyltransferase y xanthine-guanine phosphoribosyltransferase son parecidas a PrsA en cuanto a su secuencia primaria. De hecho Becerra menciona que estas enzimas son homólogas entre sí. Aunque en el presente esquema la xanthine-guanine phosphoribosyltransferase, no tiene una conexión directa con la ribose-phosphate-pyrophosphokinase, estas dos enzimas si comparten información. La diferencia es que mientras PrsA y ésta enzima comparten un solo dominio (poca información no detectada por los algoritmos de comparación de secuencias del KEGG), la enzima PrsA comparte en cambio con la hypoxanthine-guanine phosphoribosyltransferase dos dominios.



Red 10 grande de enzimas homólogas de prsA

## Lista de Dominios compartidos en enzimas homóloga de prsA

-  Dominio UPRTase
  -  Dominio PRTase\_2
  -  Dominio Pribosyl\_synth
-  Dominio Pribosyltran
-  Dominios Pribosyltran, Helicase\_C, DEAD y ResIII
-  Dominios Helicase\_C, DEAD y ResIII
-  Dominios Helicase\_C y DEAD
  -  Dominio AAA\_22
  -  Dominio AAA\_19
  -  Dominio SNF2\_N
  -  Dominio AAA\_30
  -  Dominio CMS1
  -  Dominio DUF2075
-  Dominio SbcCD\_C, AAA\_29 y ABC\_tran
  -  Dominio Helicase\_RecD
  -  Dominio Zot
  -  Dominio HA2
  -  Dominio PhoH
  -  Dominio NTPase\_1

### CAPÍTULO 3. RESULTADOS

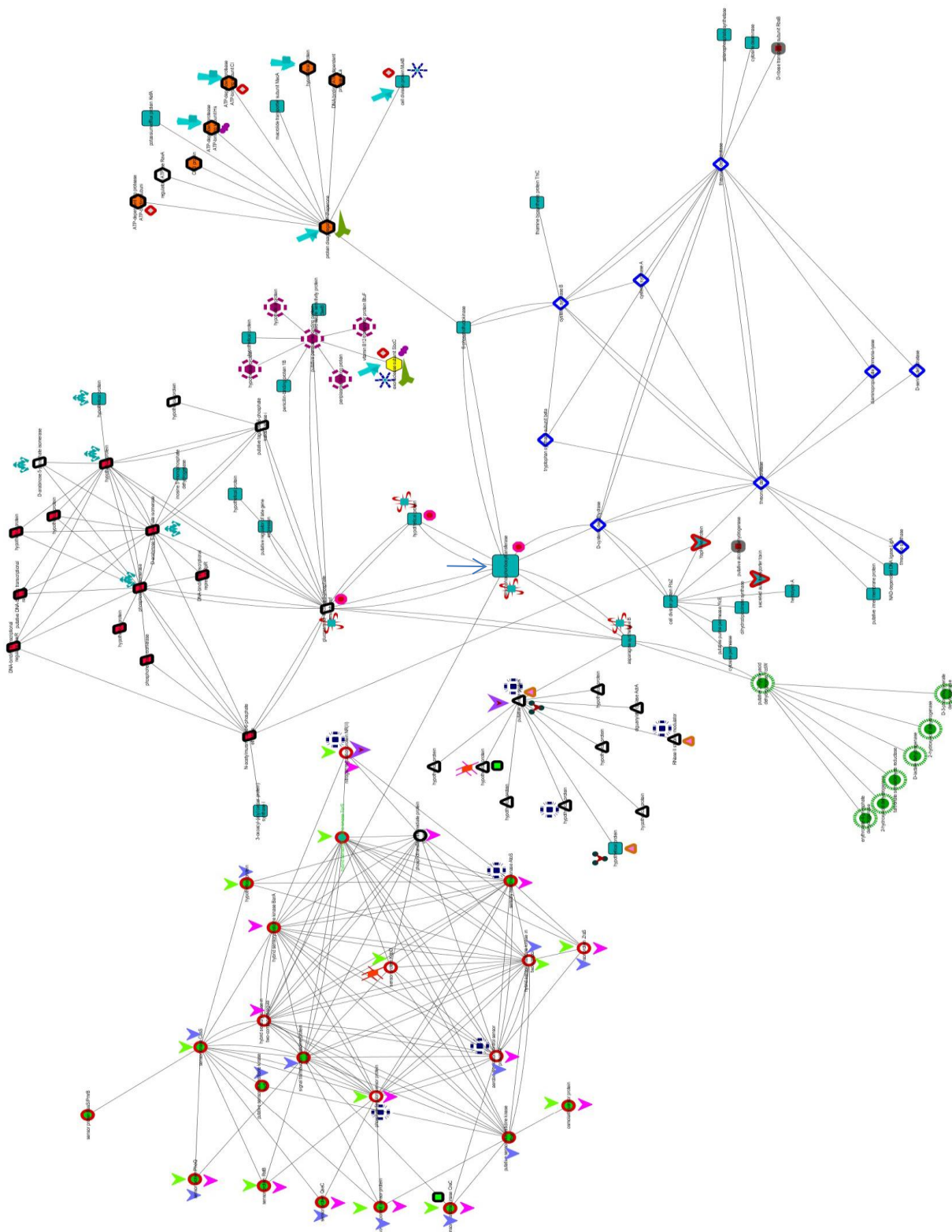
Nombre	KO	EC	Ruta metabólica	Función
ecc:c0384 xanthine-guanine phosphoribosyltransferase	K00769	(EC:2.4.2.22)	Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Transferasa
ecc:c0880 ATP-dependent RNA helicase RhlE	K11927	[EC:3.6.4.13]	RNA degradation	Helicasa
ecc:c4190 gluconate periplasmic binding protein	NO	NO	NO	Binding Protein
ecc:c0154 hypoxanthine-guanine phosphoribosyltransferase	K00760	(EC:2.4.2.8)	Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Transferasa
ecc:c1665 prsA; ribose-phosphate pyrophosphokinase	K00948	(EC:2.7.6.1)	Pentose phosphate pathway, Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments, Carbon metabolism, Biosynthesis of amino acids	Kinasa

Tabla 3. 13 Principales nodos de la red grande directamente relacionados con prsA en el genoma de E.Coli.

### 3.2.2 Amidophosphoribosyltransferase(PurF)

Los nodos principales de la red grande de purF, son los mismos que se observan en la red pequeña, purF y glms. Los primeros vecinos de purF en esta red son una enzima llamada hybrid sensory kinase 6 fosfofructoquinasa, una enzima D-cisteina sulfhidrase, una enzima asparagine sintasa B y una enzima hipotética. Esta red también se distingue por que cada uno de los primeros vecinos de purF está conectado con 1 o 2 módulos.

La red grande de PurF está compuesta de 7 módulos en total, de los cuales el programa cluster one solo identifica 3, los de mayor importancia. El primer módulo está compuesto de dos partes, la primera incluye a PurF y las enzimas más relacionadas (directamente conectadas) con esta a nivel de estructura primaria. Este grupo de enzimas (nodos) forman en la red un tetrámero de dobles conexiones (Red 12) lo que indica que hay un grado de similitud entre todas las secuencias de dichas enzimas. Algunos de los miembros de dicho grupo comparten los dominios GATase\_2 GATase\_6 y GATase\_7, otros comparten el dominio GATase\_4 (Red 12). La segunda parte es un grupo de enzimas que comparten los dominios SIS y SIS\_2 de las cuales solo 3 están conectados con la primera parte de dicho módulo. El segundo módulo de esta red está conectado con la enzima PurF, y está compuesto por varios nodos que comparten el dominio PALP. Finalmente, el tercer módulo está compuesto por un conjunto de varias enzimas que comparten todas principalmente el dominio HATPase\_c. Algunos de estos nodos también comparten otros dominios como los dominios HATPase\_c\_2 HATPase\_c\_3 HATPase\_c\_5 y los dominios HisKA y HAMP.



Red 11 grande de enzimas homólogas de purF

## Lista de Dominios compartidos en enzimas homóloga de purF

- Dominio HATPase\_c
- Dominio HATPase\_c y HisKA
- Dominio HATPase\_c, HisKA y HAMP
- ∩ Dominio SIS
- ▮ Dominio SIS Y SIS\_2
  - ▼ Dominio HATPase\_c\_3
  - ▼ Dominio HATPase\_c\_2
  - ▼ Dominio HATPase\_c\_5
- △ Dominio GGDEF
- ◇ Dominio PALP
- Dominios AAA, AAA\_16 y AAA\_5
- Dominios AAA\_17 y AAA\_22
- Dominios 2-Hacid\_dh\_C y NAD\_binding\_2
  - ▣ Dominios PAS y PAS\_9
  - ↘ Dominio ABC\_tran
  - ↗ Dominio CBS
- ⊗ Dominio Peripla\_BP\_2
  - ✦ Dominios GATase\_2, GATase\_6 y GATase\_7
- ♥ Dominio ESPR
  - ◇ Dominio AAA\_29
  - ▲ Dominio EAL
  - Dominio GATase\_4
  - Dominio GAF\_2
  - ✂ Dominio GAF\_3
  - ▼ Dominio MASE1
  - ↖ Dominio MobB
  - ◆ Dominio NTPase\_1
  - ▼ Dominio PAS\_10
  - Dominio Peripla\_BP\_6
    - \* Dominio SbcCD\_C

### CAPÍTULO 3. RESULTADOS

Nombre	KO	EC	Ruta metabólica	Función
ecc:c0505 exonuclease subunit SbcC	K03546	NO	NO PATHWAY	Exonuclease
ecc:c0749 asparagine synthetase B	K01953	(EC:6.3.5.4)	Alanine, aspartate and glutamate metabolism, Nitrogen metabolism, Metabolic pathways, Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments	Synthetase
ecc:c0373 hypothetical protein	K07008	NO	NO PATHWAY	Hipotetica
ecc:c2333 D-cysteine desulfhydrase	K05396	(EC:4.4.1.15)	Cysteine and methionine metabolism	Desulfhydrase
ecc:c4654 glucosamine-fructose-6-phosphate aminotransf	K00820	(EC:2.6.1.16)	Alanine, aspartate and glutamate metabolism, Amino sugar and nucleotide sugar metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Transferasa
ecc:c2856 amidophosphoribosyltransferase	K00764	(EC:2.4.2.14)	Purine metabolism, Alanine, aspartate and glutamate metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Transferasa

Tabla 3. 14 Principales nodos de la red grande directamente relacionados con PurF en el genoma de E.coli.



### 3.2.3 Phosphoribosylamine–glycine ligase (PurD)

En estos resultados podemos ver que la red de PurD grande está compuesta por 7 módulos aproximadamente, de los cuales solo hay tres de mayor importancia: el módulo principal, que incluye a PurD, y otros dos más que contienen cada uno una enzima que tiene dobles conexiones con PurD.

El módulo principal está compuesto de 7 enzimas en su mayoría interconectadas, que comparten diferentes dominios del clan ATP-grasp. Los dominios que aparecen con mayor frecuencia son: ATP-grasp\_3, GARS\_A, ATP-grasp\_Ter.

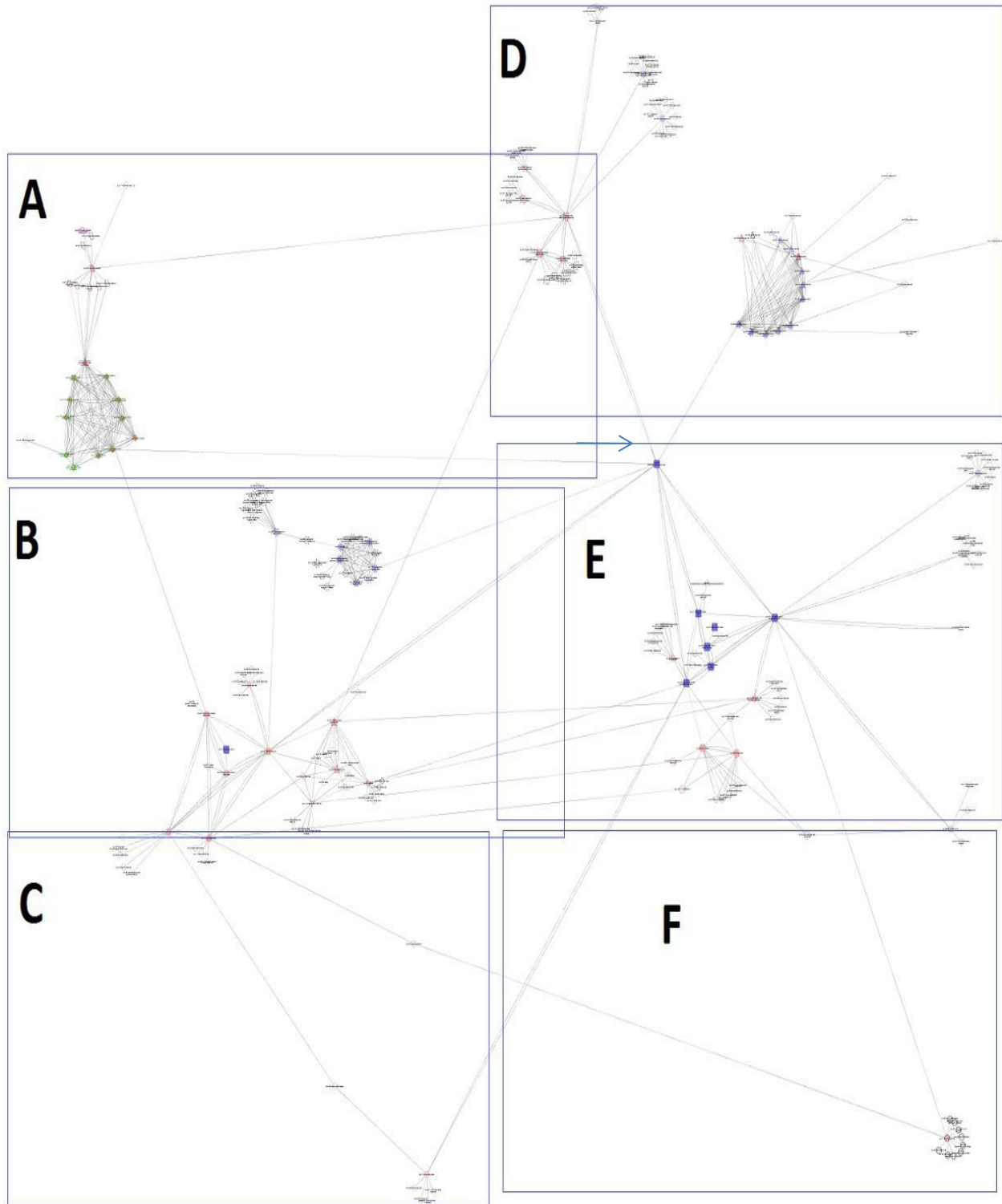
El segundo módulo importante de esta red contiene 5 enzimas que comparten los dominios Ligase\_CoA y Succ\_CoA\_lig pertenecientes al clan Succ\_CoA\_synth, y el dominio CoA\_binding del clan NADP\_Rossman. Por otro lado estos dominios presentan interacciones funcionales entre sí en diferentes enzimas: por ejemplo mientras que las enzimas del primer clan suelen interactuar con el dominio CoA\_binding del segundo clan, las enzimas del clan NADP\_Rossman suelen interactuar con los dominios Succ\_CoA\_lig y Ligase\_CoA del primer clan. Por último ambos clanes no solo tienen interacciones funcionales limitadas entre sí, sino que además presentan interacciones funcionales con uno de los dominios más importantes de la enzima PurD, el ATP-grasp-2.

El tercer módulo importante de la red está compuesto por 5 enzimas principales, las cuales comparten los dominios NMT1, el SBP\_bac\_3, y el phosphonate-bd, quienes pertenecen al clan PBP. De este módulo la enzima alkanesulfonate transporter (ecc:c1079) es la que está conectada con PurD (doble conexión) sin embargo en este caso, ninguno de sus dominios está relacionado con los de PurD ni siquiera funcionalmente, lo cual sucede pocas veces en otras redes que se revisaron.

Finalmente nos encontramos con que las enzimas del módulo principal de la red grande de PurD, pertenecen a muy diferentes rutas metabólicas, como se ve en la tabla (3.15) que incluye nucleótidos, carbohidratos y aminoácidos. Sin embargo las enzimas (4) más conectadas de dicho




























módulo excluyendo a PurD, tienen en común que participan en más de una ruta metabólica y todas están involucradas con rutas relacionadas con la síntesis de carbohidratos. Tres de ellas participan además en la síntesis de Alanina (1) y D-Alanina (2).

Por otra parte dos de las 3 enzimas que aparecen directamente relacionadas con PurD en mis resultados son: La acetyl-CoA carboxylase o Biotin carboxylase que es muy importante por que participa en 3 diferentes rutas metabólicas, y la Carbamoyl-phosphate synthase large chain que también participa en diferentes rutas entre ellas, la síntesis de pirimidinas en donde a partir de glutamina y PRPP se produce UMP. Por último, el módulo principal de la PurD se caracteriza por compartir el primer número enzimático 6 correspondiente a las enzimas que realizan la función de ligasas.



Red 12 grande de enzimas homólogos de purD

## Lista de Dominios compartidos en las enzimas homólogas de pur D

-  Dominio ATPgrasp
-  Dominio Coa\_binding
-  Dominio FIL\_HAEMAGG
-  Autotransporter
-  Dominio BIG
-  Dominio FLG\_BB
-  Dominio Usher y PapC
-  Dominio DUF
-  Dominio AMP-binding
-  Dominio NMT1, SBP\_bac y phosphonate
-  Dominio NMT1, SBP\_bac o phosphonate
-  Dominio SBP\_bac
-  Dominio ABC\_tran, SMC\_N y DUF258
-  Dominio ABC\_tran y SMC\_N
-  Dominio ABC\_tran
-  Dominio DDE\_Tnp\_IS66, ZF-IS66 y LZ\_TNP\_IS66
-  Dominio DDE\_THP\_IS66 y ZFIS66
-  Dominio DDE\_TNP\_IS66
-  Dominio LZ\_TNP\_IS66 y ZF-IS66
-  Dominio LZ\_TNP\_IS66
-  Dominio AldedH
-  Dominio AldedH, LuxC y DUF1487
-  Dominio AldedH y DUF
-  Dominio AldedH y LuxC
-  Dominio AldedH y FE-ADH
-  Dominio FE-ADH
-  Dominio HTH

Nombre	KO	EC	Ruta metabólica	Función
ecc:c4963	K01945	(EC:6.3.4.13)	Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Ligase
ecc:c0041	K01955	(EC:6.3.5.5)	Pyrimidine metabolism, Alanine, aspartate and glutamate metabolism, Metabolic pathways	Synthase
ecc:c0110	K01921	(EC:6.3.2.4)	D-Alanine metabolism, Peptidoglycan biosynthesis, Metabolic pathways	Ligase
ecc:c4012	K01961	(EC:6.4.1.2)	Fatty acid biosynthesis, Pyruvate metabolism, Propanoate metabolism, Metabolic pathways, Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments, Carbon metabolism	Carboxylase
ecc:c1079	K15553	NO	ABC transporters	Transporter
ecc:c4767	NO	NO	NO PATHWAY	Hypotetical
ecc:c1253	K06148	NO	ABC transporters	Transporter
ecc:c4534	K13922	NO	Glycerolipid metabolism Ethanolamine Utilization	Protein
ecc:c3658	NO	NO	NO PATHWAY	Hypotetical

Tabla 3. 15 Principales nodos de la red grande directamente relacionadas a PurD en el genoma de E.coli

### 3.2.4 Phosphoribosylglycinamide formyltransferase 1 (PurN)

En estos resultados se pudo ver que la red de PurN grande está compuesta por 7 módulos de los cuales solo hay 2 de mayor importancia (Red 14) El módulo más pequeño (5 enzimas en total) es el que contiene a PurN, como enzima central conectada a otras 3 enzimas mediante conexiones simples (Tabla 3.16) Sus componentes comparten el dominio Formyl\_trans\_N, siendo el único que posee PurN que pertenece a la familia del mismo nombre, la cual no pertenece a ningún clan. Por otra parte los nodos de dicho módulo, excepto uno, el ecc:c 1696 formyltetrahydrofolate deformylase (EC:3.5.1.10), comparten la primera posición y la segunda del número enzimático EC, 1 y 2 respectivamente, donde está representada la función de oxidoreductasa (1) y los sustratos sobre los que actúan estas enzimas: Aldehído o Cetona (2). Además, estos nodos tienen en común que pertenecen a rutas metabólicas relacionadas con la biosíntesis de carbohidratos y la de vitaminas y cofactores, habiendo 3 de cuatro que pertenecen a la ruta metabólica “One carbon pool by folate” incluyendo a PurN. Finalmente todas estas enzimas pertenecen a más de una ruta metabólica.

El segundo módulo contiene 13 enzimas interconectadas entre sí, todas ellas se distinguen por compartir el dominio Epimerasa, que pertenece al clan NADP\_Rossmann, además de tener otros dominios de este grupo que tienen en común varias enzimas, pero no todas. En este segundo módulo está presente la enzima (hipotética yfbG) bifuncional UDP-glucuronic acid dehydrogenase que está directamente conectada a PurN, perteneciendo también al primer módulo, por lo que tiene dominios del primer y el segundo conjunto de enzimas de esta red. La mayor parte de estas enzimas, aunque algunas son hipotéticas (5) y no se conoce el metabolismo al que pertenecen; otras están relacionadas (4) con el metabolismo de carbohidratos o con la síntesis de Terpenoides y Polyketides (2) (Tabla 3.16) Estas enzimas no hipotéticas tienden a participar en dos rutas metabólicas cada una, y no comparten entre todas una cifra específica del número EC, sin embargo, posiblemente hay una tendencia con respecto a este número dentro del presente módulo, encontrándose únicamente tres enzimas 5.1.3 - - - , dos 4.2.1 - - - y otras dos 1 - - - - -. Finalmente las enzimas de este último cluster se caracterizan por interactuar con sustratos que son o que contienen nucleótidos (tienen nombres que incluyen nucleótidos como UDP.)



## Lista de dominios compartidos en enzimas homólogas de PurN

- Dominio Epimerase
- Dominio Epimerase y adh\_short
- Dominios Epimerase, adh\_short y RmID\_sub\_bind
- Dominios Epimerase y RmID\_sub\_bind
- ▲ Dominio Aldedh
- Dominios HlyD\_2 y HlyD\_3
- Dominio HlyD\_2, HlyD\_3, HlyD y Biotin\_lipoyl\_2
- Dominio HlyD\_2, HlyD\_3, HlyD, Biotin\_lipoyl\_2 y Biotin\_lipoyl
- ◆ Dominio Amidohydro\_3, Amidohydro\_4 y Amidohydro\_5
- Dominio DEAD y Helicase\_C
- ▼ Dominio KR
- ◆ Dominio NAD\_binding\_10
- ✿ Dominio Polysacc\_synt\_2
- ⚙ Dominios 3Beta\_HSD y NAD\_binding\_4
- 🧬 Dominio NmrA
- ✂ Dominio Formyl\_trans\_N
- ✂ Dominio Formyl\_trans\_C

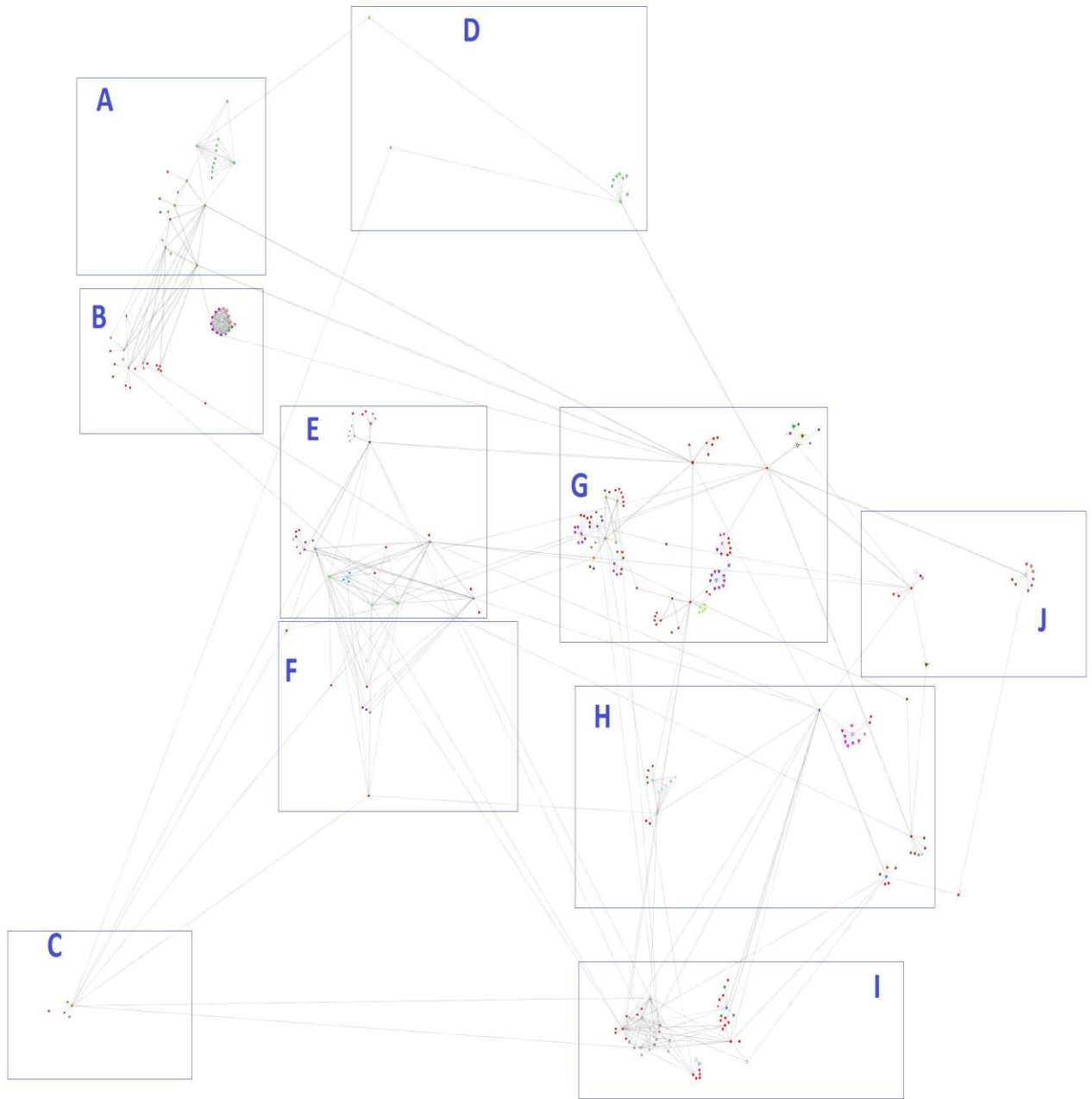


Nombre	KO	EC	Ruta metabólica	Función
ecc:c3018 phosphoribosylglycine formyltransferase	K11175	(EC:2.1.2.2)	Purine metabolism, One carbon pool by folate, Metabolic pathways, Biosynthesis of secondary metabolites	Formyltransferase
ecc:c1696 formyltetrahydrofolate deformylase	K01433	(EC:3.5.1.10)	Glyoxylate and dicarboxylate metabolism, One carbon pool by folate	Deformylase
ecc:c2797 bifunctional UDP-glucuronic acid decarboxylase	K10011	[EC:2.1.2.13 1.1.1.305]	Amino sugar and nucleotide sugar metabolism, Biosynthesis of secondary metabolites	Decarboxylase
ecc:c4048 methionyl-tRNA formyltransferase	K00604	(EC:2.1.2.9)	One carbon pool by folate, Aminoacyl-tRNA Biosynthesis	Formyltransferase

Tabla 3. 16 Principales nodos de la red grande directamente relacionadas a PurN en el genoma de E.coli

### **3.2.5 Phosphoribosylformylglycinamide synthase (purL)**

Esta red se compone de varios módulos siendo una de las redes más extensas de este trabajo. Aunque en esta se hizo un cuarto nivel de similitud de enzimas a comparación de las otras, que solo se hicieron hasta un tercer nivel. En este caso purL se encuentra conectada a 8 enzimas pertenecientes a diferentes módulos mediante dobles conexiones, como si ella misma no perteneciera a alguno en particular, sin embargo está directamente relacionada con la enzima hypE (ecc:c3290 hydrogenase isoenzyme formation protein) que es parte de un módulo de cuatro enzimas interconectadas entre sí, las cuales junto a purL, comparten todas el dominio AIRS, siendo todas de esta misma familia (pfam)[46]; que no pertenece a ningún clan (Red 15; Tabla 3.17). Este último módulo se distingue por tener solo dos enzimas que pertenecen a la misma ruta metabólica (PurL y PurM de la biosíntesis de purinas) mientras que las demás pertenecen a otras rutas diferentes. Finalmente, dos de ellas comparten los primeros 2 dígitos del número EC, siendo estos el 6 y el 3, mientras que las dos últimas enzimas comparten las cifras 2 y 7 de su número enzimático.



Red 14 grande de enzimas homólogas de purL

## Lista de Dominios compartidos en enzimas homólogas de purL

- Dominio AIRS
- Dominio Ldh\_2
- Dominio HSP70
- ▲ Dominio HSP70 y MreB\_Mbl
- ✦ Dominio HSP70 y Hydantoinase\_A
- ▲ Dominio HSP70 y DDR
- Dominio Ligase-CoA y CoA\_binding
- Dominio TPP\_enzyme
- ▼ Dominio TPP\_enzyme y POR\_N
- ✦ Dominio TPP\_enzyme y Co\_dh
- ✦ Dominio TPP\_enzyme, POR\_N y Co\_dh
- Dominio Aconitase
- Dominio Sigma
- Dominio DUF
- Dominio UvrD y AAA
- Dominio UvrD, AAA y viral\_helicase1
- ✦ Dominio Amidase\_3 y AMIN
- Dominio Amidase\_3 y TAT-Signal
- ✦ Dominio Amidase\_3, AMIN y TAT-Signal
- Dominio Fer4
- ✦ Dominio Fer4, Pyr\_redox, NAD\_binding, FAD\_binding, DAO, GIDA y 3HCDH
- ✦ Dominio Pyr\_redox, NAD\_binding, FAD\_binding y DAO
- Dominio PfkB
- Dominio PfkB y Phos\_pyr\_kin
- ✦ Dominio PfkB, Phos\_pyt\_kin y carb\_kinase
- ✦ Dominio PfkB, Phos\_pyr\_kin, Carb\_kinase, HTH y MarR
- Dominio PfkB, HTH y MarR
- ▼ Dominio Mannitol\_dh
- Dominio Rotamase
- Dominio Rotamase y SurA
- Dominio TMP\_2 y Tape\_meas\_lam
- Dominio Ms\_channel
- ✦ Dominio Molybdopterin
- Dominio HATPase
- Dominio Peptidase
- Dominio Peptidase, LysM y Biotin\_lipoyl
- Dominio Laminin
- Dominio Laminin, Phage\_tail\_N y CarboxypepD\_reg
- Dominio Phage\_tail\_N y CarboxypepD\_reg
- Dominio Asma
- Dominio BIG
- Dominio BIG y CarboxypepD\_reg
- Dominio CarboxypepD\_reg
- ✦ Dominio Peptidase, Autotransporter y ESPR
- ✦ Dominio Autotransporter y ESPR
- ✦ Dominio Autotransporter
- Dominio ESPR
- Dominio Alpha-amylase
- Dominio Alpha-amylase y Glyco\_hydro
- Dominio BenE
- ▼ Dominio BenE y Xan\_ur\_permease
- ▼ Dominio Xan\_ur\_permease
- Dominio ABhydrolase
- Dominio AA\_permease
- Dominio HTH y MarR
- Dominio HTH, TPR
- Dominio TPR
- ▲ Dominio ABC2\_membrane
- ▲ Dominio BPD\_transp

Nombre	KO	EC	Ruta metabólica	Función
ecc:c0557 peptidyl-prolyl cis-trans isomerase (rotamase D)	NO	NO	NO	Hypothetica
ecc:c0771 flavodoxin FldA	K07289	NO	NO	Assembly
ecc:c0933 hypothetical protein	NO	NO	NO	Hypothetica
ecc:c1745 aconitate hydratase	K03839	NO	NO	Flavodoxin
ecc:c2095 hypothetical protein	K01681	(EC:4.2.1.3)	Citrate cycle (TCA cycle), Glyoxylate and dicarboxylate metabolism, Metabolic pathways, Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments, Carbon metabolism, 2-Oxocarboxylic acid metabolism, Biosynthesis of amino acids	Hydratase
ecc:c2591 assembly protein	K00313	[EC:1.5.5.-]	NO	Hypothetica
ecc:c2643 hypothetical protein	K04655	NO	NO	Hydrogenase
ecc:c3080 phosphoribosyl-formylglycinamide synthase	K00313	(EC:1.5.5.1)	NO	Reductase
ecc:c3290 hydrogenase isoenzyme formation protein hypE	K00852	(EC:2.7.1.15)	Pentose phosphate pathway	Kinase
ecc:c3325 electron transfer flavoprotein-quinone oxidoreductase ygcN	K03770	(EC:5.2.1.8)	NO	Isomerase
ecc:c4680 ribokinase	K01952	(EC:6.3.5.3)	Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Synthase

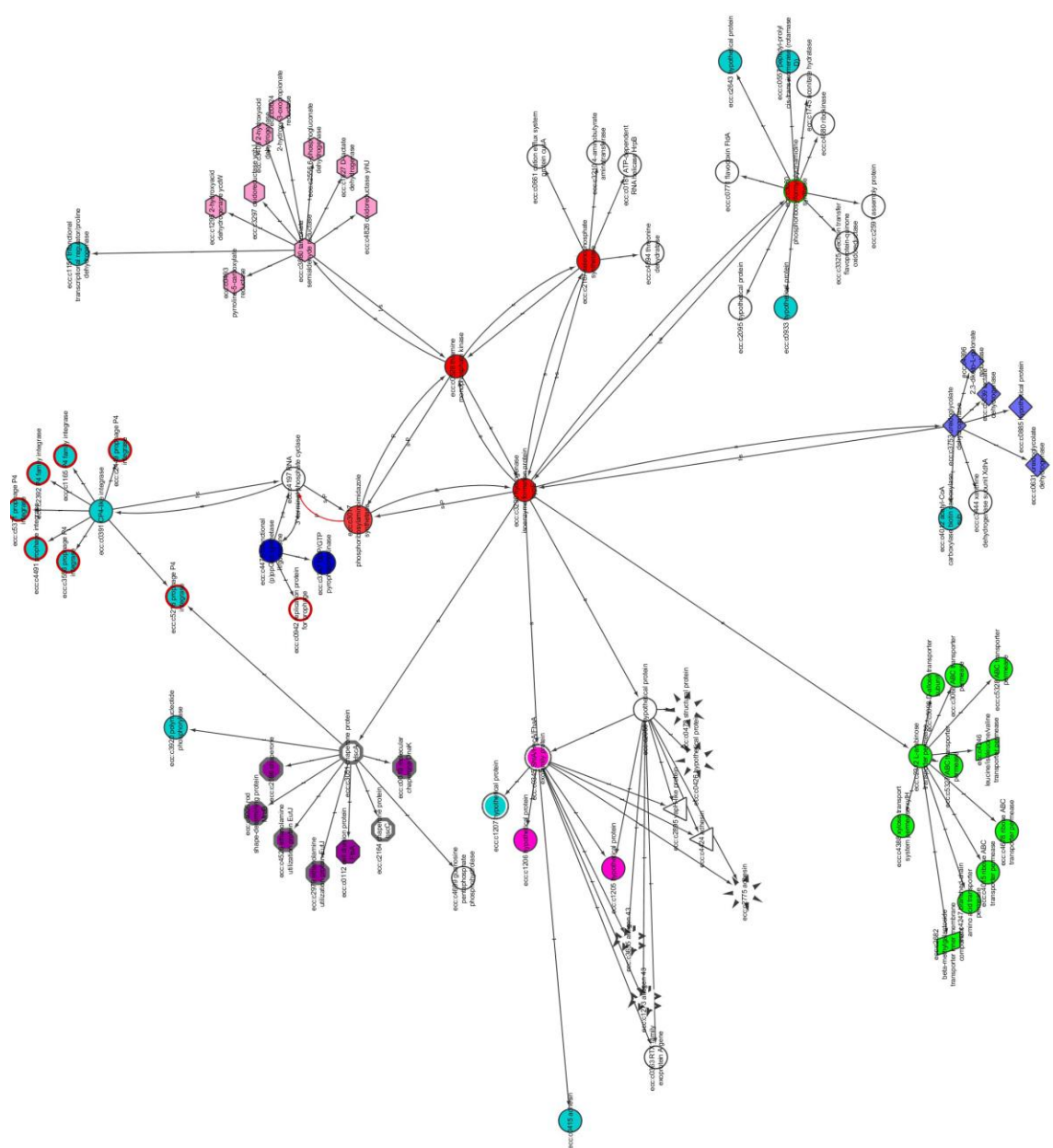
Tabla 3. 17 Principales nodos de la red grande directamente relacionados con PurL en el genoma de E.coli

### **3.2.6 Phosphoribosylformylglycinamide cyclo-ligase (purM)**

Esta red tiene 4 módulos incluyendo al principal, que contiene 4 enzimas, de las cuales cuatro (Tabla 3.18) son las mismas que se encuentran en el módulo principal de PurL y están conectadas del mismo modo. Cabe señalar que esta red fue construida a partir de la enzima hypE y no directamente de PurM, ya que en el kegg esta enzima no presentó enzimas homólogas directamente relacionadas con ella misma.

En esta red la enzima PurL también está conectada justo con las mismas enzimas que en su misma red.

Los puntos mencionados anteriormente son un indicio de que este método funciona, pues refleja correctamente que PurL y PurM al ser enzimas homólogas, compartieron los mismos homólogos entre sí, y a pesar de ser diferentes redes a las que perteneció cada enzima, compartieron la mayor parte de sus conexiones.



Red 15 grande de enzimas homólogas de purM

Lista de Dominios compartidos en enzimas homólogas de purM

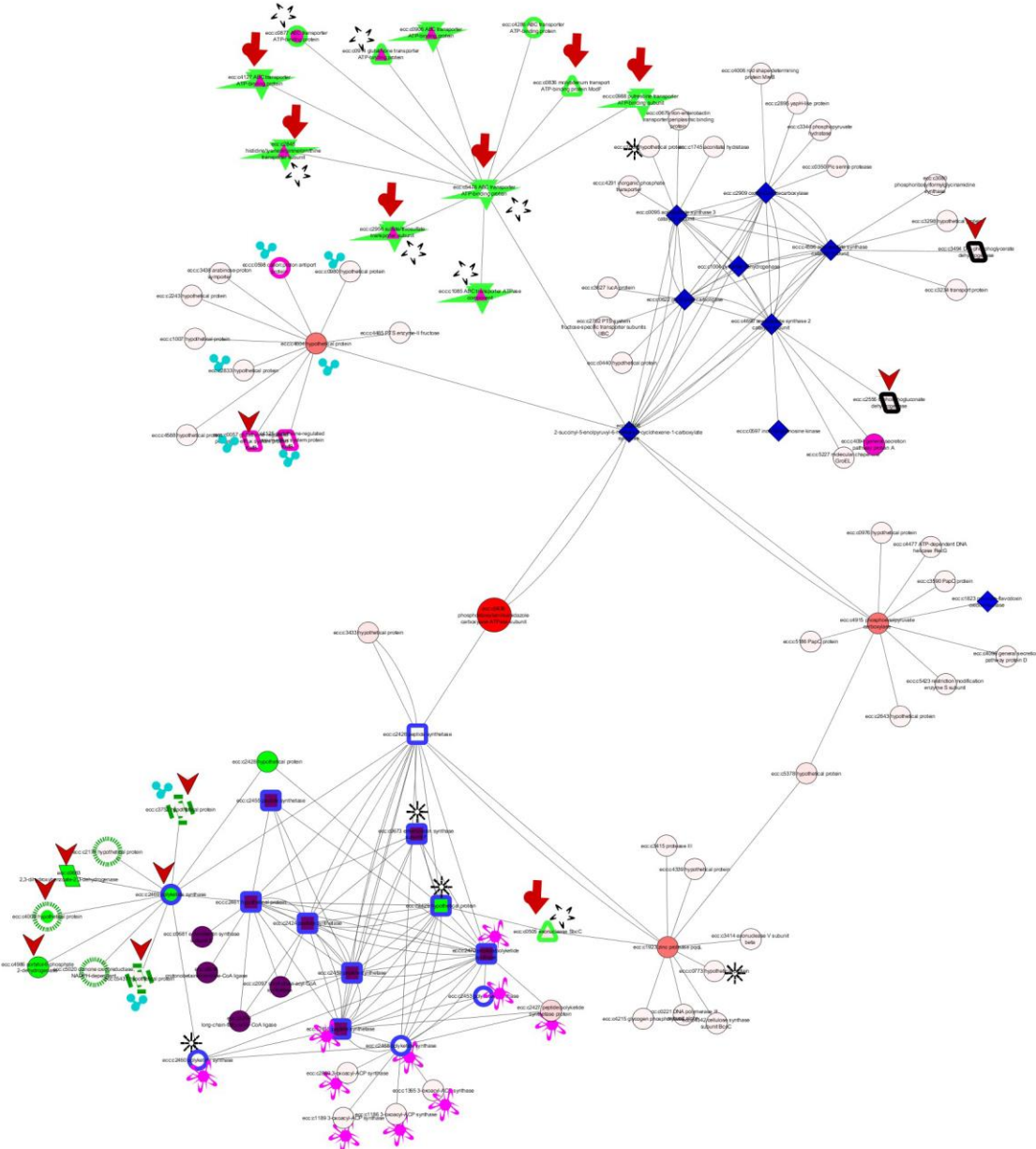
- Dominio AIRS
- △ Dominio ESPR
- Dominio Fil\_haemagg
- Dominio Toxin
- ✱ Dominio Autotransporter
- Dominio BPD\_transp\_2
- ▮ Dominio BPD\_transp\_2 y DUF
- ◆ Dominio Ldh\_2
- Dominio AIRS y GATase
- Dominio DUF
- ◇ Dominio NAD\_binding y 2-Hacid\_dh\_C
- Dominio Phage
- Dominio DUF y Phage
- Dominio HD\_4, RelA\_SpoT, TGS y ACT
- Dominio Mreb\_Mbl y HSP70
- Dominio Mreb\_Mbl y Hydantoinase\_A
- Dominio MreB\_Mbl, HSP70 y Hydantoinase



Nombre	KO	EC	Ruta metabólica	Función
ecc:c3017	K01933	(EC:6.3.3.1)	Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Synthetase
ecc:c3290	K04655	NO	NO	Hydrogenase
ecc:c4197	K01974	(EC:6.5.1.4)	NO	Cyclase
ecc:c0528	K00946	(EC:2.7.4.16)	Thiamine metabolism, Metabolic pathways	Kinase

Tabla 3. 18 Resultados de la red de enzimas homologas directamente relacionadas con Pur M dentro del genoma de *E.coli*.

### 3.2.7 5-(carboxyamino) imidazole ribonucleotide synthase (purK)



Red 16 grande de enzimas homólogas de purK

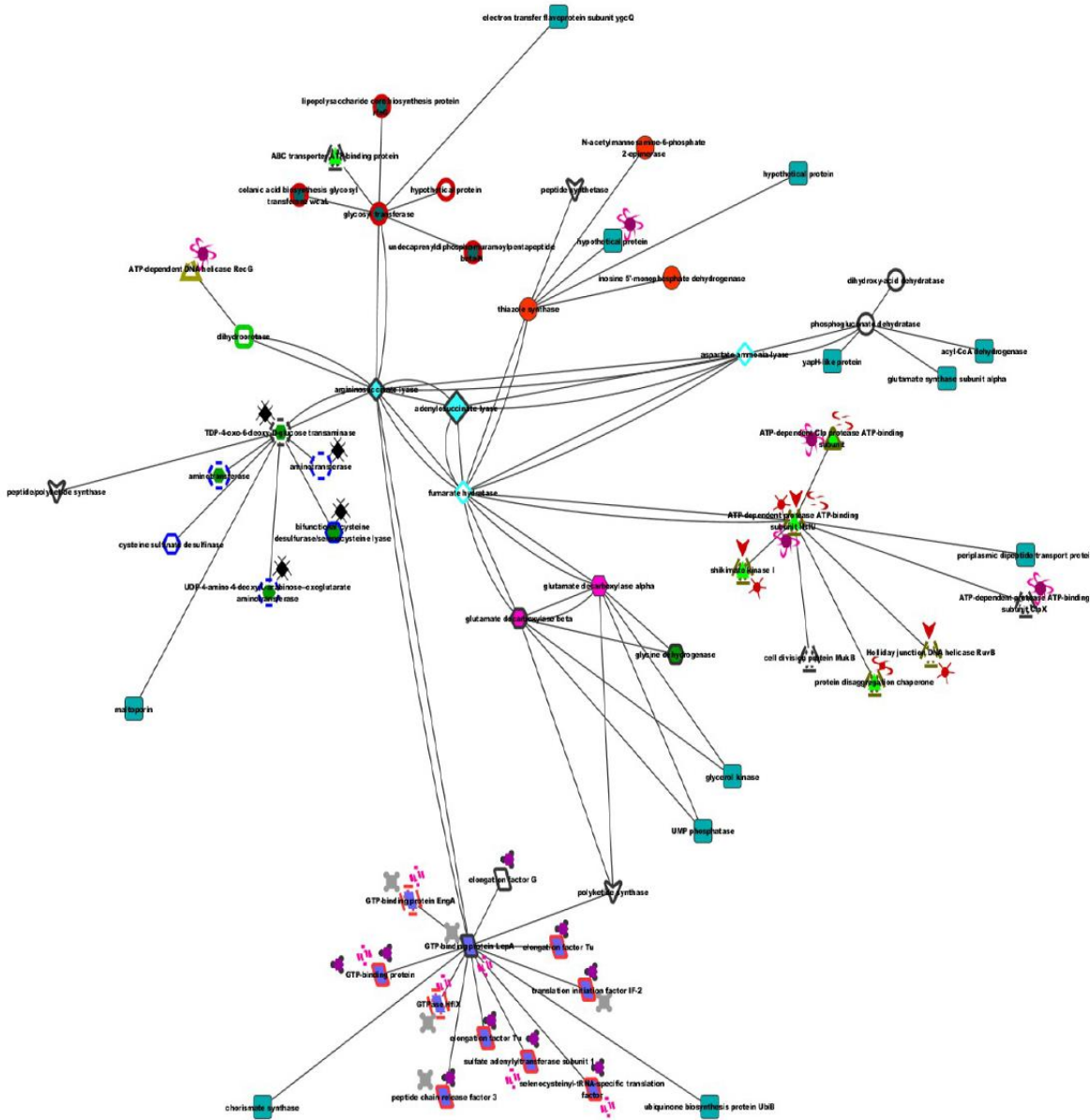
## Lista de Dominios compartidos en enzimas homólogas de purK

- Dominio HATPase\_c
- Dominio HATPase\_c y HisKA
- Dominio HATPase\_c, HisKA y HAMP
- ▽ Dominio SIS
- ▾ Dominio SIS Y SIS\_2
  - ▼ Dominio HATPase\_c\_3
  - ▼ Dominio HATPase\_c\_2
  - ▼ Dominio HATPase\_c\_5
- △ Dominio GGDEF
- ◇ Dominio PALP
- Dominios AAA, AAA\_16 y AAA\_5
- Dominios AAA\_17 y AAA\_22
- Dominios 2-Hacid\_dh\_C y NAD\_binding\_2
  - ⊞ Dominios PAS y PAS\_9
  - ↳ Dominio ABC\_tran
  - ↳ Dominio CBS
- ⊞ Dominio Peripla\_BP\_2
  - ✂ Dominios GATase\_2, GATase\_6 y GATase\_7
- ♥ Dominio ESPR
  - ◇ Dominio AAA\_29
  - △ Dominio EAL
  - Dominio GATase\_4
  - Dominio GAF\_2
  - ✂ Dominio GAF\_3
  - ♥ Dominio MASE1
  - ↳ Dominio MobB
  - ♥ Dominio NTPase\_1
  - ▼ Dominio PAS\_10
- Dominio Peripla\_BP\_6
  - \* Dominio SbcCD\_C

Nombre	KO	EC	Ruta metabólica	Función
ecc:c0636 phosphoribosylaminoimidazole carboxylase ATPase subunit	K01589	(EC:4.1.1.21)	Purine metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Carboxylase
ecc:c2426 peptide synthetase	NO	NO	NO	Synthetase
ecc:c2808 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase	K02551	(EC:4.1.1.71)	Ubiquinone and other terpenoid-quinone biosynthesis, Metabolic pathways, Biosynthesis of secondary metabolites	Synthase

Tabla 3. 19 Principales nodos de la red grande directamente relacionados con PurK en el genoma de *E.coli*

### 3.2.9 Adenylosuccinate lyase (purB)



Red 17 grande de enzimas homólogas de PurB

## Lista de Dominios compartidos en las enzimas homólogas de PurB

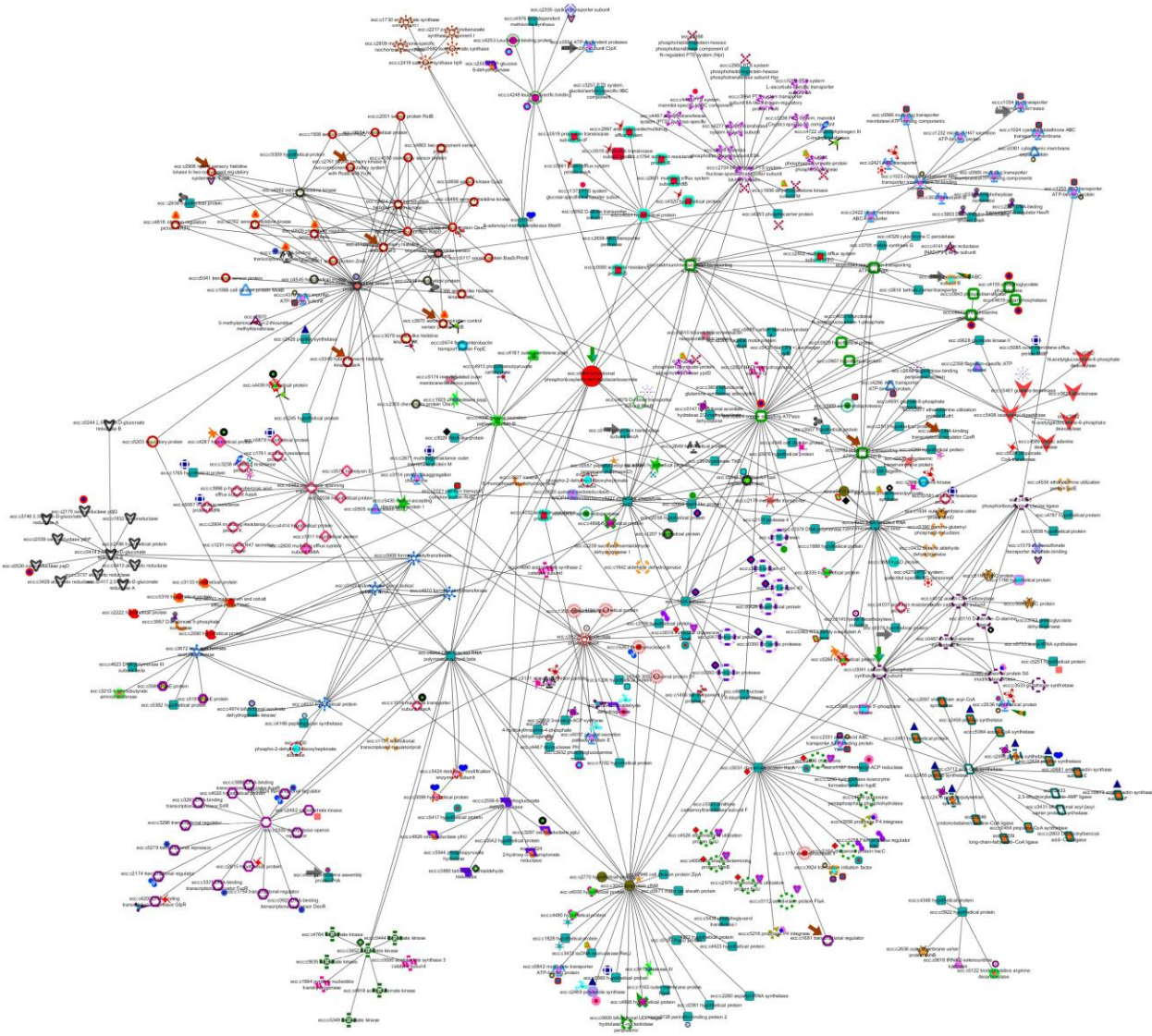
- ◇ Dominio Lyase\_1
- Dominio ASL\_C o ASL\_C2
- Dominio FumaraseC\_C
- △ Dominio AAA
- △<sub>1</sub>△<sub>2</sub>△<sub>3</sub> Dominio ABC\_tran
- ▲ Dominio Zeta\_toxin
- ▲ Dominio Sigma54\_activat, IstB\_IS21    ▲ Dominio IstB\_IS21 (solo)
- ↪ Dominio ClpB\_D2-small
- ▼ Dominio Mg\_chelatase
- ✳ Dominio RNA\_helicase
- Dominio Aminotran\_5
- Dominio Pyridoxal\_deC
- Dominio Amidohydro\_1 o Amidohydro\_2
- ▭ Dominio GTP\_EFTU
- ▭ Dominio MMR\_HSR1
- ▭ Dominio Miro
- ✳ Dominio SRPRB
- ◀ Dominio GTP\_EFTU\_D2
- ✳ Dominio Ras
- Dominio Aminotran\_1\_2
- Dominio Beta\_elim\_lyase
- Dominio Cys\_Met\_Meta\_PP
- ✳ Dominio DegT\_DnrJ\_EryC1
- ▽ Dominios ketoacyl-synt, ketoacyl-synt\_C, Thiolase\_N, PP-binding
- Dominio ILVD\_EDD
- ✳ Dominio ResIII
- Dominios ThiG, IMPDH, NanE
- Dominios Glyco\_trans\_1\_4
- Dominio Glyco\_trans\_4\_4

### CAPÍTULO 3. RESULTADOS

Nombre	KO	EC	Ruta metabólica	Función
ecc:c1510 adenylosuccinate lyase	K01756	(EC:4.3.2.2)	Purine metabolism, Alanine, aspartate and glutamate metabolism, Metabolic pathways, Biosynthesis of secondary metabolites	Lyase
ecc:c4919 argininosuccinate lyase	K01755	(EC:4.3.2.1)	Alanine, aspartate and glutamate metabolism, Arginine and proline metabolism, Metabolic pathways, Biosynthesis of secondary metabolites, Biosynthesis of amino acids	Lyase
ecc:c2003 fumarate hydratase	K01679	(EC:4.2.1.2)	Citrate cycle (TCA cycle), Metabolic pathways, Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments, Carbon metabolism	Hydratase
ecc:c5222 aspartate ammonia-lyase	K01744	(EC:4.3.1.1)	Alanine, aspartate and glutamate metabolism, Metabolic pathways	Lyase

Tabla 3. 20 Principales nodos de la red grande directamente relacionadas con purB en el genoma de *E.coli*

### 3.2.9 Phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase (purH)



Red 18 grande de enzimas homólogas de purH



## Lista de Dominios compartidos en las enzimas homólogas de PurH

- Dominio HATPase\_c
- Dominio HATPase\_c y PilJ
- Dominio HATPase\_c y HisKA
- △ Dominio ABC\_tran
- ▲ Dominios ABC\_tran y AAA\_22
- ▲ Dominio AAA\_22
- ▲ Dominios ABC\_tran, AAA\_22 y AAA\_16
- ▲ Dominio AAA\_16
- ▲ Dominios AAA\_16 y AAA\_22
- ◇ Dominios Biotin\_lipoyl\_2, HlyD y HlyD\_3
- ◇ Dominios Biotin\_lipoyl\_2, HlyD, HlyD\_3, PTS\_EIIA\_1 y FliJ
- Dominio AMP-binding
- Dominios AMP-binding y AMP-binding\_C
- Dominios AMP-binding, AMP-binding\_C, Thiolase\_N y ketoacyl-synt
- Dominio HTH\_11
- ▽ Dominio Aldo\_ket\_red
- Dominios HAD, Hydrolase y Hydrolase\_3
  - Dominio MMR\_HSR1
- ✳ Dominio PTS\_EIIA\_2
- Dominio SecD\_SecF
- Dominios SecD\_SecF y ACR\_tran
  - Dominio MreB\_Mbl
  - Dominio Miro
  - ▲ Dominio PP-binding
- ▽ Dominios Amidohydro\_1, Amidohydro\_3, Amidohydro\_4 y Amidohydro\_5
- ⌘ Dominios ATPgrasp\_Ter y ATP-grasp\_4
- Dominio Autotransporter
- Dominio CBS
- Dominio CBS y CorC\_HlyC
- Dominio CBS, CorC\_HlyC y TerC
- Dominio DDR
- ▽ Dominio F420\_oxidored
- ▽ Dominios F420\_oxidored, 2-Hacid\_dh\_C y 3HCDH\_N
- ▽ Dominios F420\_oxidored, 2-Hacid\_dh\_C, 3HCDH\_N y Shikimate\_DH
  - Dominio HSP70
  - Dominio OEP
  - ⌘ Dominio PTS-HPr
  - ▲ Dominio Response\_reg
- ▲ Dominio Aldeth
  - ⇒ Dominio DEAD
  - Dominio ESPR
  - Dominio HAD\_2
  - ▽ Dominio Hydrantoinase\_A
- Dominio S1
- Dominio S1, OB\_RNB y RNB
- Dominio AA\_kinase
- ☀ Dominio Chorismate\_bind
- ✳ Dominio Gly\_radical
- Dominio HTH\_17
- ▲ Dominio PAS, PAS\_8 y PAS\_9
- ▲ Dominio PEP-utilizers
- Dominio FGGY\_C
- Dominio Methyltransf\_31
- ▽ Dominio MMPL
- ▽ Dominio PapC\_C, PapC\_N y Usher
  - Dominio Peptidase\_M23
  - ▲ Dominios SIS y SIS\_2
  - ▲ Dominio SRP54
- ✳ Dominio TPP\_enzyme\_M
- ✳ Dominio TPP\_enzyme\_M, TPP\_enzyme\_N, TPP\_enzyme\_C y POR\_N
- ▽ Dominio 5\_nucleotid\_C y Metallophos
- ▽ Dominio 5\_nucleotid\_C, Metallophos y PGA\_cap
- ▲ Dominio AAA\_31
- ▽ Dominio AsmA\_2
  - ▽ Dominio Band\_7
  - Dominio CarboxypepD\_reg
  - ✳ Dominio CoA\_trans
  - ▲ Dominio DUF3816
  - ▽ Dominio Phage\_integrase y CENP-W
    - ▲ Dominio DUF4102
    - Dominio Epimerase
    - Dominio HTH\_12
    - ▲ Dominio IncA
  - ▽ Dominio Ketoacyl-synt\_C
  - ▲ Dominio NAD\_binding\_10
- Dominio Peptidase\_S46
- Dominios Peptidase\_S46, PDZ, PDZ\_2, Trypsin, Trypsin\_2
  - ✳ Dominio Peripla\_BP\_1
- Dominio Peripla\_BP\_6
- Dominio Peripla\_BP\_6, ANF\_receptor, Fe-ADH y Peripla\_BP\_5
  - Dominio PGM\_PMM\_I
  - ▽ Dominio SBP\_bac\_3
  - Dominio AceK
  - Dominio ADH\_N
  - Dominio AlaDh\_PNT\_N
  - Dominio Aminotran\_1\_2
  - Dominio AP\_endonuc\_2
    - Dominio Big\_1
    - ▲ Dominio Big\_3\_4
    - Dominio Carb\_kinase
    - Dominio Complex1\_51K
  - Dominio DAHP\_synth\_1
  - Dominio DUF2061
  - Dominio DUF2776
  - Dominio DUF3650, Toxin\_50 y PT-VENN
    - ▲ Dominio DUF748
    - ▽ Dominio DXP\_reductoisom
  - Dominios EndoU\_bacteria, Fimbrial y Lipocalin
    - Dominio Fil\_haemagg\_2
  - Dominios GNVR y Wzz
  - Dominio IF2\_N
    - ▽ Dominio Kinesin
- Dominio MacB\_PCD
- ▽ Dominio MCPsignal
- ✳ Dominio MFS\_1
  - ▲ Dominio MG1
  - ▼ Dominio MGS
  - ▲ Dominio NAD\_binding\_8 y Pyr\_redox\_3
    - Dominio OKR\_DC\_1, OKR\_DC\_1\_C y OKR\_DC\_1\_N
    - Dominio PduV-EutP
    - ▽ Dominio PEPcase\_2
  - ▲ Dominios Peptidase\_M16 y Peptidase\_M16\_C
  - ▲ Dominios Plug y TonB\_dep\_Rec
  - Dominio Phosphonate-bd y NMT1
    - Dominio PPOK\_N
    - Dominio PTS\_EIIC
  - Dominio Reo\_sigmaC
    - Dominio RNase\_PH y RNase\_PH\_C
  - Dominio Secretin y Secretin\_N
  - Dominio SLBB
    - ▲ Dominio Spc7
    - × Dominio TPR\_MLP1\_2
    - Dominio TrwB\_AAD\_bind
    - Dominio Trypan\_PARP
  - ▲ Dominios Eco57I, MTS y UPF0020
    - ✳ Dominio Y\_Y\_Y
    - Dominio Transpeptidase

Nombre	KO	EC	Ruta metabólica	Función
ecc:c4964 bifunctional phosphoribosylaminoimidazolecarboxamide	K00602	(EC:2.1.2.3 3.5.4.10)	Purine metabolism, One carbon pool by folate, Metabolic pathways, Biosynthesis of secondary metabolites	Transferase
ecc:c0041 carbamoyl phosphate synthase large subunit	K01955	(EC:6.3.5.5)	Pyrimidine metabolism, Alanine, aspartate and glutamate metabolism, Metabolic pathways	Synthase
ecc:c4262 zinc/cadmium/mercury/lead-transporting ATPase	K01534	(EC:3.6.3.3 3.6.3.5)	NO	ATPase
ecc:c0100 S-adenosyl-methyltransferase MraW	K03438	(EC:2.1.1.- )	NO	Transferase
ecc:c4944 DNA-directed RNA polymerase subunit beta	K03043	(EC:2.7.7.6)	Purine metabolism, Pyrimidine metabolism, Metabolic pathways, RNA polymerase	Polymerase
ecc:c0908 formate acetyltransferase 3	K00656	(EC:2.3.1.54)	Pyruvate metabolism, Propanoate metabolism, Butanoate metabolism, Metabolic pathways	Transferase
ecc:c1682 nitrate/nitrite sensor protein NarX	K07673	(EC:2.7.3.- )	Two-component system	Kinase
ecc:c5227 molecular chaperone GroEL	K04077	NO	RNA degradation	Chaperone

Tabla 3. 21 Principales nodos de la red grande directamente relacionados con PurH en el genoma de E.Coli

## 4. Discusión

### 4.1 Discusión general de la metodología

Pudimos aprender que la diferencia entre los valores de comparación de secuencias de cada metodología nos dio una resolución distinta en los resultados obtenidos. Mientras que con la primera metodología que uso valores más estrictos, la escala de comparación fue de enzimas completas, y esto nos permitió identificar duplicaciones totales; con la segunda, que utilizó parámetros de similitud más laxos, la escala de comparación fue a nivel de dominios con lo que nos fue posible identificar la presencia de duplicaciones a este nivel.

### 4.2 Discusión de los aportes de la primera metodología

Lo que nos aportó el análisis realizado con la primera metodología fue, en primer lugar, que pudimos conocer tanto las principales enzimas homólogas con diferente función (muy limitadas en número) como las enzimas fusionadas de la ruta de la biosíntesis *de novo* de purinas. Por otro lado nos permitió conocer su distribución filogenética y su conservación en los clados principales de los tres dominios (y por tanto su antigüedad relativa) y la de sus homólogos, así como la de sus enzimas fusionadas. También pudimos identificar el nivel de redundancia (número de copias) de estas enzimas en cada organismo de los considerados.

A pesar de los buenos resultados que se obtuvieron con este método, hubo ciertas desventajas:

- 1) Al utilizarse únicamente las secuencias de la RBNP de *E. coli* como base de búsqueda de cada enzima de la ruta de biosíntesis *de novo* de purinas en otros organismos, quedaron excluidas del estudio las variantes típicamente utilizadas por las arqueas (purT, purP y purO) y los eucariontes (PurE clase II, etc.).

- 2) El tamaño de la muestra de los organismos utilizados fue insuficiente para poder identificar un

patrón más amplio y fidedigno de la distribución filogenética de los homólogos encontrados de las enzimas de la RBNP.

Por lo tanto se recomienda en análisis posteriores que la muestra sea más amplia (un número representativo de organismos por dominio) y que sean consideradas para la comparación las secuencias no sólo de una bacteria, sino las de un eucarionte y las de una arquea. Si se pretende hacer un trabajo de redes de similitud de secuencias, no conviene hacer una búsqueda de similitud de secuencias utilizando parámetros estrictos (Por ejemplo: una cobertura mayor o igual a 80%), al menos para rutas muy conservadas como la RBNP.

### **4.3 Discusión de los aportes de la segunda metodología**

Lo que nos aportó el análisis de la segunda metodología, fue el descubrimiento de una gran cantidad de información, no sólo referente a la perspectiva evolutiva de la ruta de RBNP, sino también, con respecto a la estructura de las redes de similitud y su significado biológico y evolutivo.

En cuanto a sus aportaciones prácticas, la ventaja más importante de esta metodología propuesta, con respecto a otras similares de trabajos recientes, fue que las redes de similitud de secuencias, construidas en esta investigación, se caracterizaron por ser direccionadas, lo cual tuvo la utilidad de poder corroborar la similitud de dos enzimas teóricamente relacionadas, con bajo nivel de conservación de la secuencia, cada vez que se encuentra una conexión bidireccional entre dos nodos.

Otra de las ventajas de la segunda metodología de este trabajo, fue la de haber construido dichas redes de similitud a partir de información limitada, y no como en el caso de redes que se han hecho con demasiada información y han sido difíciles de procesar. También tuvo la ventaja, de que las redes de similitud fueran construidas con niveles subsecuentes de enzimas “parálogas” en vez de utilizar una gran cantidad de enzimas sin conocer el tipo de homología (ortólogos y

parálogos) de las secuencias utilizadas [50]; y la de haber utilizado para ello, como fuente de información, los datos ya clasificados, previamente analizados (comparación de sec.) y filtrados de la base de datos Kegg (2011). El mismo hecho de haber utilizado la versión 2011 de la base de datos Kegg, fue otra cualidad metodológica, ya que sus porcentajes de similitud de comparación de secuencias resultaron ser los precisos para encontrar similitudes entre las secuencias de enzimas, a nivel de dominio, y en menor escala. Un último atributo de la segunda metodología fue que nos permitió entender de manera más clara las relaciones de similitud entre las enzimas conectadas, gracias a la identificación y el señalamiento de los dominios de las enzimas de las redes construidas.

Sin embargo, dicha metodología también presentó algunas desventajas, como la de estar limitada a la inclusión de las secuencias de un solo organismo.

#### **4.4 Discusión de la organización de la estructura de las redes de similitud de secuencias**

Lo que observamos en las redes grandes al darnos cuenta de que hay varios grupos de enzimas muy interconectados entre sí similar a lo que se ha encontrado en otros trabajo [50]; lo que podría significar que tienen una estructura modular.

Dichos módulos como se vio en los resultados se distinguieron por compartir todos un promedio de dos dominios entre sus componentes. Lo que significa que aparentemente la resolución mínima de este método podría ser en general de 2 dominios, como se ve en el caso del primer módulo de la de red grande de PrsA, donde el método falla en identificar la similitud de dos enzimas que comparten un solo dominio [7].

En cuanto a que se encontraron solamente pocos dominios, que estuvieron ampliamente distribuidos en diferentes enzimas de varios módulos, en cada red, podríamos pensar que la distribución general de estos dominios en diferentes enzimas podría concordar con una ley de potencias [35]; aunque esto también se podría deber a la información de los datos, ya que se desconoce mucho de estos dominios.

Es de destacar que el hallazgo más importante de esta sección, el hecho de que las enzimas de un mismo módulo tuvieran la misma clase (función enzimática general) concuerda con los resultados de Patricia Babbitz [50]; quien menciona que es posible que las enzimas dentro de un mismo cluster de una red de similitud de secuencias tengan la misma función. Asimismo Atkinson [51]; explica que este comportamiento es consistente con el “dogma” de la biología estructural, que establece que la conservación de la secuencia implica una conservación estructural y a su vez esta última implica una conservación funcional entre las enzimas.

Por otro lado el hecho mencionado de que las enzimas de un mismo módulo comparten mínimamente 2 dominios en común, además de la primera cifra del número enzimático EC, podría significar que la función general de las enzimas está determinada por una parte relativamente pequeña de su propia secuencia.

Con respecto al hallazgo de algunos módulos, conectados entre sí a través de dobles aristas, que no compartieron dominios relacionados evolutivamente (del mismo clan por ejemplo) sino dominios diferentes que pueden presentar interacciones funcionales entre sí, frecuentes entre sus familias (sección de interacciones, Pfam) [46]; o en un solo grupo más específico de enzimas (sección de arquitecturas, Pfam) [46]; podemos decir que para que se den estas interacciones estructurales entre dominios diferentes es posible que tenga que haber entre ellos secuencias mínimas similares para que estos puedan ensamblarse o interactuar.

También se encontraron ciertos casos en los que dos módulos conectados bidireccionalmente entre sí, no compartieron enzimas con dominios relacionados evolutivamente, ni enzimas con dominios que presentaran interacciones funcionales entre sus familias, lo que podría explicarse porque compartieron secuencias similares muy pequeñas que pudieran ser importantes como las correspondientes a motivos enzimáticos, o secuencias muy pequeñas y tal vez irrelevantes. Una segunda explicación al respecto es que posiblemente algunos nodos de diferentes módulos con dobles conexiones entre si estén presentando una relación de similitud individual entre ellos, siendo estos los únicos entre los componentes de cada módulo que comparten algún parecido entre sus secuencias, posiblemente de menor grado.

En conclusión, con respecto a las enzimas que presentaron dominios de la misma familia, dentro de su estructura y de las que compartieron dominios estando dentro de un mismo módulo, así como a las que estuvieron relacionadas estando en diferentes módulos, podemos decir que probablemente los 3 casos se explican porque en todos ellos se han dado duplicaciones a nivel de dominio o menores.

## **Discusión de observaciones adicionales**

En dos observaciones breves que se hicieron al revisar las conexiones entre diferentes módulos, para identificar qué tipos de enzimas son hubs dentro de las redes, se encontró lo siguiente: a) En primer lugar, con este método aparentemente se pueden identificar enzimas que han adquirido múltiples dominios a partir de identificarlas como hubs, como se ve en la red grande de purN, donde se observa que hay una sola enzima (nodo) que se encuentra conectada con diferentes nodos correspondientes a diferentes módulos (2 principales) cada uno con diferentes tipos de dominios, siendo esta enzima la única en tener los diferentes dominios de cada uno de los módulos mencionados. b) En segundo lugar, se identificó (red grande purL) que uno de los módulos con una mayor cantidad de nodos conectados con otros de diferentes clusters, fue uno que contenía únicamente enzimas de membrana, lo cual posiblemente muestra que este tipo de enzimas son las que más comparten información a nivel de secuencia (especialmente pedazos pequeños, menores a un dominio) con otras enzimas pudiendo ser muy redundantes en este sentido.

### **4.5 Discusión sobre las implicaciones evolutivas de los resultados del primer análisis realizado (Redes Chicas)**

Con respecto a los hallazgos de este primer estudio, en cuanto a que las enzimas de la biosíntesis de purinas hayan sido prácticamente repeticiones de cada una, teniendo el mismo número KO y probablemente la misma función en los genomas de los organismos estudiados, encontrándose en general, muy pocas enzimas homólogas con diferente función que han presentado nula o poca divergencia. Podría ser, en primer lugar, debido a que la ruta de biosíntesis de purinas está altamente conservada, y en segundo lugar a que las funciones de las enzimas de esta misma ruta

se encuentran tan especializadas que salvo en algunos casos, sus homólogos pueden tener funciones diferentes.

A pesar de lo anterior, se encontró que las diferentes enzimas de la ruta de purinas presentan diferentes grados de conservación, distribución filogenética y nivel de redundancia, lo cual podría estar relacionado con el grado de dificultad de las reacciones químicas que procesan y la especificidad de las mismas o de su sustrato.

Por último, se encontró que algunas enzimas de la ruta de purinas se encuentran fusionadas entre sí, y que estas presentan una distribución peculiar (como se discutirá más adelante) tanto como aquellas que no están fusionadas.

#### **4.5.1 Discusión sobre las diferencias del grado de conservación evolutiva y distribución filogenética entre las enzimas de la ruta de biosíntesis de purinas**

Con respecto a lo mencionado sobre las diferencias en el grado de conservación y distribución filogenética que mostraron los componentes de la ruta de la biosíntesis *de novo* de purinas, a pesar de su conservación podemos decir que la PrsA, por ejemplo, es la enzima más conservada de la ruta, ya que se encontró presente en el 99% de los organismos que se consideraron en este trabajo. Esto significa que no hay una diferencia significativa entre las PrsA de los 3 dominios, sugiriendo que esta enzima no ha cambiado mucho a lo largo de su evolución. El nivel de conservación de PrsA es tal, que no presenta en ningún dominio, proteínas homólogas que realicen funciones diferentes a ella misma, viéndose únicamente copias de la misma PrsA en varios organismos. En sí, es curioso que se hayan encontrado este tipo de duplicaciones de PrsA y que ninguna de ellas haya divergido, no se sabe porque puede suceder esto, y si tiene que ver con algo en especial, por ejemplo, como que PrsA al ser tan importante para el metabolismo, se haya favorecido su redundancia en bacterias y eucariontes. La última explicación de esto es que haya habido alguna presión selectiva o una limitante estructural que no hayan permitido la divergencia de PrsA.



Otra enzima muy conservada de la ruta de purinas es PurF, pues se encuentra presente en todas las bacterias consideradas y en la mayor parte de las arqueas. En los eucariontes, sin embargo, la enzima purF se encuentra totalmente ausente en algunos casos, mientras que en muchos otros en lugar de la purF normal aparecen organismos con enzimas hipotéticas, algunas de ellas con el mismo ko.

En conclusión, mis resultados muestran, bajo la luz del presente análisis, que purF tiene un mayor grado de conservación en las bacterias a diferencia de los otros dominios.

Otras dos enzimas con un grado de conservación aparentemente similar al de purF fueron purL y purB. Con respecto a la primera, podemos decir, que está en general muy conservada ya que no se encontró ausente en bacterias, en las arqueas faltó en pocos casos (7) al igual que en los eucariontes, aunque estos últimos presentaron varias enzimas hipotéticas similares a purL y un tipo de purL (purL larga) un poco diferente a la que tuvieron la mayor parte de las bacterias y todas las Arqueas (purL corta), lo que coincide con información ya reportada [42]. Por último, purL solo presentó una enzima aparentemente homóloga que fue purQ. Con respecto a purB al igual que purL y purF también podemos decir que es una enzima que está muy conservada, por las pocas ausencias que presentó en los tres dominios.

Por otra parte, revisando la distribución filogenética y el grado de conservación evolutiva de otras enzimas de la RBNP menos conservadas, tenemos lo siguiente. En un primer grupo de enzimas, formado por purD, purN y purM se observó, como ya se había descrito [42]; que en las bacterias y las arqueas dichas enzimas se encuentran en forma individual a comparación de lo que se ha visto en la mayor parte de los eucariontes. PurD, purN, y purM también tuvieron en común un alto nivel de conservación (sin ausencias) y poca redundancia en el grupo de las bacterias, sin embargo, estas tres enzimas presentaron diferentes grados de conservación en las arqueas y en los eucariontes, siendo iguales en este sentido purD y purM, ya que se encontraron en los mismos eucariontes (solamente en plantas) y en las mismas arqueas (15 euriarqueotas y 3 crenarqueotas), a diferencia de purN, que presentó una menor conservación en dichos dominios. Probablemente esto significa que purD y purM son enzimas que han coevolucionado a lo largo de su historia evolutiva. El hecho de que solo purD y purM estén conservadas únicamente como enzimas individuales en las plantas, a diferencia del resto de los eucariontes, quiere decir:

a) Que probablemente las plantas adquirieron la purD y la purM a partir del ancestro de los cloroplastos.

b) Que por alguna razón desconocida, las plantas habiendo heredado este par de enzimas de sus ancestros arqueanos, fueron las únicas dentro de los eucariontes que se adaptaron a vivir con estas enzimas en su forma individual y no fusionadas como se encuentran en el resto de los eucariontes, aunque se requiere una mayor evidencia para corroborar esta suposición.

La razón por la que PurN, se encuentra menos conservada en las arqueas y ausente en los eucariontes (a comparación de purD y purM) podría explicarse porque la reacción que en su caso realizaría purN en ambos grupos de organismos, es realizada por diferentes enzimas.

Muchas arqueas, por ejemplo, como se menciona en el artículo “Purine biosynthesis in arqueas”, [9]; no usan la enzima purN, hacen uso de purT, su enzima análoga. Sin embargo, curiosamente hay arqueas en las que no se encuentra ninguna de estas dos enzimas, además de que la distribución de purN y purT dentro de las arqueas no parece estar relacionada con su filogenia [9]. En cuanto a la ausencia de purN en los eucariontes podría explicarse porque las enzimas que llevan a cabo esta reacción en dichos organismos son en su lugar: purN (K00601), Gart y purNH. Lo que probablemente significa que la acción que realiza purN dentro de la ruta de biosíntesis de purinas, es algo que a lo largo de la evolución se ha realizado de diferentes maneras.

En cuanto a un segundo grupo de enzimas en general menos conservadas que las del grupo anterior formado por purK/purE clase I y purC, se observó que la enzima purC está muy conservada en las bacterias, en menor grado en la arqueas y estuvo prácticamente ausente en los eucariontes, presentando dos homólogos que fueron fusiones (PAICS y purC/E). En cambio, la enzima purK clase I se encontró en general poco conservada, siendo aparentemente la enzima con menor distribución filogenética de todas las enzimas de la ruta de purinas ya que estuvo prácticamente ausente en los eucariontes, en la mayoría de las arqueas y en muchas bacterias, y

presento 2 homólogos que fueron fusiones de purK con otras enzimas (purE clase II y purC/E) y solo un homólogo no fusionado (purT).

Finalmente se encontró que el grado de conservación y distribución de la última enzima de la ruta de la RBNP, purH, resulto, en primer lugar, mayor en las bacterias, al igual que otras enzimas de la ruta de purinas, en segundo lugar, dicha enzima se encontró más conservada en los eucariontes a diferencia de la mayor parte de las enzimas de la ruta de purinas. Aunque muchas de las enzimas que presentaron los eucariontes no fueron purH, sino proteínas hipotéticas con el mismo KO y EC que esta enzima. Por último, la enzima purH se encontró muy poco conservada en el grupo de las arqueas, hallándose únicamente en 9 de estos organismos. Además, se vio que esta enzima no presentó ninguna secuencia homóloga.

En conclusión, la mayor parte de las enzimas de la ruta de biosíntesis de purinas se encontraron más conservadas en el grupo de las bacterias que en los otros dominios, y aparentemente las enzimas de esta ruta que no se encuentran fusionadas, están más conservadas que las que si lo están. Por último, podemos decir que la enzima más conservada y distribuida de la RBNP fue prsA, mientras que purK fue la que menos lo estuvo. Por su parte purN fue la enzima con mayor divergencia y número de enzimas homólogas.

#### **4.5.2 Enzimas fusionadas (importancia e implicaciones evolutivas de su distribución filogenética)**

Entre los homólogos encontrados de las enzimas de la ruta de purinas, pudimos ver con sorpresa, que una gran parte, no correspondieron a enzimas duplicadas, sino enzimas fusionadas, que se identificaron especialmente en los eucariontes. Este fenómeno debe de ser analizado en estudios posteriores.

Un primer grupo de enzimas fusionadas encontrado lo forman la enzima ADE 5 (KO 11788, EC:6.3.4.13 6.3.3.1) que corresponde a la enzima purD fusionada con la enzima purM (ver resultados purD y purM, redes chicas) y la enzima GART (KO 11787,

EC:6.3.4.13 2.1.2.2 6.3.3.1) en la que purD, purN y purM están fusionadas. (Ver resultados de redes purD, purN y purM).

En cuanto a los resultados que se obtuvieron sobre la presencia de la enzima ADE 5 en las redes chicas de purD y purM, lo que se encontró es que esta enzima está presente únicamente en hongos (2 ascomycetes) y en 2 protistas. Este hallazgo coincide con lo que nos muestra el artículo “*Structural biology of Purine Biosynthesis*” [42]; donde ya se había identificado esta enzima en los hongos y las levaduras aunque en el artículo además se encontró en los reptiles y en este trabajo se encontró en *Dyctiostelium discoideum* y en *T.pseudonana*.

Con respecto a la enzima GART sabemos que está ya se había identificado en humanos, otros mamíferos y en drosófila [42]. Sin embargo, en mis resultados (purD, purN, purM, redes chicas) esta enzima se encontró además en otro vertebrado no mamífero, el pez cebra; también en los animales invertebrados incluidos en este trabajo: 1 placozooario, 1 cnidario, 1 ascidia, 1 lanceolado, 1 equinodermo, 1 nematodo y 1 artrópodo (excepto en 1 :una planaria) ; y en los coanoflagelados.

Este resultado sugiere que probablemente todos los animales heredaron la enzima GART a partir los coanoflagelados, en los que posiblemente surgió esta enzima fusionada.

Por último, me parece interesante que la enzima GART y la enzima ADE5 a pesar de tener componentes comunes tengan distintas distribuciones filogenéticas entre los mismos eucariontes, lo cual significa que posiblemente hay en general diferentes estrategias evolutivas para hacer los pasos que realizan dichas enzimas dentro de la ruta de biosíntesis de purinas y este fenómeno podría no ser exclusivo para estas enzimas.

Otro grupo de enzimas fusionadas son las enzimas purK y purC, que presentaron una enzima homóloga en común y 2 enzimas diferentes pero relacionadas, que comparten el mismo número EC y se describen a continuación:

a) Porphoribosylaminoimidazole carboxylase o purE clase II (K11808)

Esta es una enzima en la que se encuentran fusionadas las enzimas purE y purK en una sola y es homóloga a purK. En los resultados correspondientes a la red chica de purK, se encontró que la enzima purE clase II solo está presente en eucariontes (11), como se había descrito anteriormente pero a diferencia de lo reportado, de que las levaduras como las plantas tenían las enzimas purE clase I y purK fusionadas, en lugar de la purE clase II (reportada solo en vertebrados) [42]: esta enzima fue encontrada en 2 ascomycetes y 6 plantas (diferentes tipos). En tres de estos últimos organismos se vieron enzimas similares con el mismo número K y EC que el de la enzima purE clase II (ver resultados purK red chica).

b) Phosphoribosylaminoimidazole carboxylase/phosphoribosylaminoimidazole-succinocarboxamide synthase (PAICS)

Esta fusión está dada entre la purE clase II y la enzima purC y aparece en los resultados de purC. Los organismos que presentaron la enzima PAICS fueron 2 vertebrados: el ratón y el pez cebra. El resto de los animales (8 invertebrados) además de un coanoflagelado presentaron enzimas con los mismos números K y EC que la enzima PAICS, lo que indica que tanto esta última proteína como el resto de las enzimas con los mismos números K y EC podría ser exclusiva del grupo de los animales y pudo haber sido heredada a estos a través de los coanoflagelados purE y purK.

En conclusión, a pesar de que las enzimas purE clase II y PAICS son el resultado de la fusión de enzimas comunes, estas no aparecen en mis resultados como enzimas claramente homólogas. Finalmente resulta interesante que dichas enzimas, que realizan la misma función se encuentren en mis resultados distribuidos en diferentes tipos de organismos.

c) purC/E; phosphoribosylaminoimidazole carboxylase.

Esta enzima resulta de una fusión de la enzima purC con la enzima purE clase II. Únicamente fue encontrada en el protista *Dictyostelium discoideum*, pero curiosamente esta enzima se encontró en los resultados tanto de la enzima purC como en los de la purK. Por último, podría ser que las enzimas PAICS, purE clase II y purEC sean homólogas entre sí, lo cual no pudo ser detectado en este análisis, debido a la resolución de la metodología utilizada.

Llama la atención que en la ruta de biosíntesis de purinas se encuentren especialmente varias enzimas fusionadas. Lo que se ha sugerido sobre estas fusiones es que podrían ser el resultado de procesos de adaptación, adquiridos por los eucariontes para evitar la rápida degradación de ciertos metabolitos inestables [42]. Sin embargo también se encontró que mientras que algunos de los pasos en la RBNP son catalizados por una sola enzima individual en todos los organismos, otros pasos se caracterizan porque son catalizados por más de una enzima diferente y coincide Sin embargo también se encontró que mientras que algunos de los pasos en la RBNP son catalizados por una sola enzima individual en todos los organismos, otros pasos se caracterizan porque son catalizados por más de una enzima diferente (cada una, en diferentes grupos de organismos) y coincide con que estos últimos son especialmente los que catalizan las enzimas fusionadas de la ruta de purinas. Un ejemplo es el caso del paso que realiza la enzima PurN que puede ser realizado por diferentes enzimas como purN (K00601), Gart, purNH y purT. Y el caso de purM que también es realizado por ADE5 y Gart. Este fenómeno puede estar relacionado con la función y la flexibilidad de la reacción de dichos pasos, y no solo con lo propuesto por Zhang, Y., Morar, M., [42]. Sin embargo, es necesario realizar otro tipo de análisis para verificar dichas hipótesis.

Por último como se vio en los resultados, solo ciertas enzimas de la RBNP se puedan fusionar con otras muy específicas, como sucede con purM que solo se encuentra fusionada con purD y purN, mas no con cualquier otra enzima de la ruta. Esto concuerda con lo que se ha dicho anteriormente sobre que las fusiones de enzimas característicamente están compuestas de proteínas que funcionan de manera concertada como sucede, por ejemplo, con las enzimas que catalizan pasos sucesivos dentro de las rutas metabólicas [52]. Y también coincide con la hipótesis de que una fusión de enzimas es fijada durante su proceso evolutivo solamente cuando estas enzimas pueden cooperar funcionalmente en conjunto [52].

Finalmente se ha visto que las enzimas fusionadas muestran frecuentemente una distribución filogenética dispersa [52]; tal y como se puede constatar en mis resultados.

### 4.5.3 Duplicaciones

A continuación, se discute la importancia de las duplicaciones encontradas de cada enzima de la ruta de la biosíntesis de purinas, correspondientes a los resultados del primer análisis realizado (redes chicas).

#### **purF**

a) La asparagine synthase (AsnB) fue hallada únicamente en la Euriarqueota *thermococcus kodakarensis*. Sin embargo, esta enzima está presente en una mayor cantidad de organismos en la base de datos KEGG, por lo que debe tener una distribución más amplia. En este estudio no pudo ser identificada en más organismos debido a que los parámetros de búsqueda fueron muy estrictos.

b) La glucosamine-fructose-6-phosphate aminotransferase: el homólogo principal de purF, una de las duplicaciones más interesantes de este trabajo, que al encontrarse en todas las arqueas, en la mayoría de las bacterias y solamente en 9 eucariontes (ver resultados de la red chica de purF) se podría considerar como una duplicación muy antigua que posiblemente pudo haber existido en el último ancestro común. Dentro de estos mismos resultados fue curioso encontrar que la mayor parte de los eucariontes que no presentaron la enzima glmS (todos los animales invertebrados, las plantas, un hongo y un parabasárido) poseen proteínas hipotéticas equivalentes a esta (con el mismo número KO, ver resultados purF red chica) que probablemente tengan la misma función.

#### **purD, purM**

Otras dos enzimas que presentaron muy pocas duplicaciones no hipotéticas y solo en un genoma cada una fueron: purD una (carB) y purM otra (HypE). Es posible que estas duplicaciones se encuentren distribuidas en una mayor cantidad de organismos, como lo muestra la base de datos Kegg.

#### **purK, purC**

Dos enzimas contiguas que al igual que las anteriores se distinguieron por presentar muy pocas secuencias homólogas (no fusionadas) con funciones diferentes fueron las enzimas purK y purC. La enzima 5-(carboxamino) imidazole ribonucleotide synthetase (purK) presentó 2 enzimas homólogas en total, una enzima homóloga individual y una que fue encontrada en 20 de organismos.

a) ppp:PHYPADRAFT\_158224 hypothetical protein. Esta es una enzima hipotética muy parecida a purK que podría ser muy reciente o resultado de transporte horizontal pues los eucariontes no presentan purK.

B) purT: Esta enzima a pesar de ser el homólogo más importante que presento purK, tuvo una distribución poco clara y no se encontró en muchos organismos, hallándose únicamente en 12 bacterias y 6 arqueas (una de ellas en duda) y un eucarionte con una enzima hipotética, pero con el mismo KO que el de purT.

Esta homología puede explicarse porque tanto purK como purT tienen el dominio ATPgrasp aunque son las dos únicas enzimas en donde se observa este dominio en mis resultados, lo que difiere de lo planteado en el artículo escrito por Zhang y colaboradores (2008), donde se sugiere que todas las enzimas de la ruta *de novo* de biosíntesis de purinas que tiene este dominio podrían ser homólogas.

## **Pur B**

Una de las enzimas que se presentó un mayor número de secuencias homólogas hipotéticas, aunque con su mismo número KO fue la enzima purB. Dichas secuencias homólogas solo estuvieron presentes en eucariontes.

## **PurN**

A diferencia de las enzimas anteriores, purN resultó tener la mayor cantidad de enzimas homólogas (4). Un ejemplo de esto, son las 5 enzimas encontradas que son similares a la 10-formyltetrahydrofolate dehydrogenase /purT. Estas 5 enzimas se encuentran presentes en



animales (1 vertebrado y 4 invertebrados) y no en arqueas que son las que utilizan purT. El hecho de haber encontrado estas enzimas en un grupo específico de animales, sugiere que hubo una duplicación paráloga y que las proteínas resultantes tienen una función secundaria.

Otro ejemplo interesante de homólogos encontrados de purN fue la enzima phosphoribosylglycinamide formyltransferase (EC:2.1.2.2 K00601), que tiene el mismo nombre que purN y purT, pero el número K es diferente y solo se encuentra en eucariontes (6 plantas, 2 hongos y 1 protista) Esto podría ser explicado por una sustitución no homóloga o a una rápida divergencia en estructura primaria, que puede ser verificada por medio de la comparación de cristales de las proteínas.

Los últimos 2 homólogos que presento purN fueron la enzima fmt: methionyl-tRNA formyltransferase (EC:2.1.2.9, K00604) y la enzima purU: formyltetrahydrofolate deformylase (EC:3.5.1.10); K01433. Estas dos enzimas fueron, después de la glmS, las duplicaciones más interesantes que se encontraron en este trabajo por su distribución (aunque esta no fue clara) y a comparación de otros casos, estas 2 duplicaciones fueron encontradas en más de 25 organismos. La mayoría bacterias de diferentes tipos: la enzima fmt se presentó en 15 organismos y la purU en 18. La distribución de estas enzimas en los eucariontes fue la siguiente: La fmt se vio en 12 eucariontes (8 de ellos plantas) 2 stramenofilos y mus musculus y la purU en 10 eucariontes. En las bacterias se ve que esta última enzima tiene una distribución amplia, pues aparece en la mayor parte de los grupos de bacterias considerados. Los únicos grupos que son filogenéticamente cercanos en donde no está presente purU son: Elusimicrobia, Fusobacteria, Gammatimonadetes y Synergistetes. Al parecer esta enzima tiene algunas pérdidas como en algunas protobacterias y Bacteroidetes. Por otra parte al encontrarse únicamente en una arqueas, esto sea resultado de un evento de transporte horizontal.

Por último, purU tiene una distribución poco clara en los eucariontes (encontrándose en 3 stramenofilos, 6 plantas y un hongo) sin embargo aún en estos organismos no es clara su presencia pues en algunos de ellos se presenta en vez de la purU, una enzima hipotética con el mismo número KO y el mismo nombre que esta enzima.

## **5. Conclusiones**

### **5.1 Conclusiones del análisis de las redes chicas**

En la primera parte de esta investigación se llegó a varias conclusiones: Que la ruta de biosíntesis *de novo* de purinas está muy conservada, por ello se encontraron pocas duplicaciones. Sin embargo, hubo algunas diferencias entre la cantidad de homólogos que presentó cada una de las enzimas, lo que no necesariamente tuvo una relación con su grado de conservación filogenética.

La ruta de biosíntesis *de novo* de purinas, además de haber evolucionado por el modelo de Patchwork, ha presentado otros mecanismos de evolución: mediante duplicaciones de enzimas que no han presentado divergencias y mediante enzimas que se han fusionado.

### **5.2 Conclusiones del análisis de las redes grandes**

En la segunda parte de la investigación (redes grandes) la base de datos del Kegg de 2011 resulto ser útil para encontrar similitudes entre las secuencias de enzimas, tanto a nivel de dominios como en una menor escala, por lo que hubiera sido útil actualizarla sin cambiar sus parámetros de comparación, sobre todo pensando en trabajos futuros con redes de similitud de secuencias.

Al trabajar con las redes de similitud, descubrimos un enfoque diferente sobre la manera de estudiar los procesos evolutivos de las enzimas. Pudimos ver que se podían encontrar las relaciones de similitud (y su grado) entre las secuencias de diferentes grupos de enzimas, observando cada una de estas enzimas en cada una de sus partes divisibles y no como unidades completas.

En trabajos posteriores, se podrían estudiar mejor las características topológicas de las redes de similitud de secuencias, y conocer a profundidad su significado evolutivo, así como saber si en estas redes, al igual que en las redes de interacción molecular, se cumple una relación entre conectividad y antigüedad. En si esta investigación abre nuevas preguntas sobre lo que podemos conocer de la evolución de las enzimas a partir de las redes de similitud de secuencias

Por último, las redes mencionadas sirvieron de manera efectiva para visualizar e identificar a nivel grupal las relaciones de similitud de la estructura primaria de las enzimas de la RBNP a nivel de dominio con enzimas de diferentes rutas metabólicas.

### **5.3 Conclusiones finales**

La evolución por duplicación y divergencia de las enzimas podría ser más importante de lo que se había considerado, no solo por su frecuencia sino porque además se vio que se da comúnmente a otros niveles, además del de las enzimas completas, por ejemplo, a nivel de dominio(s) y a una menor escala de este.

El modelo de evolución de Patchwork no solo se da a nivel de enzimas completas, sino que también se da a nivel de dominio, como lo afirma Sara Teichmann (2000). Sin embargo, este supuesto se tendría que poner a revisión, dado que la parte teórica del modelo Patchwork está hecha en base a duplicaciones de las enzimas completas, y no de subpartes de estas, por lo que dicha afirmación tendría que ser revisada y quizá replanteada.

# Bibliografía

- [1] Almas E, Vázquez A and László Barabási. SCALE-FREE NETWORKS IN BIOLOGY.
- [2] Anat Kreimer, Elhanan Borenstein, Uri Gophna, Eytan Ruppin, 2007, The evolution of modularity in bacterial metabolic networks. PNAS vol. 105 no. 19 6976-6981
- [3] Barabasi A. L. & Réka A. 1999. Emergence of scaling on random networks. Science 286, 509-512
- [4] Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., & Danchin, A. (2009). From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. Microbiology, 155(6), 1758-1775. <http://bsubcyc.org>
- [5] Barve A., Wagner A. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature. 500, 203-206.
- [6] Becerra, A. 1997 Evolución Molecular de las Rutas de Salvamento de Nucleótidos de Purinas. Facultad de Ciencias. UNAM.
- [7] Becerra, A., Lazcano, A. 1998. The Role of Gene Duplication in the Evolution of Purine Nucleotide Salvage Pathways. Origin Life Evol. Biosph. Vol 28, Issue 4, pp 539-553
- [8] Berens, R. L., Krug, E. C., & MARR, J. J. 1995. Purine and Pyrimidine Metabolism. Biochemistry and molecular biology of parasites, 89.
- [9] Brown, A. M., Hoopes, S. L., White, R. H., & Sarisky, C. A. 2011. Purine biosynthesis in archaea: variations on a theme. Biol Direct, 6(1), 63.

- [10] Delaye, L., Becerra, A., & Lazcano, A. (2005). The last common ancestor: What's in a name?. *Origins of Life and Evolution of Biospheres*, 35(6), 537-554.
- [11] Díaz-Mejía JJ, Ernesto Pérez-Rueda Ernesto and Segovia Lorenzo. 2007 A network perspective on the evolution of metabolism by gene duplication. *Genome Biology*, 8(2), R26
- [12] Emiliani, G., Fondi, M., LiÅš, P., & Fani, R. 2010. Evolution of metabolic pathways and evolution of genomes (pp. 37-68). Springer Netherlands.
- [13] Fani, R., & Fondi, M. 2009. Origin and evolution of metabolic pathways. *Physics of life reviews*, 6(1), 23-52.
- [14] Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99 (12), 7821-7826.
- [15] Granick, S. 1957. Speculations on the origins and evolution of photosynthesis. *Annals of the New York Academy of Sciences*, 69(2), 292-308.
- [16] Henderson JF, Paterson AR. 1973. *Nucleotides and metabolism: an introduction* Ed. Academic Press.
- [17] Hernández-Montes, G., Díaz-Mejía, J. J., Pérez-Rueda, E., & Segovia, L. 2008. The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution. *Genome Biol*, 9(6), R95.
- [18] Horowitz, N. H. 1945. On the evolution of biochemical syntheses. *Proceedings of the National Academy of Sciences of the United States of America*, 31 (6), 153.
- [19] Jensen, R. A.: 1976, *Annu. Rev. Microbiol.* 30, 409.
- [20] Kappock TJ, Ealick SE, Stubbe J. 2000 Modular evolution of the purine biosynthetic pathway. *Curr Opin Chem Biol*. Vol. 4(5):567-72.
- [21] Kepes Francois, 2007 *Biological networks*. World Scientific Publishing Co. Pte. Ltd. 3
- [22] Keseler et al., *Nuc Acids Res*, 39:D583-590 2011.

- [23] Lazcano, A., & Miller, S. L. 1999. On the origin of metabolic pathways. *Journal of molecular evolution*, 49(4), 424-431.
- [24] Matthew J Betts, Robert B Russell, 2009. A more structured metabolome. *Nature Structural & Molecular Biology* 16, 1125-1126
- [25] McMurry, J., & Begley, T. P. 2005. *The organic chemistry of biological pathways*. Roberts and Company Publishers.
- [26] Nannemann DP, Kaufmann KW, Meiler J, Bachmann BO. 2010. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng Des Sel.* 23(8):607-16
- [27] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL: Hierarchical organization of modularity in metabolic networks. *Science* 2002, 297:1551-1555.
- [28] Rison SC, Thornton JM. 2002. Pathway evolution, structurally speaking. *Curr Opin Struct Biol.* 12(3):374-82.
- [29] Robert H. White, 1997, Purine Biosynthesis in the Domain Archaea without Folates or Modified Folates, *Journal of Bacteriology*. Vol. 179, No. 10 p. 3374-3377
- [30] Roy, S. 1999. Multifunctional enzymes and evolution of biosynthetic pathways: Retro-Revolution by jumps. *Proteins: Structure, Function, and Bioinformatics*, 37(2), 303-309.
- [31] Schmidt, S., Sunyaev, S., Bork, P., & Dandekar, T. 2003. Metabolites: a helping hand for pathway evolution?. *Trends in biochemical sciences*, 28(6), 336-341.
- [32] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003 Nov; 13(11):2498-504
- [33] Solé 2009 *Redes Complejas Del genoma a Internet*. Tusquets editores. España

- [34] Tatusov, R. L., Mushegian, A. R., Bork, P., Brown, N. P., Hayes, M. S., Borodovsky, M., Rudd, K. E. and Koonin, E. V.: 1996, Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli. Current Biol. 6, 279.
- [35] Teichmann, S. A., Rison, S. C., Thornton, J. M., Riley, M., Gough, J., & Chothia, C. (2001). The evolution and structural anatomy of the small molecule metabolic pathways in Escherichia coli. Journal of molecular biology, 311(4), 693-708.
- [36] Tham, T. N., Ferris, S., Kovacic, R., Montagnier, L. and Blanchard, A., 1993 Identification of Mycoplasma pirum genes involved in the salvage pathways for nucleosides. J. Bacteriol. 175, 5281
- [37] Vázquez Hernández, C. D. (2011). Reconstrucción y estudio topológico de las redes de regulación y del metabolismo central de Bacillus subtilis. Universidad Nacional Autónoma de México.
- [38] Voet, D., & Voet, J. G. 2006. Bioquímica. Ed. Médica Panamericana.
- [39] Ycas, M.: 1974, J. Theoret. Biol. 44, 145.
- [40] Vázquez Salazar Alberto. 2015. Evolución molecular de la biosíntesis de histidina y otros compuestos imidazólicos. Tesis de Licenciatura. Facultad de Ciencias. UNAM
- [41] Ullrich, A., Rohrschneider, M., Scheuermann, G., Stadler, P. F., & Flamm, C. 2011. In silico evolution of early metabolism. Artificial life, 17(2), 87-108.
- [42] Zhang, Y., Morar, M., & Ealick, S. E. 2008. Structural biology of the purine biosynthetic pathway. Cellular and molecular life sciences, 65(23), 3699-3724.
- [43] <http://themedicalbiochemistrypage.org/es/nucleotide-metabolism-sp.php>
- [44] <http://library.med.utah.edu/NetBiochem/pupyr/pp.htm>
- [45] <http://www.genome.jp/kegg/docs/plea.html>, datos tomados en el 2011.
- [46] <http://pfam.xfam.org>

- [47] Programa en PERL desarrollado por Ricardo Hernández, Laboratorio de Origen de la Vida, Facultad de Ciencias UNAM
- [48] Programa en PERL desarrollado por Luis Delaye y Arturo Becerra, Laboratorio de Origen de la Vida, Facultad de Ciencias.
- [49] Susumu Ohno (1970). Evolution by gene duplication. Springer-Verlag. ISBN 0-04-575015-7)
- [50] Patricia Babbit. 2009. [efi.igb.illinois.edu/efi-est/tutorial\\_references.php](http://efi.igb.illinois.edu/efi-est/tutorial_references.php)
- [51] Atkinson, Holly J., John H. Morris. 2009. Using Sequence Similarity Networks for Visualization of Relationship Across Diverse Protein Superfamilies, PlosOne. <https://doi.org/10.1371/journal.pone.0000434>.
- [52] Itai Yanai,<sup>1</sup> Yuri I Wolf,<sup>2</sup> Genome Biol. 2002; 3(5): research0024.1 research0024.13. Published online 2002 Apr 26. Evolution of gene fusions: horizontal transfer versus independent even



