

UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

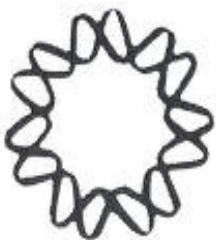
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE BIOTECNOLOGÍA

SUBDIVISIÓN DE GRUPOS COG_s CON
PRESENCIA DE PARÁLOGOS EN SUBFAMILIAS
DE PROTEÍNAS

T E S I S
QUE PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS BIOQUÍMICAS
P R E S E N T A

Mario Alberto Martínez Núñez

Director de Tesis
Dr. Enrique Merino Pérez



Cuernavaca, Morelos, México. 2007.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

El presente trabajo se realizó bajo la asesoría del Dr. Enrique Merino Pérez en el Departamento de Microbiología Molecular del Instituto de Biotecnología de la Universidad Nacional Autónoma de México, en Cuernavaca, Morelos. México.

El comité tutorial estuvo conformado por:

Dr. Lorenzo Segovia (IBt, UNAM)

Dr. Alejandro Garcarrubio (IBt, UNAM)

Durante el desarrollo del presente proyecto el autor recibió el apoyo del Consejo Nacional de Ciencia y Tecnología (CONACYT) como becario.

El análisis computacional se realizó en el Cluster de Computadores del Instituto de Biotecnología-UNAM, donado por el Macroproyecto de Tecnologías de la Información y la Computación-UNAM.

El proyecto realizado recibió financiamiento de donativos otorgados por CONACYT y Pappit al Dr. Enrique Merino.

Agradecimientos

Al Dr. Enrique Merino Pérez, por la enseñanza de vida, los bailes y los infaltables chistes (**¡No, aún no!**). ¡Chido!

Al comité tutorial por su tiempo, sus comentarios, los cuales fueron dando forma a este proyecto, y los sustos semestrales.

Dr. Lorenzo Segovia

Dr. Alejandro Garcarrubio

Al jurado por su tiempo y sus comentarios.

Dr. Enrique Morett Sánchez

Dr. Humberto Flores Soto

Dr. Miguel Ángel Ramírez

Dr. Ernesto Pérez Rueda

A l@s integrantes del laboratorio de la Dr. Guadalupe Espín por la tolerancia a mis actos, y su amistad.

A l@s integrantes del Grupo de Genómica Computacional del Dr. Enrique Merino por su amistad, tiempo y sus chistes en los seminarios.

La escritura de esta Tesis ha sido posible gracias al apoyo de una multitud de personas.

Algunas jamás en su vida han escuchado o escucharán hablar del autor de este trabajo, mucho menos conocerlo. Como por ejemplo Robert Taylor y a J. C. R. Licklider, o bien, Richard Stallman, Linus Torvalds e Ian Murdock, o también los participantes de las comunidades y foros del Internet. A todos ellos mi agradecimiento, todas esas noches y todos esos días han sido de aprendizaje.

La comunidad de la blogósfera, y en especial al consejo editorial del No-mames.blogspot.com, contribuyo con muy buenos ratos de humor, asombro, y demás sentimientos que te hacen exclamar ¡No mames!. Bob Esponja Mexicano, ¡No mames!. Para ell@s mi agradecimiento por tantas alegrías.

Los que alguna vez coincidimos en La Azotea o en las pozas de Amatlán (Ivette, Brenda, Oviedo, Iskra, Paty, Maga, Edna, Sergio, etc.) fueron de mucha valía para que no decayeran los animos. Así como también l@s que gustan de caminar por las montañas, ya sea a Chalma o a Zempoala, por los paisajes y las aventuras, los cuales dieron muchas satisfacciones con tantos retos compartidos.

A Donald Knuth (el cual no me conoce) y Felipe Contreras mi agradecimiento por \TeX y \LaTeX , así como a José Luis Gutiérrez Sánchez, porque lo que aprendemos en diferentes lugares y momentos se conecta de las formas más extrañas con nuestro presente, en nuestras vidas no-lineales.

Los que se quedaron y a los que nos dejaron en el camino. Lo aprendido con ellos siempre enriqueció y enriquecerá lo que hacemos y hagamos.

Para todos ustedes, vaya pues esta página y este trabajo.

Para Nelly y Andrea. Porque siempre han estado en mis pensamientos, y yo en el de ustedes.

Índice

1. Introducción	1
1.1. Base de Datos COG (Cluster of Orthologous Groups)	6
1.2. Algoritmos de agrupamiento (“clustering”)	9
2. Hipótesis	12
3. Objetivos	12
4. Metodología	13
4.1. Identificación de hits bidireccionales	16
4.2. Aplicación del criterio de <i>Tatusov et al.</i>	16
4.3. Eliminación de genomas redundantes	20
4.4. Agrupamiento al interior de los COGs	22
4.5. Selección de COGs divididos por MCL	24
4.6. Matrices de peso-posición	26
5. Resultados: Significado biológico de los agrupamientos	29
5.1. Comparación contra la base de datos Pfam	29
5.2. Comparación contra la base de datos PROSITE	32
5.3. Documentación manual de COGs	34
5.3.1. COG0095. Lipoate-protein ligase A	35
5.3.2. COG0558. Phosphatidylglycerophosphate synthase PgsA	40
5.3.3. COG1393. Arsenate reductase and related proteins, glutaredoxin family	43
5.3.4. COG0147. Anthranilate/para-aminobenzoate synthases component I	49
6. Discusión	54
7. Conclusión	61
8. Perspectivas	62

9. Apendice 1: Cuadro COGs eliminados en 1er. filtro	63
10.Apendice 2: Cuadro COGs eliminados en 2do. filtro	98
11.Apendice 3: Cuadro COGs eliminados en 3er. filtro	117
12.Apendice 4: Cuadro COGs finales	252
13.Apendice 5: Cuadro COGs anotados	278
14.Apendice 6: División de perfiles Pfam PF04715 y PF01066	289

Índice de figuras

1.	Tipos de homología entre genes	2
2.	Enfoques computacionales para la predicción de funciones	2
3.	Piedra Rosetta	4
4.	Vecindad de genes	4
5.	Operones	4
6.	Perfiles filogenéticos	5
7.	Mejor hit bidireccional (BBeT)	6
8.	COG mínimo	7
9.	COG1697-DNA topoisomerase VI	7
10.	COG2927-DNA polymerase III	8
11.	Clustering jerárquico y de partición	10
12.	Algoritmo MCL	11
13.	Relaciones de bidireccionalidad de los COGs	17
14.	COG2515 (1-Aminocyclopropane-1-carboxylate deaminase)	17
15.	COG1306 (Uncharacterized conserved protein)	18
16.	Porcentaje de parálogos en los COGS	19
17.	Eliminación de genomas redundantes	21
18.	Porcentaje de COGs con proteínas parálogas	22
19.	Distribución de grupos al interior de los COGs	24
20.	Selección de grupos con al menos 5 secuencias	25
21.	Selección de COGs con al menos dos grupos	26
22.	Diagrama de flujo para la obtención de los MOGs	28
23.	COGs finales con grupos cubiertos por Pfam	30
24.	Perfiles Pfam para los grupos del COG0262	31
25.	COGs finales con grupos cubiertos por PROSITE	33
26.	Firmas PROSITE para los grupos del COG1304	34

27.	Árbol evolutivo de LplA, LipB y BPL	36
28.	Alineación de las secuencias de BPL, LplA y LipB	37
29.	Ruta biosintética de la cardiolipina	40
30.	Enzimas que utilizan CDP-DAG	41
31.	Alineamiento de las CDP-alcohol fosfatidiltransferasas	42
32.	Transporte y reducción de arsenato en <i>E. coli</i>	43
33.	Clados de la arsenato reductasa	44
34.	Regulación negativa de Spx	45
35.	Secuencia de Spx	46
36.	Estructuras de Spx y ArsC	47
37.	Ruta del shikimato, folato y triptófano	49
38.	Ruta de bisíntesis del triptófano	50
39.	Ruta de bisíntesis del folato	51
40.	8 COGs con grupos anotados por Pfam y PROSITE	55
41.	Firma PROSITE y Perfil Pfam para el grupo 1 del COG0262	56
42.	COGs anotados en total por PFAM o PROSITE	57
43.	Diferencia de porcentaje de cobertura en los COGs	57
44.	Contexto genómico del grupo 1 del Pfam PF04715	290
45.	Contexto genómico del grupo 2 del Pfam PF04715	291
46.	Motivos de la familia PgsA encontrada en Pfam	292
47.	Motivos de la familia Pss encontrada en Pfam	293
48.	Motivos de la familia Pis encontrada en Pfam	294
49.	Motivos de la familia de fosfatidiltransferasas de eucariotes	295
50.	Motivos de la familia Pcs encontrada en Pfam	296

Índice de cuadros

1.	Genomas usados en el presente trabajo	13
2.	Número de COGs después del primer filtro	19
3.	Número de COGs sin y con parálogos	20
4.	Número de COGs después del segundo filtro	21
5.	Número de COGs sin y con parálogos. 2do filtro	21
6.	Valores de granularidad (I) probados	23
7.	Distribución de grupos formados por MCL	24
8.	Motivos de LplA en el COG0095	38
9.	Motivos de LplA en secuencias recuperadas por MAST	38
10.	Porcentaje de proteínas con los motivos de LplA	39
11.	Proteínas con los motivos de PgsA y Pis	41
12.	Proteínas recuperadas con los motivos de PgsA y Pis	43
13.	Proteínas con los motivos de ArsC y Spx	47
14.	Proteínas recuperadas con los motivos de ArsC y Spx	48
15.	Porcentaje de aparición de los motivos de ArsC y Spx	49
16.	Anotaciones para las secuencias del COG0147	52
17.	Anotaciones para las secuencias recuperadas del COG0147	52
18.	Comparación de las anotaciones del COG0147	53
19.	COGs eliminados en el 1er. filtro	63
20.	COGs eliminados en el 2do. filtro	98
21.	COGs eliminados en el 3er. filtro	117
22.	COGs finales	252
23.	COGs anotados	278

Resumen

La base de datos COG agrupa a las proteínas en familias de ortólogos. Sin embargo, algunos de los grupos COGs presentan proteínas parálogas en su interior, las cuales pueden dar origen a subfamilias funcionalmente diferentes a las ortólogas. La presencia de parálogos en los COGs puede provocar errores al ser utilizados en la anotación de proteínas de genomas recién secuenciados. En el presente trabajo se analizaron 584 COGs, que presentaron parálogos en su interior. A estos 584 COGs se les aplicó un algoritmo de agrupamiento basado en cadenas de Markov (MCL) para dividirlos en subfamilias funcionalmente diferentes. La documentación de las subfamilias encontradas en los COGs divididos por MCL se realizó de manera manual y comparándolos contra las bases de datos PROSITE y Pfam. Para el 6.33 % de estos 584 COGs, se encontraron firmas PROSITE específicas para cada subfamilia contenida en su interior, mientras que un 7.19 % de los COGs presentaron modelos Pfam específicos para las subfamilias contenidas en ellos. Estas firmas PROSITE o modelos Pfam diferencian a los grupos encontrados por MCL como subfamilias diferentes. Para 4 grupos COGs (COG0095, COG0558, COG0147, COG1393), se realizó un análisis detallado en base a la documentación publicada, y se encontraron evidencias bioquímicas que apoyan la división de estos COGs en subfamilias a su interior. Nuestro enfoque de división de familias de proteínas con presencia de parálogos fue probado exitosamente en Pfam, con lo que se demostró que nuestra metodología puede ser aplicada a otras bases de datos de proteínas homólogas.

1. Introducción

El incremento constante de datos provenientes de proyectos genómicos y proteómicos, ha favorecido el surgimiento de diversas herramientas para su análisis, tanto experimentales como computacionales (revisado en [25], [43], [47], [65]). Uno de los objetivos de dichas herramientas es la predicción de las funciones de las proteínas codificadas en los genomas de los organismos. Los métodos computacionales para la asignación de funciones, frecuentemente están basados en la suposición de que proteínas con secuencias similares (homólogas) llevan a cabo funciones moleculares similares [61]. Los genes (y sus proteínas) que descienden de un ancestro en común son llamados homólogos. La homología puede dividirse en al menos tres tipos: ortología, paralogía y xenología. Los genes ortólogos son aquellos que se encuentran en diferentes especies, y su separación se ha dado por un proceso de especiación [55] [22] (Figura 1). Los genes o proteínas ortólogos conservan por lo general su función después de ocurrida la especiación, lo que permite la predicción de las funciones codificadas en un genoma secuenciado mediante la transferencia de la información funcional de los ortólogos ya conocidos y caracterizados [79]. La paralogía es la duplicación de un gene (Figura 1) dentro de un mismo genoma, donde una de las copias puede modificar parcialmente su función de la original, como por ejemplo el reconocimiento de un sustrato diferente en una reacción enzimática. Mientras que la xenología es la transferencia horizontal de los genes entre organismos [22] (Figura 1).

Los métodos predictivos basados en la homología (Figura 2), infieren la función de una proteína comparando una secuencia de aminoácidos contra una base de datos de proteínas ya conocidas. Otro método, como el alineamiento de proteínas ortólogas, permite observar patrones de conservación de aminoácidos. Con estos patrones, es posible obtener motivos que capturen las características funcionales de la familia de proteínas, para después buscar los motivos en otras secuencias polipeptídicas que probablemente lleven a cabo la misma función. Los alineamientos múltiples de secuencias ortólogas son también utilizados para la construcción de perfiles (matrices de posición) o modelos ocultos de Markov (HMM), los cuales se emplean después en la detección de proteínas que pertenezcan a la misma familia de la cual se derivó el perfil o modelo de Markov [28].

Otros enfoques no recurren a la comparación directa de secuencias de proteínas o genes para la pre-

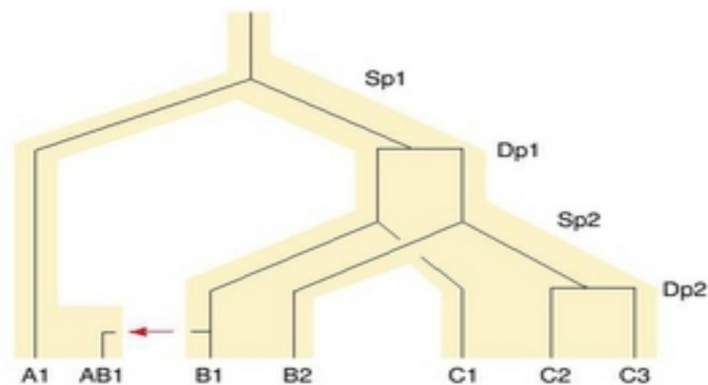


Figura 1: Tipos de homología que se establecen entre los genes por los procesos de especiación y duplicación. Los 2 eventos de especiación (Sp1 y Sp2) se encuentran en la bifurcación en forma de ‘Y’ invertida del árbol. Los 2 eventos de duplicación (Dp1 y Dp2) se muestran como barras horizontales. Dos genes cuyo ancestro común reside en la bifurcación ‘Y’ serán ortólogos, mientras que dos genes cuyo ancestro común reside en la barra horizontal serán parálogos. La flecha horizontal denota un evento de transferencia horizontal (xenología). Figura tomada de [22].

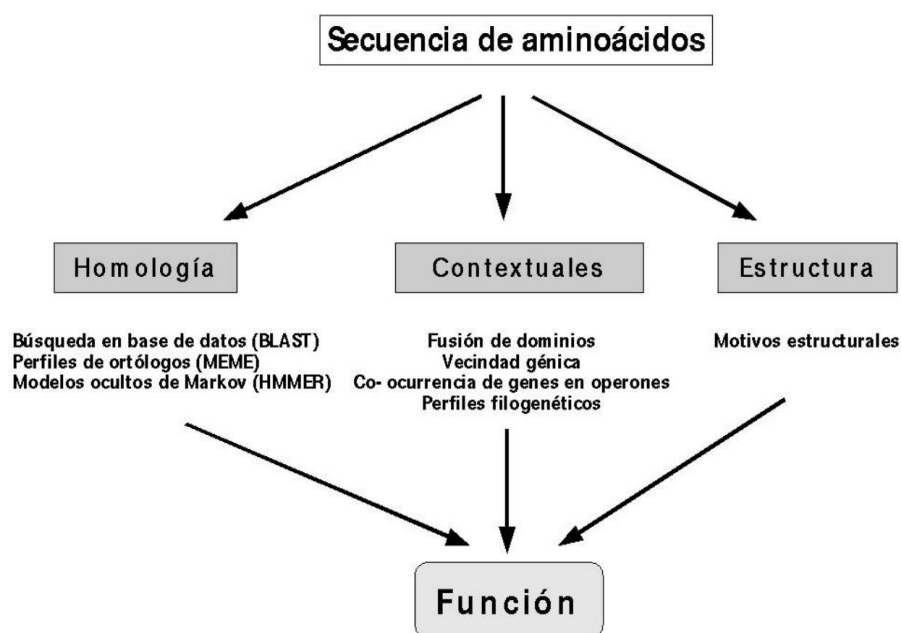


Figura 2: Diferentes enfoques computacionales que permiten la predicción de funciones en los genes o proteínas de organismos procariontes. En la parte izquierda se encuentran los métodos que utilizan la homología. En el centro los métodos no-homólogos o contextuales. En la parte derecha los métodos que utilizan la estructura de las proteínas. Figura modificada de [51].

dicción de funciones [24]. Estos métodos independientes de la homología, utilizan principalmente el contexto en el cual se encuentran los genes, por lo cual se han llamado métodos contextuales (Figura 2). Ejemplos de estos métodos son la Piedra Rosetta, la vecindad génica, la co-ocurrencia de genes en operones y los perfiles filogenéticos (**Caja 1**). El método de la Piedra Rosetta realiza asignaciones funcionales explorando la fusión de dominios que se dan entre las proteínas (Figura 3). Proteínas que interactúan físicamente o que están funcionalmente relacionadas en algunos organismos, y se expresan de manera separada, se encuentran fusionadas en una sola cadena polipeptídica en otros [61]. Un ejemplo son los genes que codifican para las proteínas GyrA y GyrB, las cuales conforman la DNA girasa en *Escherichia coli*, dichos genes se encuentran fusionados y expresados como una sola cadena polipeptídica (topoisomerasa II) en levaduras [51]. En el método de la vecindad génica, se utiliza el supuesto de que dos proteínas pueden estar funcionalmente relacionadas si sus genes se encuentran como vecinos en el genoma de diferentes organismos [24] (Figura 4). En procariotes los genes pueden ser transcritos en mRNA policistrónicos (operones), los cuales codifican para genes co-regulados y normalmente relacionados funcionalmente, como en alguna ruta bioquímica particular. Este arreglo de los genes permite la utilización del método de la co-ocurrencia de genes en operones, para la asignación de funciones de proteínas cuyos genes se encuentren en operones ya conocidos o caracterizados [17] (Figura 5). Los perfiles filogenéticos permiten una predicción funcional, basados en la hipótesis de que proteínas ligadas funcionalmente evolucionan de manera correlacionada, y que por lo tanto sus homólogos se presentan juntos en los genomas de otros organismos [66], [34] (Figura 6).

La utilización de motivos estructurales que puedan presentar las proteínas, es otro método utilizado en la asignación de funciones [6] (Figura 2). En este caso la búsqueda está orientada a fragmentos de secuencias que formen un motivo estructural que acompañe a la función, como por ejemplo, los sitios de unión a un cofactor, motivos estructurales empleados en la unión a DNA, o bien sitios de modificaciones postranscripcionales.

Las bases de datos usadas para la predicción de funciones de proteínas, son herramientas que agrupan a secuencias de aminoácidos homólogas en familias, como por ejemplo, Pfam, PROSITE, COGs [10],

Caja 1: Métodos Contextuales

Piedra Rosetta: La predicción de proteínas que interactúan, ya sea física o funcionalmente, se lleva a cabo identificando pares de proteínas que estén fusionadas en un organismo, pero separadas en otros. El valor del método se encuentran en la asignación funcional de uno de los dos componentes, para el cual se conoce poco o nada de su función, a partir del otro ya caracterizado [24], [51], [61].

Piedra Rosetta

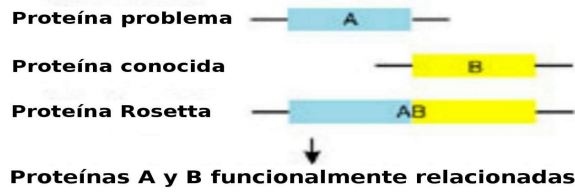


Figura 3: Piedra Rosetta. Figura tomada de [48].

Vecindad de genes: Debido a la alta tasa de recombinación de los genomas bacterianos, el orden de los genes en los organismos no es significativo. Sin embargo, cuando el orden es observado en diversos genomas, esto puede deberse a una presión de selección positiva. La conservación del orden puede mantenerse porque las proteínas codificadas por dichos genes interactúan física o funcionalmente, lo que permite realizar asignaciones funcionales a uno de los dos componentes del arreglo, conociendo la función del otro [18], [63].

Vecindad de genes

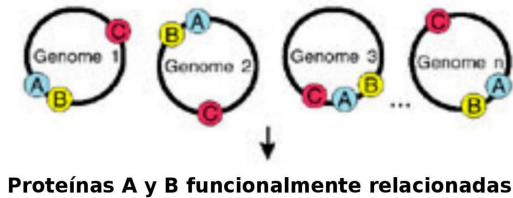


Figura 4: Vecindad de genes. Figura tomada de [48].

Co-ocurrencia de genes en operones: En procariontes, proteínas que interactúan funcionalmente son codificadas en mRNA policistrónicos, este agrupamiento de genes es conocido como operón. La predicción de funciones se realiza asociando aquellos genes cuyos productos polipeptídicos no están bien caracterizados, con la función que lleva a cabo el operón en donde se encuentran [17], [61].

Agrupamiento de genes



Figura 5: Operones. Figura tomada de [48].

Caja 1 (Continuación)

Perfiles filogenéticos: Genes funcionalmente relacionados evolucionan de manera correlacionada en los genomas de los organismos. Esto implica que los genes pueden coexistir en diversos organismos que lleven a cabo la función en la cual participan. Los perfiles filogenéticos exploran la presencia o ausencia en un grupo de organismos, de genes relacionados funcionalmente. En este método, cada gene es representado por un perfil filogenético, el cual es una cadena de n caracteres, donde n es el número de organismos considerados. La presencia o ausencia de los genes estudiados, es representada por un 1 o un 0, respectivamente. Los genes son agrupados dependiendo del perfil filogenético que presenten [66], [34].

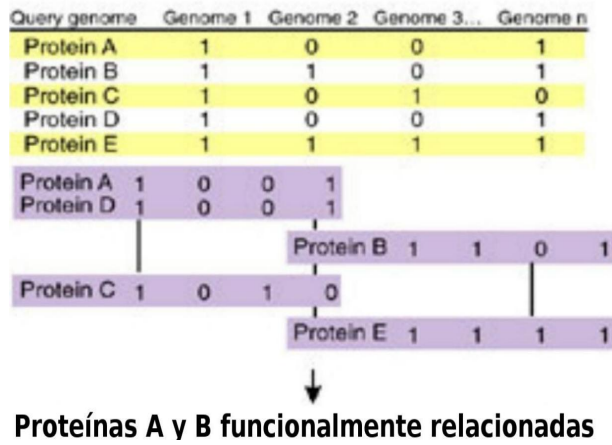
Perfil filogenético

Figura 6: Perfiles filogenéticos. Figura tomada de [48].

[33], [80]. La pertenencia de una proteína a una familia particular, usualmente indica la función general que lleva a cabo dicha proteína [1]. Proteínas homólogas que pertenecen a una familia de proteínas diversas, frecuentemente adquieren pequeñas diferencias funcionales. Por ejemplo, las enzimas tripsina, quimotripsina y elastasa son homólogas, pertenecientes a la familia de la serina proteasa. Estas enzimas catalizan la misma reacción, que es una hidrólisis de un enlace peptídico, sin embargo, reconocen y se unen a diferentes sustratos [3], por lo que representan diferentes subfamilias. Una subfamilia puede ser vista como un conjunto de proteínas con una línea particular de evolución dentro de una familia, las cuales han modificado su función original [1]. Cada subfamilia realiza una función específica, además de la función general de la familia a la cual pertenece. Las duplicaciones de genes son fuente de divergencia funcional, permitiendo a una de las copias adquirir una nueva función, mientras que el otro gene mantiene la función original [49], [84]. Este mecanismo de duplicación, el cual origina nuevas funciones en las proteínas, puede dar origen a nuevas subfamilias. El análisis de las familias de proteínas para la identificación de sus subfamilias, las cuales pueden divergir funcionalmente, permitirá una mejor

anotación funcional de secuencias de proteínas.

1.1. Base de Datos COG (Cluster of Orthologous Groups)

La base de datos de grupos de proteínas ortólogas COG (Cluster of Orthologous Groups), se utiliza principalmente para la asignación de funciones de proteínas en genomas [80]. El criterio bajo el cual se agrupan a las proteínas en esta base, es el de la ortología. Los ortólogos son definidos mediante la comparación de las secuencias de las proteínas o dominios provenientes de genomas secuenciados en su totalidad, estableciendo una relación de bidireccionalidad entre un par de proteínas llamada “mejor hit bidireccional” (Bidirectional Best Hit, BBeT). El BBeT es la similitud a nivel de secuencia obtenida por un par de proteínas “a” y “b” pertenecientes a los genomas “A” y “B” respectivamente, donde al comparar “a” contra todo el genoma de “B” el mejor alineamiento para “a” es el de la proteína “b” y viceversa [80] (Figura 7).

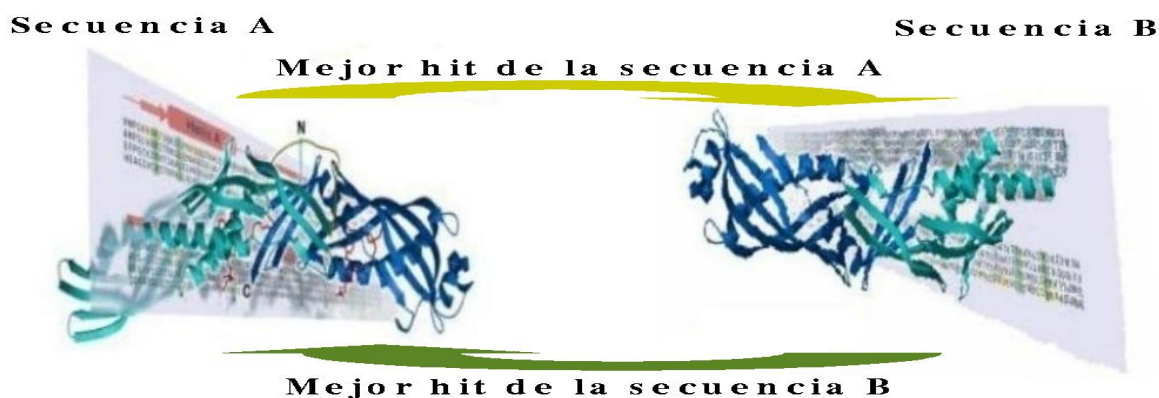


Figura 7: Relación de bidireccionalidad de hits a nivel de secuencia entre un par de proteínas provenientes de genomas diferentes (A y B) secuenciados en su totalidad. La flecha indica el mejor hit de cada proteína.

Una vez definidas las relaciones de ortología mediante los BBeT entre las secuencias, el criterio de construcción de los COGs establece que las proteínas contenidas en ellos deben presentar un triángulo formado de BBeT. Las proteínas que forman el triángulo deben de pertenecer a organismos de familias taxónomicas diferentes, como las proteínas KatG de *E. coli* (familia Enterobacteriaceae), sll1987 de *Synechocystis* sp. (orden Chroococcales) y YKR066c de *Saccharomyces cerevisiae* (familia Saccharomycetaceae), representadas en la figura 8 [80].

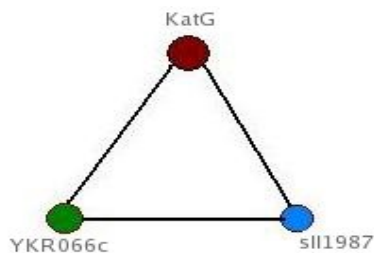


Figura 8: COG mínimo. Las líneas representan un BBeT y las proteínas en el triángulo KatG, sll1987 y YKR066c pertenecen a familias taxónomicas distintas (Enterobacteriaceae, Chroococcales –orden–, Saccharomycetaceae, respectivamente). Figura tomada de [80].

La unión de triángulos que comparten un lado en común da origen al crecimiento de los COGs. Por ejemplo, en el COG1697 (Figura 9) (DNA topoisomerase VI, subunit A) [86], el triángulo formado por las proteínas APE0703 de *Aeropyrum pernix* (familia Desulfurococcaceae), YHL022c de *Saccharomyces cerevisiae* (familia Saccharomycetaceae) y VNG0884G de *Halobacterium salinarum* (familia Halobacteriaceae) comparte un lado común —formado por APE0703 y VNG0884G— con el triángulo formado por las proteínas PAB2411 de *Pyrococcus abyssi* (familia Thermococcaceae), VNG0884G y APE0703.

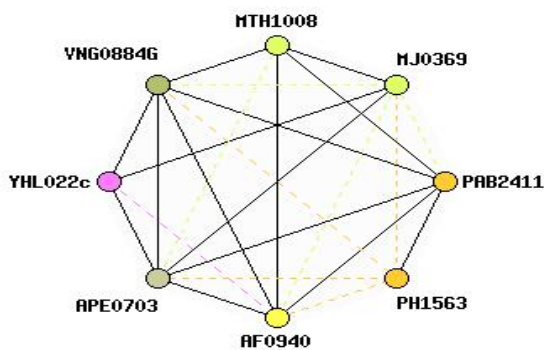


Figura 9: COG1697-DNA topoisomerase VI. Las líneas continuas representan BBeTs, las líneas punteadas representan hits unidireccionales. PH1563 *Pyrococcus horikoshii* (familia Thermococcaceae), AF0940 *Archaeoglobus fulgidus* (familia Archaeoglobaceae), MJ0369 *Methanocaldococcus jannaschii* (familia Methanocaldococcaceae), MTH1008 *Methanothermobacter thermautotrophicus* (familia Methanobacteriaceae), para las proteínas APE0703, YHL022c, VNG0884G y PAB2411 ver texto. Figura tomada de [86].

La utilización de los COGs para la asignación de funciones se ha llevado a cabo tanto en genomas procariotes como en genomas de eucariotes [42], [60]. Esta asignación funcional no se ha limitado solamente a genomas individuales. Recientemente ha sido utilizada en el análisis de metagenomas, como el del Mar de los Sargazos, comunidades de organismos en mar profundo y suelos agrícolas [81]. En muchos casos durante el agrupamiento de triángulos se incluyen proteínas parálogas, como las proteínas NMB1568 y NMA1757 de *Neisseria meningitidis*, pertenecientes al COG2927 (DNA polymerase III chi subunit) [87] (Figura 10). Debido a la existencia de parálogos en los COGs pueden existir proteínas que no formen un triángulo de BBeTs, y que además presenten relaciones de mejor hit (Best hit, BeT) unidireccional con las otras proteínas del COG [80], como la proteína NMB1568 (Figura 10). La presencia de proteínas parálogas en los COGs puede llevar a anotaciones erróneas o imprecisas, ya que los parálogos pueden dar origen subfamilias que varíen en función, con respecto de los ortólogos. Los errores en las anotaciones se incrementa conforme el número de parálogos en los COGs crece.

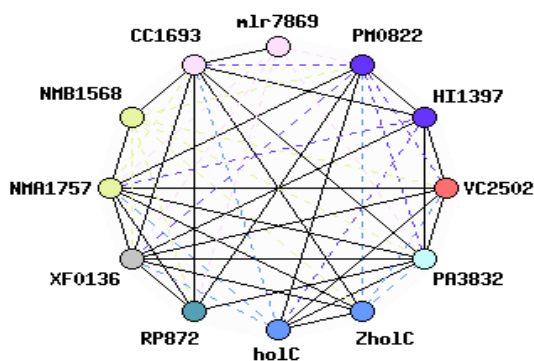


Figura 10: COG2927- DNA polymerase III chi subunit. Las proteínas NMB1568 y NMA1757 son parálogas, ambas pertenecientes al genoma de *Neisseria meningitidis*. Figura tomada de [87].

Una mejor agrupación de los COGs ya formados (con presencia de parálogos) mediante el uso de algoritmos de agrupamiento en subfamilias de proteínas diferentes, pudieran favorecer estrategias de análisis genómico, como las mencionadas anteriormente ([60], [42], [2]).

1.2. Algoritmos de agrupamiento (“clustering”)

Los algoritmos de agrupamiento (“clustering”) son herramientas que nos permiten clasificar un conjunto de datos de manera no-supervisada dentro de grupos (“clusters”) homogéneos. Los elementos que integran los clusters comparte uno o más atributos en común que los separa de los elementos de otros clusters distintos, de esta forma podemos explorar las diferencias y similitudes presentes en los datos. Los enfoques utilizados para la formación de los clusters se pueden dividir en *jerárquicos* y de *partición*. En el enfoque *jerárquico* los elementos a analizar se encuentran al inicio separados, en un primer paso se une a los elementos más cercanos para formar un cluster, posteriormente se busca el elemento más cercano al cluster recién formado y se unen formando un nuevo grupo. Este procedimiento se repite de manera iterativa hasta que se forma un árbol simple jerárquico [38], [36] (Figura 11.A). El procedimiento genera una clasificación de los grupos de manera anidada [67]. Ejemplos de los métodos *jerárquicos* son los que utilizan las medidas de distancia entre sus elementos para agrupar, como el enlace simple (single link) o el enlace completo (complete link). Los algoritmos *jerárquicos* generan un dendograma donde pueden visualizarse los grupos formados y las distancias entre sus elementos. El concepto de conectividad (*connectedness*) es aplicado en estos algoritmos, basados en la idea de que elementos vecinos comparten el mismo cluster [27].

Los algoritmos de *partición* inician con todos los elementos en un solo grupo, posteriormente van realizando divisiones que dan lugar a nuevos clusters hasta que se satisface cierto criterio (Figura 11.B). La partición genera grupos compactos (*compactness*) donde la variación dentro del cluster es mínima [27]. Ejemplos de algoritmos de partición son las K-medias, los algoritmos basados en teorías de gráficas, entre otros [36].

Los algoritmos de agrupamiento han sido usados extensivamente en el análisis de microarreglos para la agrupación de aquellos genes que comparten perfiles de expresión similares [20], [73], así como para el análisis de datos proteómicos de bacterias que infectan el conducto urinario [26], así como en la comparación de proteínas [39], [82]. El algoritmo MCL (Markov Cluster Algorithm) (<http://micans.org/~mcl/>) desarrollado por Van Dongen [83] aborda el problema de formación de clusters representando los datos como una gráfica en la cual se simulan caminatas aleatorias. La conexión entre los elementos

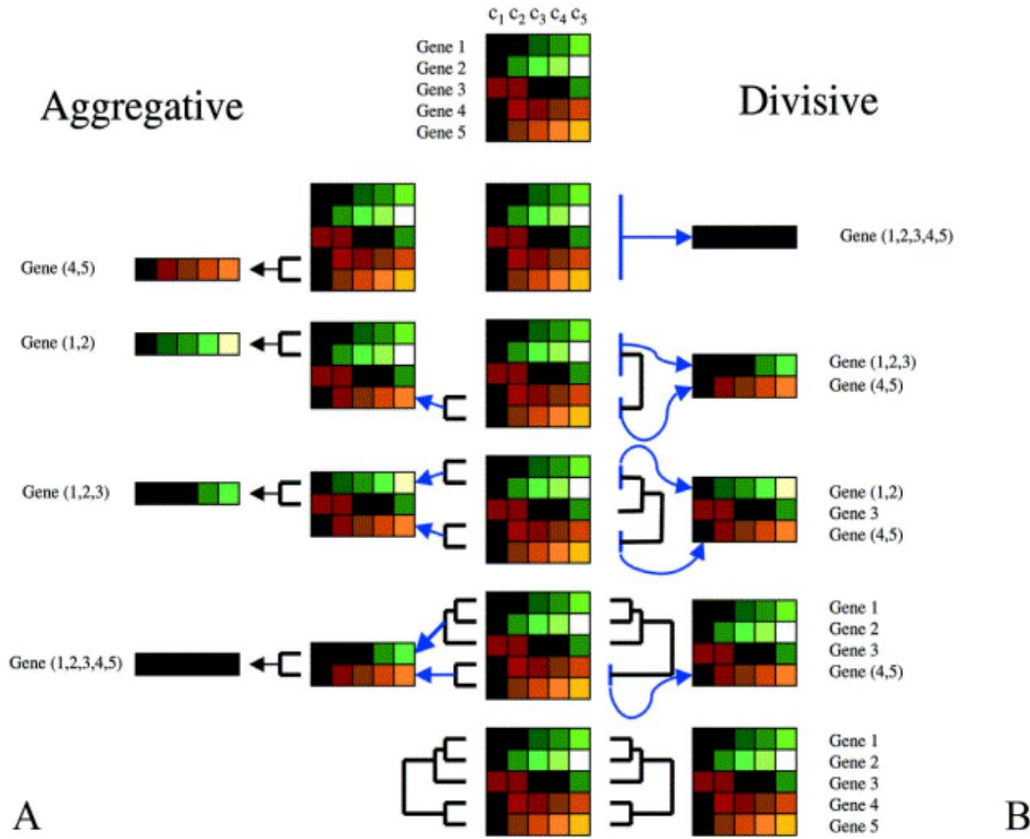


Figura 11: Ejemplificación de los procedimientos realizados en los algoritmos de agrupación jerárquica (A), y en los algoritmos de agrupación por partición (B). Figura tomada de [19].

pertenecientes a un mismo cluster es mayor a la que pudiera haber entre elementos pertenecientes a clusters distintos, por lo cual la probabilidad de caminar (saltar) entre clusters es menor a la probabilidad de caminar dentro de un mismo cluster, de esta forma MCL identifica los grupos que pudieran haber al interior de una gráfica (Figura 12).

El procedimiento por el cual MCL lleva a cabo la identificación de grupos es a través de 2 operaciones algebraicas sobre una matriz de Markov llamadas expansión e inflación. La expansión se corresponde con el cómputo de la probabilidad de una caminata aleatoria de distancia larga a través de la gráfica, mientras que la inflación fomenta y debilita las probabilidades de las rutas que hay en la gráfica. De manera iterativa se llevan a cabo la expansión e inflación hasta que no se detectan más cambios en la matriz de Markov. El parámetro principal dentro de MCL es la inflación (I), el cual determina la granularidad (tamaño) del cluster. Para valores bajos de I (1.2) se obtienen clusters con granularidad

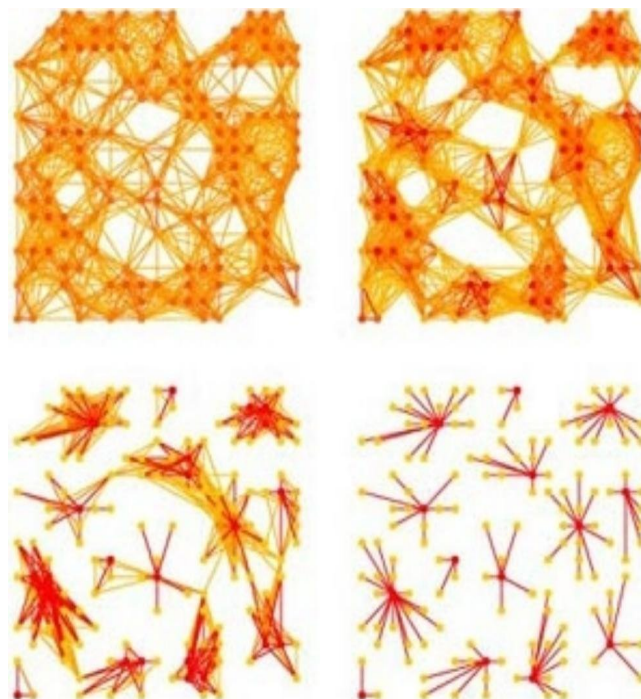


Figura 12: El algoritmo MCL simula un flujo dentro de una red, el cual permanece con mayor probabilidad dentro de los grupos que viajando de un grupo a otro. Figura tomada de [85].

gruesa, mientras que a valores altos de I (5.0), la granularidad de los clusters es fina. MCL ha sido utilizado en aplicaciones biológicas como la detección de familias de proteínas (Tribe-MCL) [21], la detección de proteínas ortólogas en eucariotes (OrthoMCL) [46], así como en el análisis microarreglos (GeneMCL) [45].

En el presente trabajo se utilizó el algoritmo MCL para la separación en grupos de subfamilias distintas de aquellos COGs que tuvieron presencia de proteínas parálogas en su interior. Se utilizaron las relaciones de BBeTs que se dan entre los miembros de los COGs para representar las conexiones de una red, y a las proteínas como nodos de la misma.

2. Hipótesis

- La exploración de las relaciones de bidireccionalidad que se dan entre las secuencias y la información contenida en ellas, pueden permitir la separación de proteínas parálogas dentro de los COGs que las presentan.
- La subdivisión de los COGs en grupos por metodologías de agrupamiento (“clustering”) podrá tener en algunos casos significado biológico detectable.

3. Objetivos

General

- Subdivisión de aquellos COGs que presenten proteínas parálogas en familias funcionales específicas.

Particulares

- Subdivisión de los grupos originales de COGs con presencia de proteínas parálogas en grupos biológicamente diferentes utilizando las relaciones de BBeT y el algoritmo de agrupamiento MCL.
- Verificación del significado biológico de los grupos generados por MCL mediante:
 - Búsqueda bibliográfica.
 - Búsqueda en bases de datos (*i.e.* Pfam, PROSITE).
- Asignación funcional mejorada a los subgrupos formados dentro de los COGs.
- Generación de perfiles (matrices peso-posición) para cada grupo encontrado por MCL.

4. Metodología

Para el desarrollo del presente trabajo se utilizaron los dominios definidos en la base de datos pública de grupos de proteínas ortólogas COG [80]. El número de COGs que conforman la base es de 4,873 donde se encuentran 138,458 proteínas provenientes de 66 genomas de organismos unicelulares secuenciados en su totalidad. A los 66 genomas tomados de los COGs se agregó para su análisis el genoma de *Salmonella typhi* (Cuadro 1).

Cuadro 1: Genomas usados en el presente trabajo. El genoma marcado con * no pertenece a la base de datos COG. Los organismos marcados con † son los que pasaron el filtro de redundancia de genomas (ver sección 4.3).

Genoma	Taxa	Familia
<i>Archaeoglobus fulgidus</i> †	Archaea	Archaeoglobaceae
<i>Halobacterium salinarum</i> † sp. NRC-1	Archaea	Halobacteriaceae
<i>Methanosarcina acetivorans</i> † str. C2A	Archaea	Methanosarcinaceae
<i>Methanothermobacter thermautotrophicus</i>	Archaea	Methanobacteriaceae
<i>Methanococcus jannaschii</i> †	Archaea	Methanocaldococcaceae
<i>Methanopyrus kandleri</i> †	Archaea	Methanopyraceae
<i>Thermoplasma acidophilum</i>	Archaea	Thermoplasmataceae
<i>Thermoplasma volcanium</i> †	Archaea	Thermoplasmataceae
<i>Pyrococcus horikoshii</i> †	Archaea	Thermococcaceae
<i>Pyrococcus abyssi</i>	Archaea	Thermococcaceae
<i>Pyrobaculum aerophilum</i> †	Archaea	Thermoproteaceae
<i>Sulfolobus solfataricus</i> †	Archaea	Sulfolobaceae
<i>Aeropyrum pernix</i> †	Archaea	Desulfurococcaceae
<i>Saccharomyces cerevisiae</i> †	Eukaryota	Saccharomycetaceae

Continúa . . .

Cuadro 1: Genomas usados en el presente trabajo (continuación)

Genoma	Taxa	Familia
<i>Schizosaccharomyces pombe</i> †	Eukaryota	Schizosaccharomycetaceae
<i>Encephalitozoon cuniculi</i> †	Eukaryota	Unikaryonidae
<i>Aquifex aeolicus</i>	Bacteria	Aquificaceae
<i>Thermotoga maritima</i> †	Bacteria	Thermotogaceae
<i>Synechocystis</i> †	Bacteria	Chroococcales - orden -
<i>Nostoc</i> sp. PCC 7120 †	Bacteria	Nostocaceae
<i>Fusobacterium nucleatum</i> †	Bacteria	Fusobacteriaceae
<i>Deinococcus radiodurans</i> †	Bacteria	Deinococcaceae
<i>Corynebacterium glutamicum</i> †	Actinobacteria	Corynebacteriaceae
<i>Mycobacterium tuberculosis</i> H37Rv	Actinobacteria	Mycobacteriaceae
<i>Mycobacterium tuberculosis</i> † CDC1551	Actinobacteria	Mycobacteriaceae
<i>Mycobacterium leprae</i>	Actinobacteria	Mycobacteriaceae
<i>Clostridium acetobutylicum</i> †	Gramplus	Clostridiaceae
<i>Lactococcus lactis</i> †	Gramplus	Streptococcaceae
<i>Streptococcus pyogenes</i> M1 GAS	Gramplus	Streptococcaceae
<i>Streptococcus pneumoniae</i> † TIGR4	Gramplus	Streptococcaceae
<i>Staphylococcus aureus</i> N315	Gramplus	Staphylococcaceae
<i>Listeria innocua</i> †	Gramplus	Listeriaceae
<i>Bacillus subtilis</i> †	Gramplus	Bacillaceae
<i>Bacillus halodurans</i> †	Gramplus	Bacillaceae
<i>Escherichia coli</i> K12	gamma	Enterobacteriaceae
<i>Escherichia coli</i> O157:H7 EDL933	gamma	Enterobacteriaceae
<i>Escherichia coli</i> † O157:H7	gamma	Enterobacteriaceae

Continúa ...

Cuadro 1: Genomas usados en el presente trabajo (continuación)

Genoma	Taxa	Familia
<i>Yersinia pestis</i>	gamma	Enterobacteriaceae
<i>Salmonella typhimurium</i> LT2	gamma	Enterobacteriaceae
<i>Salmonella typhi</i> * †	gamma	Enterobacteriaceae
<i>Buchnera</i> sp. APS	gamma	Enterobacteriaceae
<i>Vibrio cholerae</i> †	gamma	Vibrionaceae
<i>Pseudomonas aeruginosa</i> †	gamma	Pseudomonadaceae
<i>Haemophilus influenzae</i>	gamma	Pasteurellaceae
<i>Pasteurella multocida</i>	gamma	Pasteurellaceae
<i>Xylella fastidiosa</i> † 9a5c	gamma	Xanthomonadaceae
<i>Neisseria meningitidis</i> MC58	Proteobacteria	Neisseriaceae
<i>Neisseria meningitidis</i> Z2491	Proteobacteria	Neisseriaceae
<i>Ralstonia solanacearum</i> †	Proteobacteria	Burkholderiaceae
<i>Helicobacter pylori</i> † 26695	Proteobacteria	Helicobacteraceae
<i>Helicobacter pylori</i> J99	Proteobacteria	Helicobacteraceae
<i>Campylobacter jejuni</i>	Proteobacteria	Campylobacteraceae
<i>Agrobacterium tumefaciens</i> str. C58 (Cereon)	alpha	Rhizobiaceae
<i>Sinorhizobium meliloti</i> †	alpha	Rhizobiaceae
<i>Brucella melitensis</i>	alpha	Brucellaceae
<i>Mesorhizobium loti</i> †	alpha	Phyllobacteriaceae
<i>Rickettsia prowazekii</i>	alpha	Rickettsiaceae
<i>Rickettsia conorii</i> †	alpha	Rickettsiaceae
<i>Caulobacter crescentus</i> † CB15 sinonimo: <i>Caulobacter vibrioides</i> CB15	alpha	Caulobacteraceae
<i>Chlamydia trachomatis</i>	Chlam-Spir	Chlamydiaceae

Continúa ...

Cuadro 1: Genomas usados en el presente trabajo (continuación)

Genoma	Taxa	Familia
<i>Chlamydomophila pneumoniae</i> CWL029	Chlam-Spir	Chlamydiaceae
<i>Treponema pallidum</i> †	Chlam-Spir	Spirochaetaceae
<i>Borrelia burgdorferi</i> †	Chlam-Spir	Spirochaetaceae
<i>Ureaplasma urealyticum</i> †	Gramplus	Mycoplasmataceae
<i>Mycoplasma pulmonis</i> †	Gramplus	Mycoplasmataceae
<i>Mycoplasma pneumoniae</i> †	Gramplus	Mycoplasmataceae
<i>Mycoplasma genitalium</i>	Gramplus	Mycoplasmataceae
<i>Treponema pallidum</i> †		Spirochaetaceae

Fin tabla

4.1. Identificación de hits bidireccionales

Partiendo de los dominios de proteínas definidos en la base de datos del NCBI para cada COG (<http://www.ncbi.nih.gov/COG/>) y utilizando BLAST [4] para la comparación de las secuencias, se establecieron las relaciones de bidireccionalidad tomando una proteína o dominio “A” y su mejor hit “B” y verificando después que el mejor hit de la proteína o dominio “B” fuese “A” (Sección 1.1, Figura 7). De esta manera establecimos la topología de las relaciones de bidireccionalidad (BBeT) (Figura 13) que se dan entre las secuencias o dominios de las proteínas que conforman cada uno de los COGs.

4.2. Aplicación del criterio de *Tatusov et al.*

Para cada uno de los elementos de los COGs se verificó que cumplieran de manera estricta con el criterio establecido por Tatusov *et al.* [80], donde cada elemento que se encuentra en un COG forma al menos un triángulo de BBeT con otras proteínas que pertenecen a familias taxonómicas diferentes.

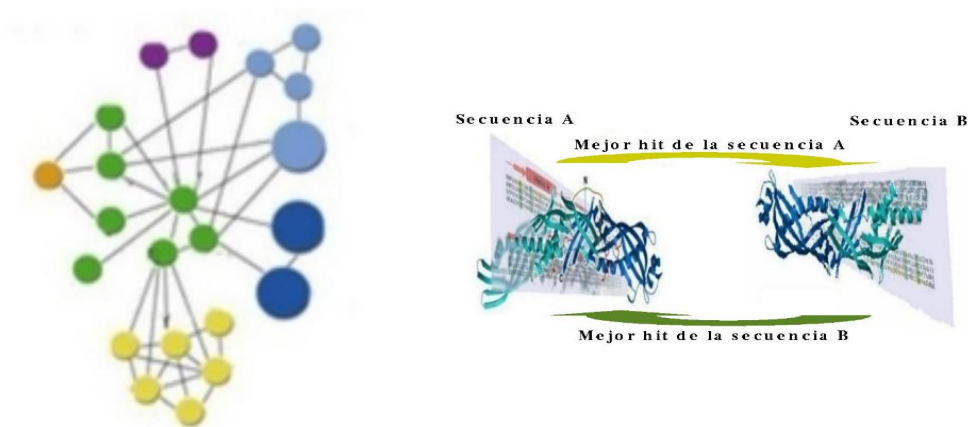


Figura 13: Identificación de las relaciones de bidireccionalidad entre las proteínas que conforman los COGs.

En la agrupación original realizada por Tatusov *et al.* [80] hay elementos dentro de los COGs que no cumplen con dicho criterio. Por ejemplo, en el COG2515 (1- Aminocyclopropane-1-carboxylate deaminase) conformado por 13 proteínas (Figura 14.A), las proteínas VC1743 de *Vibrio cholerae* y Cj0600 de *Campylobacter jejuni* sólo presentan un BBeT entre ellas, por lo cual fueron eliminadas. De esta manera después del primer filtro pasamos de 13 proteínas a 11 en el COG2515 (Figura 14.B).

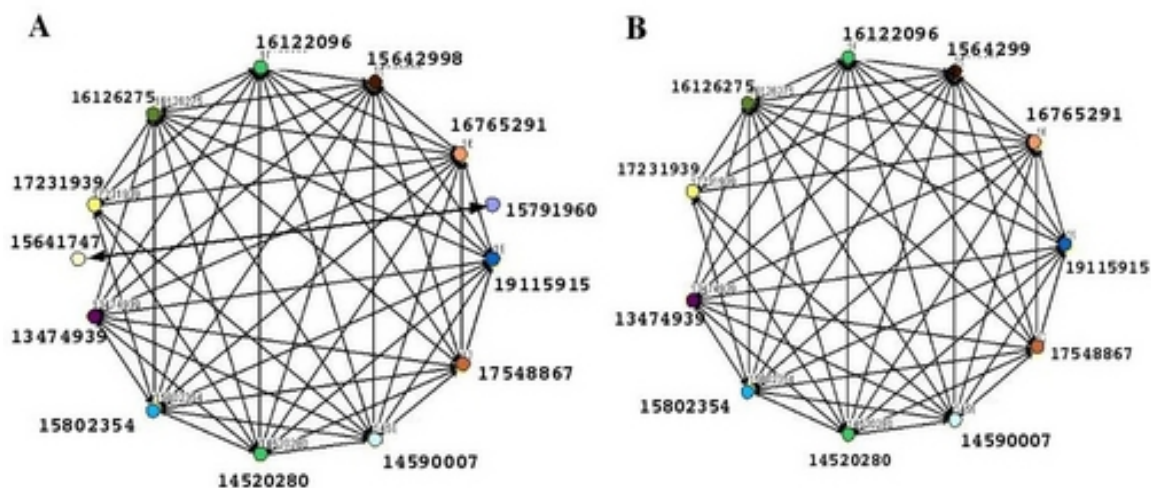


Figura 14: Disminución de elementos del COG2515. **A.** Las proteínas VC1743 (GI:15641747) y Cj0600 (GI:15791960) sólo presentan un BBeT, por lo que fueron eliminadas. **B.** Conformación del COG2515 después del primer filtro. La visualización de la topología de las relaciones de BBeT se realizó con el programa Osprey [12].

Se analizó cada una de las proteínas de los COGs, a las cuales se verificó de manera individual que formaran al menos un triángulo de BBeT conservando las demás conexiones que tuvieran. El establecimiento de este filtro disminuyó el número de elementos de los COGs, como en el ejemplo del COG2515 antes mencionado (Figura 14.B).

En 787 COGs, ninguno de sus elementos cumplió con el criterio antes descrito por lo que fueron eliminados. Una lista de los COGs eliminados se muestra en el apéndice 1 (ver Sección 9). Como ejemplo de COGs eliminados podemos mencionar al COG1306 (Uncharacterized conserved protein), el cual está constituido por tres proteínas FN0456 (*Fusobacterium nucleatum*), TP0083 (*Treponema pallidum*) y BB0165 (*Borrelia burgdorferi*) (Figura 15.A). Cada una de las proteínas tiene dos BBeTs. BB0165 y TP0083 pertenecen a la misma familia, Spirochaetaceae, por lo que una de sus conexiones fue eliminada (Figura 15.B).

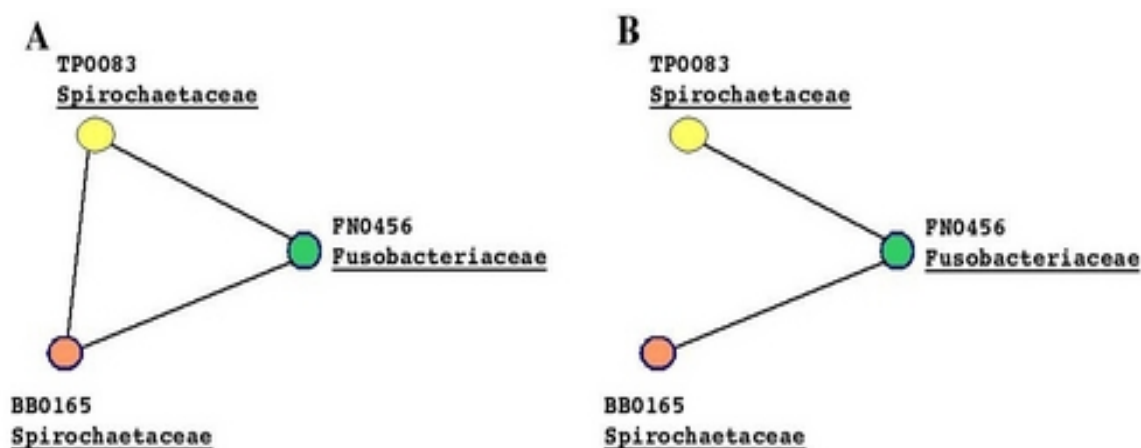


Figura 15: COG1306. Las familias a las que pertenecen cada proteína están escritas y subrayadas debajo de cada una de ellas. **A.** Topología original de BBeTs. **B.** Eliminación del BBeT entre las proteínas BB0165 y TP0083 por pertenecer a la misma familia.

Esto provocó que la topología del COG se modificara (Figura 15.B), dando como resultado que ninguna de las proteínas cumpliera con el criterio del triángulo de BBeTs. Al final de esta selección nos quedamos con 4, 086 COGs (Cuadro 2).

Cuadro 2: Número de COGs antes y después de aplicar el criterio de Tatusov [80] (ver texto).

COGs iniciales	Filtro	COGs finales
4, 873	Criterio de Tatusov <i>et al.</i> [80]. 787 COGs eliminados	4, 086

En la base de datos de los COGs tomada del NCBI tenemos 1, 828 COGs que están conformados únicamente por proteínas ortólogas, mientras que los 3, 045 restantes presentan al menos una proteína paróloga en su interior (Figura 16.A, Cuadro 3).

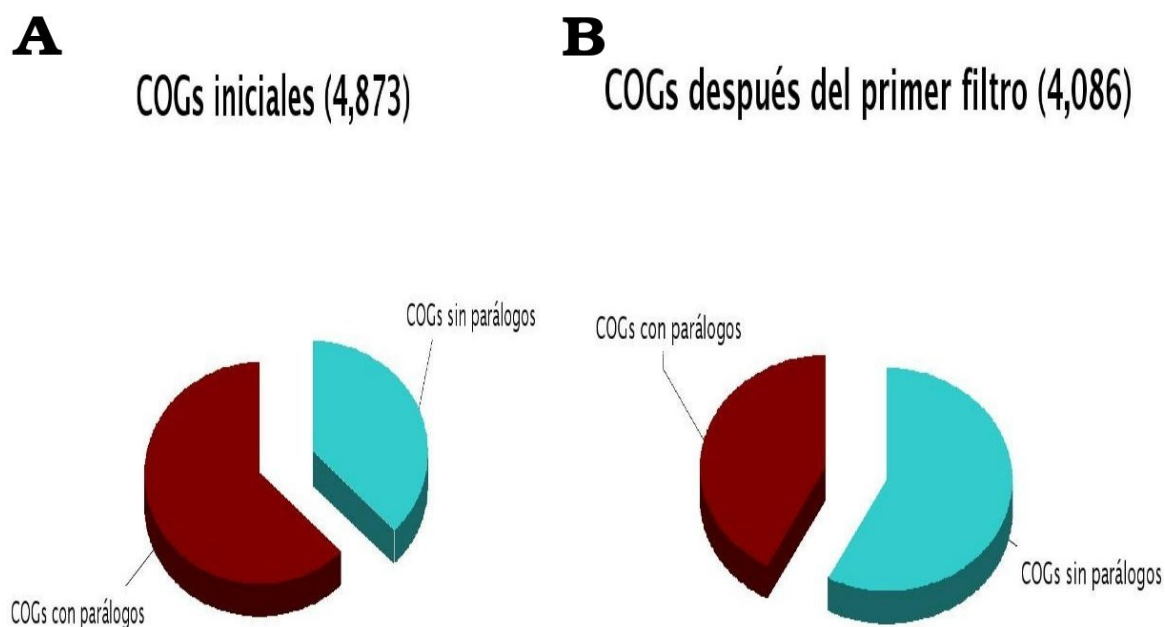


Figura 16: Porcentaje de parálogos en los COGs. **A.** Porcentaje de COGs que presentan parálogos en los archivos de inicio. **B.** Porcentaje de COGs que presentan parálogos después de la aplicación del primer filtro. En azul se representa el porcentaje de COGs sin parálogos, y en rojo los COGs que tienen parálogos.

La aplicación del primer filtro eliminándose 787 COGs en los cuales los elementos que los conformaban no cumplían con presentar un triángulo de BBET con proteínas que pertenecieran a familias taxonómicas distintas. Pasamos de tener 4, 873 COGs a tener 4, 086 COGs y pasamos de un 62.5 % de COGs con al menos una proteína paróloga a un 42.21 % (Figura 16.B, Cuadro 3).

Cuadro 3: Número de COGs sin y con proteínas parálogas.

	Inicio	Después del primer filtro
COGs sin parálogos	1, 828 (37.5 %)	2, 361 (57.78 %)
COGs con parálogos	3, 045 (62.5 %)	1, 725 (42.21 %)

4.3. Eliminación de genomas redundantes

La redundancia de los organismos filogenéticamente cercanos fue eliminada con base en el número de proteínas ortólogas que comparten. Para ello se ordenaron los genomas de acuerdo a su tamaño, de mayor a menor en una lista, y se fueron comparando sus proteínas de manera descendente, tomando al primer genoma como referencia se comparó con los de menor tamaño para encontrar los ortólogos comunes a ambos genomas, cuando el porcentaje de proteínas ortólogas compartidas era mayor o igual al 50 % con respecto al genoma de menor tamaño, éste último era eliminado. De esta manera se fue descendiendo por la lista hasta terminarla. En un segundo ciclo se repite la misma operación con la diferencia de que el genoma con el que se realizan las comparaciones, es el segundo de mayor tamaño que halla pasado el primer ciclo, dado que en el ciclo anterior fue analizado el genoma de mayor tamaño (Figura 17). El corte de 50 % de proteínas ortólogas compartidas nos permite eliminar cepas de la misma especie.

La eliminación de los genomas redundantes junto con sus proteínas, modificó la topología de las conexiones bidireccionales de las proteínas de los genomas que no fueron eliminados, por lo que se volvió a verificar que las proteínas cumplieran con tener al menos un triángulo de BBeT con proteínas pertenecientes a diferentes familias taxonómicas. En este proceso se eliminaron 423 COGs (ver apéndice 2, sección 10), y pasamos de tener 67 genomas a 43 (Cuadro 1 y Cuadro 4). El número de COGs sin proteínas parálogas al final de este filtro fue de 2, 161, mientras que el número de COGs con proteínas parálogas fue de 1, 502 (Figura 18, Cuadro 5).

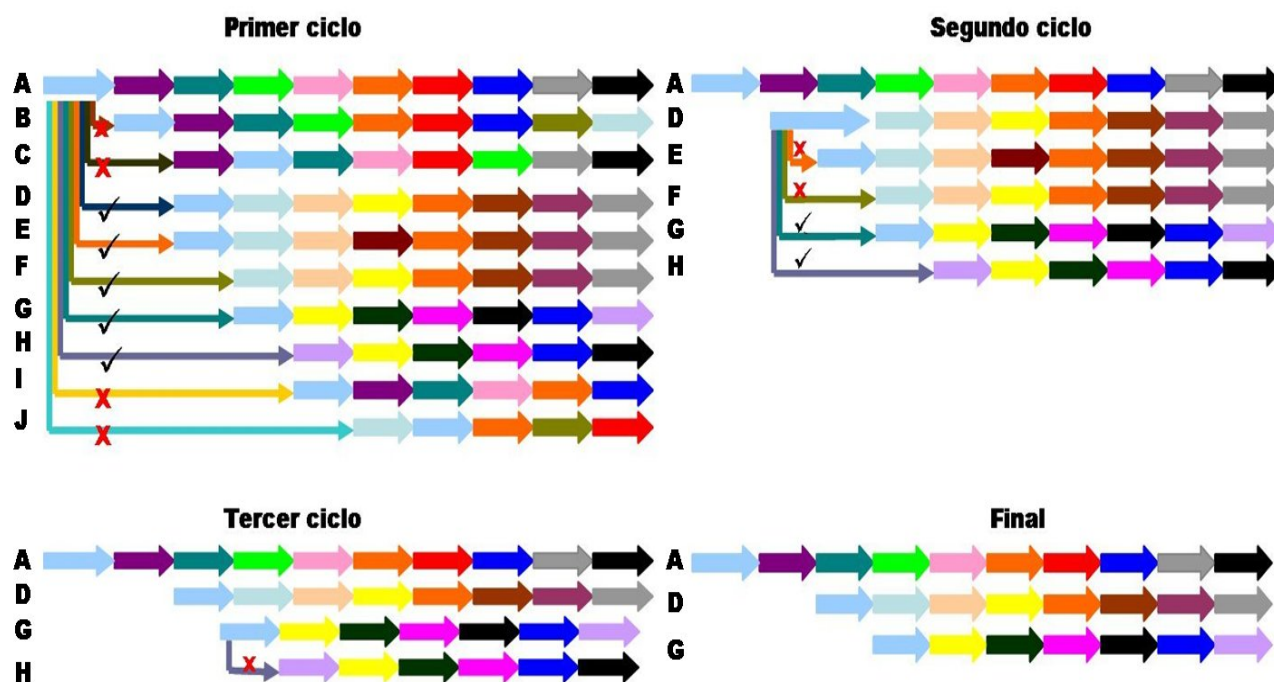


Figura 17: Eliminación de genomas redundantes. Cada línea de flechas representa a los genes contenidos en un genoma. Las flechas y sus colores representan a genes que codifican proteínas con distinta función. Genes que codifican proteínas ortólogas comunes a dos genomas se representa con flechas del mismo color en los respectivos genomas. Se ejemplifican tres ciclos. Los genomas eliminados en cada ciclo se señalan con una **x**, mientras que los genomas que pasan al siguiente ciclo se señalan con una **✓**.

Cuadro 4: Número de COGs antes y después de eliminar los genomas redundantes (ver texto).

INICIO	Filtro	FIN
4, 086 COGs	Eliminación de genomas redundantes.[80]	3, 663 COGs
67 genomas		43 genomas

Cuadro 5: Número de COGs sin y con proteínas parálogas, después del segundo filtro.

	Inicio	Después del segundo filtro
COGs sin parálogos	2, 361 (57.78 %)	2, 161 (59.0 %)
COGs con parálogos	1, 725 (42.21 %)	1, 502 (41.0 %)

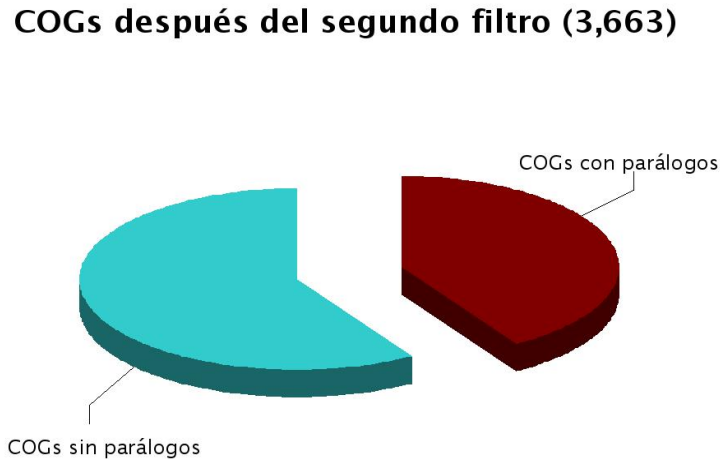


Figura 18: Porcentaje de COGs con proteínas parálogas después del segundo filtro. En azul el porcentaje de COGs sin parálogos y en rojo el porcentaje de COGs con parálogos después del segundo filtro.

4.4. Agrupamiento al interior de los COGs

Los 3, 663 COGs que restan después de aplicar los dos filtros antes descritos fueron utilizados como archivos de entrada para el programa Markov Cluster Algorithm (MCL) (<http://micans.org/mcl/>) desarrollado por Van Dongen [83], el cual ha sido utilizado en aplicaciones biológicas como la detección de familias de proteínas por Enright *et al.*[21]. MCL simula un flujo en una red de elementos interconectados partiendo de la idea de que si al interior de la red hay grupos (“clusters”) los elementos de dichos grupos se encontrarán más interconectados entre sí que con elementos que se encuentren en otros grupos. Por lo tanto, el flujo que viaja por la red tendrá más probabilidad de permanecer dentro de los grupos que saltando de un grupo a otro (Figura 12), de esta manera son identificados los “clusters” dentro de la red. La aplicación de MCL a los COGs se llevó a cabo representando de manera binaria las conexiones de BBeT entre las secuencias de aminoácidos. Diferentes valores de granularidad (I) fueron probados. Tomando al COG0147 (Anthranilate/para-aminobenzoate) como control se probaron valores de I de 1.2, 1.5, 2.0, 3.0, 4.0 y 5.0 (Cuadro 6).

El COG0147 está conformado por las proteínas parálogas antranilato sintasa (*trpE*) componente I y para-aminobenzoato sintasa (*pabB*). Ambas enzimas utilizan al corismato como sustrato para producir triptófano y ADC (4-amino-4-deoxi corismato) respectivamente (ver Sección 5.3.4). A una granulari-

Cuadro 6: Valores de granularidad (I) probados en el COG0147.

Granularidad	Clusters	Elementos máximos por cluster	Elementos mínimos por cluster
1.2	1	55	55
1.5	2	37	18
2.0	2	34	21
3.0	2	34	21
4.0	5	30	4
5.0	5	29	4

dad de 1.2 el COG no se divide, por lo que dicho valor se descarto. Para el intervalo de valores entre 1.5 y 3.0 de granularidad las secuencias se dividieron en dos clusters, como era lo esperado. Las proteínas que se encuentran dentro de cada cluster presentan en más de un 50% anotaciones diferentes para cada grupo, TrpE para uno y PabB para el otro (ver Cuadro 16). Con los valores de granularidad de 4.0 y 5.0 se formaron 5 clusters. Los 3 nuevos clusters formados con estos valores estan integrados por una mezcla de secuencias anotadas como TrpE y PabB. No se encontro alguna anotación que diera indicios de que estos grupos puedan representar a otras proteínas parálogas de TrpE, como lo es el producto del gene *phzE*, o algún otro. Con base en el análisis del COG0147 se tomo el valor de I=2.0 para ser usado en los COGs restantes.

La distribución de los grupos encontrados al interior de los COGs por MCL va desde un grupo único por COG, hasta COGs con 71 grupos en su interior (Cuadro 7, Figura 19).

El COG1028 el cual agrupa a las proteínas de la familia deshidrogenasa/reductasa presentó 71 grupos en su interior, mientras que el COG0477 presentó 70 grupos, donde se encuentran proteínas transportadoras. Ejemplos de COGs con un sólo grupo encontrado por MCL son el COG0033 el cual contiene a las proteínas fosfoglucomutasas, COG0302 en el cual se agrupan las GTP ciclohidasas I; el COG0147 es un ejemplo de COGs con dos grupos en su interior.

Cuadro 7: Distribución de grupos formados por MCL al interior de los COGs.

Número de COGs	Número de grupos
2, 697	1
493	2
196	3
99	4

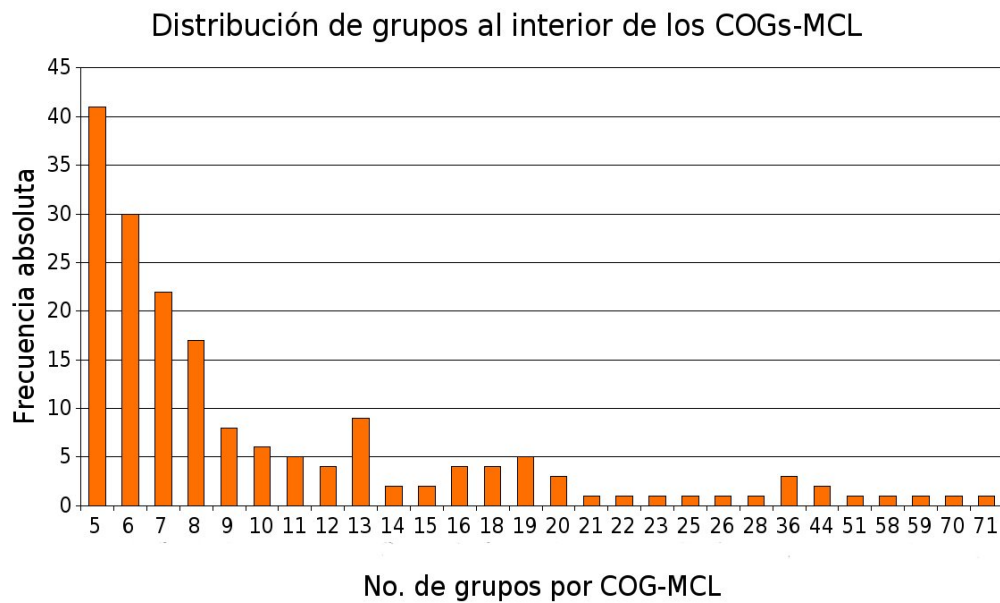


Figura 19: Distribución de grupos obtenidos al interior de los COGs después de utilizar el programa MCL. En el histograma se representa la distribución a partir de 5 grupos en adelante.

4.5. Selección de COGs divididos por MCL

Para obtener un perfil de matrices peso-posición de los grupos encontrados por MCL en los COGs se considero que cada grupo debería estar formado por al menos un número mínimo de 5 secuencias iniciales o semillas. Para lograr lo anterior se eliminarón aquellos grupos que en su interior tuvieran menos de 5 elementos. Nuestro algoritmo trabajó tomando el o los grupos en que estuviera dividido el COG verificando que el número mínimo de secuencias dentro del grupo fueran mayor o igual a cinco, eliminando a los que no cumplieran este criterio. De esta manera se eliminaron COGs (ver apéndice 3, sección 11) que estaban formados por uno o más grupos con menos de cinco secuencias cada uno

(Figura 20.a), como el COG5434 o el COG4997, conformados de un solo grupo con cuatro y tres elementos respectivamente. También se redujo el número de grupos dentro de algunos COGs (Figura 20.b). Ejemplo de esto es el COG1241, en el cual se encontraron cuatro grupos, uno de ellos con trece elementos, mientras que los restantes con tres elementos cada uno fueron eliminados.

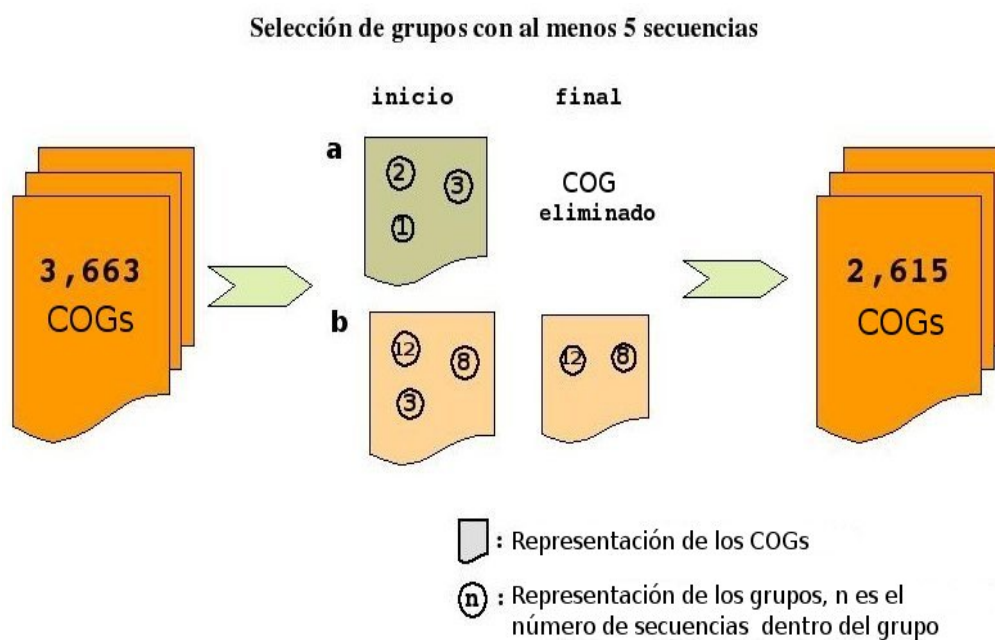


Figura 20: Selección de grupos con al menos 5 secuencias. Se ejemplifica el análisis de dos COGs en **a** y **b**. **a**) Eliminación total del COG por tener todos los grupos menos de 5 secuencias. **b**) Eliminación de algunos grupos en el COG que no tuvieron al menos 5 secuencias.

Al final de esta selección nos quedamos con 2, 615 COGs, los cuales están conformados por grupos con al menos cinco proteínas cada uno. La segunda selección que se hizo sobre estos 2, 615 COGs fue el quedarnos con aquellos que tuvieran al menos dos grupos en su interior con lo cual obtuvimos 584 COGs (Figura 21). Una lista de los COGs eliminados puede verse en el apéndice 3 (Sección 11).

Por lo anterior los 584 COGs que resultaron de los filtros aplicados presentan al menos dos grupos en su interior con cinco secuencias como mínimo conformandolos. Una lista con los 584 COGs finales puede verse en el apéndice 4 (Sección 12).

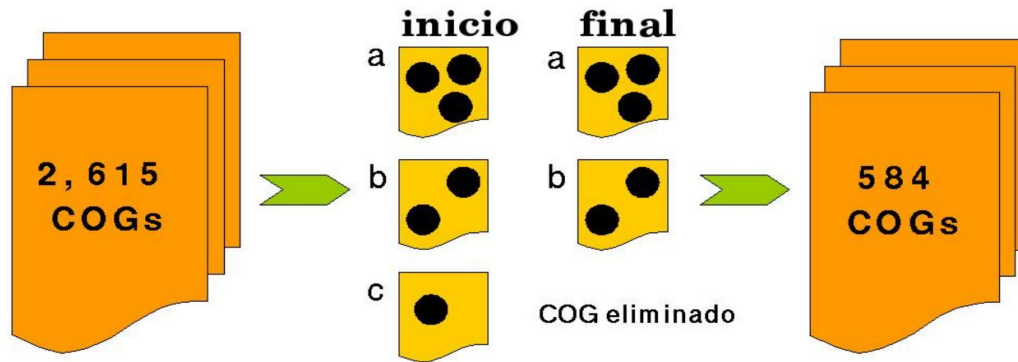


Figura 21: Selección de COGs con al menos dos grupos al interior. En el ejemplo se muestran 3 COGs de los cuales “c” es eliminado por contener menos de dos grupos.

4.6. Matrices de peso-posición

Las funciones biológicas frecuentemente son llevadas a cabo por proteínas que no comparten una secuencia de aminoácidos idéntica pero si relacionada, de tal manera que aunque los aminoácidos varían en las diferentes posiciones de las proteínas la función se mantiene. Por otro lado no toda la secuencia de la proteína es responsable de la función sino solo ciertas regiones (motivos) conservadas dentro de la secuencia. Las matrices peso-posición (Position Weight Matrices PWM) nos permiten capturar la variación de la frecuencia de aparición de los aminoácidos en las diferentes posiciones de los motivos que pudieran presentarse en las proteínas, con lo cual la información obtenida nos permite la búsqueda de secuencias relacionadas, basandonos en un modelo probabilístico más que en una secuencia única o en una expresión regular determinada. Una variación de la matrices peso-posición (PWM) son aquellas donde se toma en cuenta también a los aminoácidos que están fuera del motivo, para lo cual se genera otra matriz que describa la frecuencia de aparición de estos aminoácidos. De esta manera, tenemos una matriz A describiendo la variación de los aminoácidos dentro del motivo y una matriz B describiendo la variación de los aminoácidos fuera del motivo. La información que se obtiene con estas dos matrices de la proteína es más global. Los métodos que utilizan la estrategia de maximización de lo esperado (expectation maximization EM) generan un modelo de matrices el cual pueda reconstruir a la secuencia de entrada de dicho modelo. En un primer paso es generada la matriz A del motivo así como la matriz B de los aminoácidos fuera del motivo. Posteriormente conforme

más secuencias son analizadas, las matrices A y B son optimizadas con la nueva información hasta llegar a converger a una solución o matrices A y B que mejor describan a las secuencias de entrada. El programa MEME (Multiple EM for Motif Elicitation) adopta este enfoque, generando un conjunto de matrices (perfil) para las secuencias que son analizadas con él. Cada uno de los subgrupos (2, 044) de los 584 COGs finales, fueron utilizados como entrada para el programa MEME, generandose perfiles para cada uno de ellos. Estos perfiles contienen la información característica de los subgrupos formados dentro de un COG determinado, la cual puede ser utilizada para identificar la pertenencia de un elemento dentro del subgrupo al que corresponde. Las matrices generadas para cada subgrupo dentro de un COG fueron usadas con el programa MAST (Motif Alignment and Search Tool). MAST toma como entrada un conjunto de matrices, con las cuales busca en una base de datos aquellas secuencias que mejor se adecúen al perfil de las matrices peso-posición de entrada. La búsqueda que se realizó con MAST usando las matrices de cada subgrupo de un COG determinado, se llevó a cabo sobre todas las secuencias que conforman el COG en su totalidad, incluyendo aquellas secuencias eliminadas durante la aplicación de los filtros antes descritos. De esta manera, los 584 COGs seleccionados con dos subgrupos como mínimo en su interior fueron caracterizados utilizando matrices de peso-posición, y sus perfiles de matrices utilizados para la búsqueda de secuencias relacionadas. A este conjunto de 2, 044 grupos obtenidos se les nombro **MOGs** por la utilización de los programas antes descritos y de la base de datos de la cual se parte (MCL-MEME - MAST-COGs) (Figura 22).

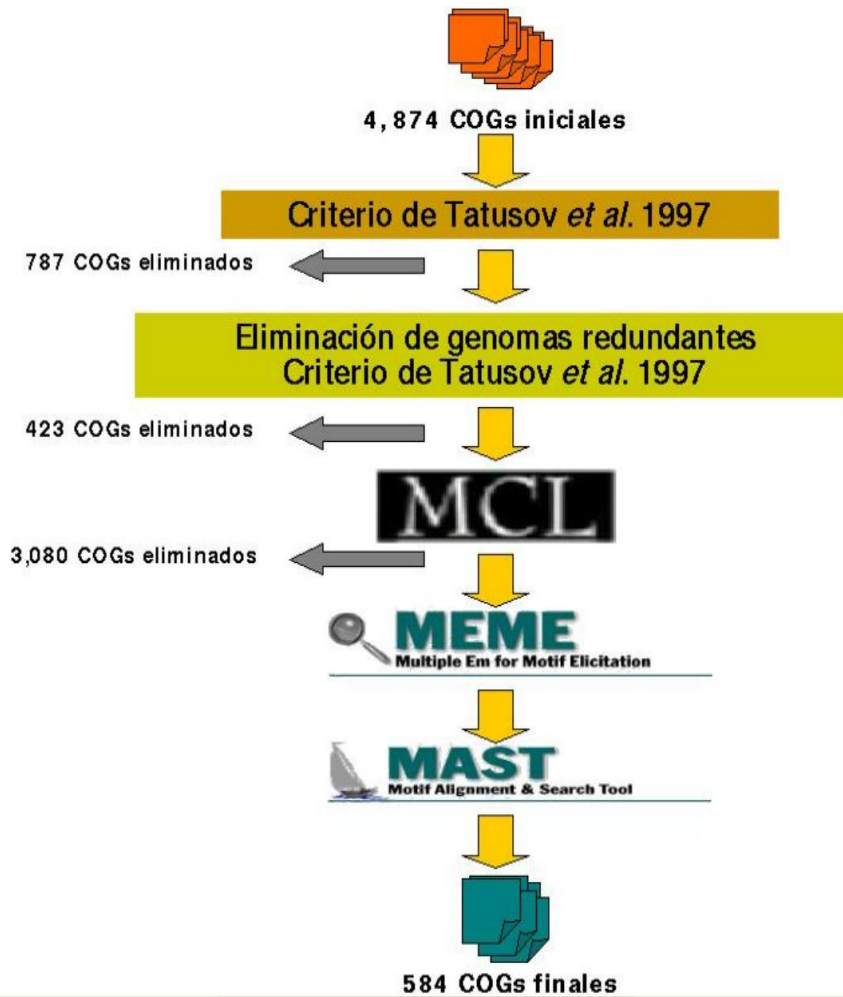


Figura 22: Diagrama que muestra el número de COGs iniciales y el número de COGs finales obtenidos después de aplicar los filtros descritos. Los 584 COGs finales presentan parálogos en su interior

5. Resultados: Significado biológico de los agrupamientos

La división de las proteínas obtenida en los COGs se basa en la exploración de las relaciones de similitud bidireccional que se dan entre las secuencias de aminoácidos. La congruencia de dichos agrupamientos en relación al conocimiento biológico fue analizada explorando la base de datos de familias de proteínas Pfam, así como la base de datos de motivos de proteínas PROSITE. De igual manera se buscó información publicada en la literatura para documentar la separación de algunos COGs. En las subsecciones y secciones siguientes, al mencionar a los COGs, haremos referencia a los 584 COGs obtenidos después de aplicar los filtros antes descritos, los cuales presentan parálogos.

5.1. Comparación contra la base de datos Pfam

Pfam es una colección extensa de familias y dominios de proteínas. En su última versión (21.0) Pfam contiene 8,957 familias (<http://www.sanger.ac.uk/Software/Pfam/>), las cuales cubren un 75 % de las proteínas contenidas en las bases de secuencias de proteínas Swiss-Prot y TrEMBL [10]. Cada familia en Pfam está representada por dos alineamientos múltiples y dos perfiles generados por Modelos Ocultos de Markov (HMM), además de un archivo con la anotación disponible para cada familia [8], [11]. Uno de los dos alineamientos corresponde a las secuencias semillas, donde se encuentran miembros representativos de cada familia, mientras que el segundo alineamiento se corresponde a todos los miembros de la familia detectados con el perfil HMM realizado con las secuencias semillas [9].

Uno de los principales usos de Pfam es la anotación genómica. Como en el trabajo realizado por Batema *et al.* [11], en el cual utilizan el Pfam para la anotación de los proteomas de *E. coli*, *Rickettsia prowazekii*, *Methanococcus jannaschii*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* y *Homo sapiens*. Pfam ha sido utilizado también para la predicción de splicing alternativo por Hiller *et al.* [30], en la minería de datos para validar nuevos métodos de búsqueda de información de proteínas en la base de datos de referencias pública PubMed (<http://www.ncbi.nlm.nih.gov>) y su posterior agrupamiento en familias [50], en la asignación y comprobación experimental de la doble función de clorito dismutasa y de monooxigenasa en la síntesis de antibióticos de la proteína PitA de *Haloferax volcanii*

[7], en estudios de transferencia horizontal de genes (HGT) [16], propuesta como base de datos para obtener de ella 5,000 familias de proteínas que puedan analizarse en el proyecto internacional Structural Genomics, el cual trata de determinar la estructura tridimensional de macromoléculas biológicas importantes [14], entre otras.

Los perfiles HMM generados por Pfam para cada familia o dominio contenida en su base de datos, fueron utilizados para la identificación de familias funcionalmente distintas en los grupos obtenidos dentro de los 584 COGs finales. Los 584 COGs divididos presentaron un perfil HMM para al menos alguno de sus grupos. Para 542 COGs los perfiles HMM se repitieron en los grupos formados dentro del COG. 42 COGs de los 584, presentaron perfiles HMM específicos (diferentes) para alguno de los grupos (MOGs) encontrado por MCL al interior del COG (ver apéndice 5, sección 13). Estos 42 COGs representan un 7.19 % del total de los 584 COGs finales, para los cuales sus grupos tiene perfiles HMM específicos (diferentes) o no repetidos (Fig. 23).

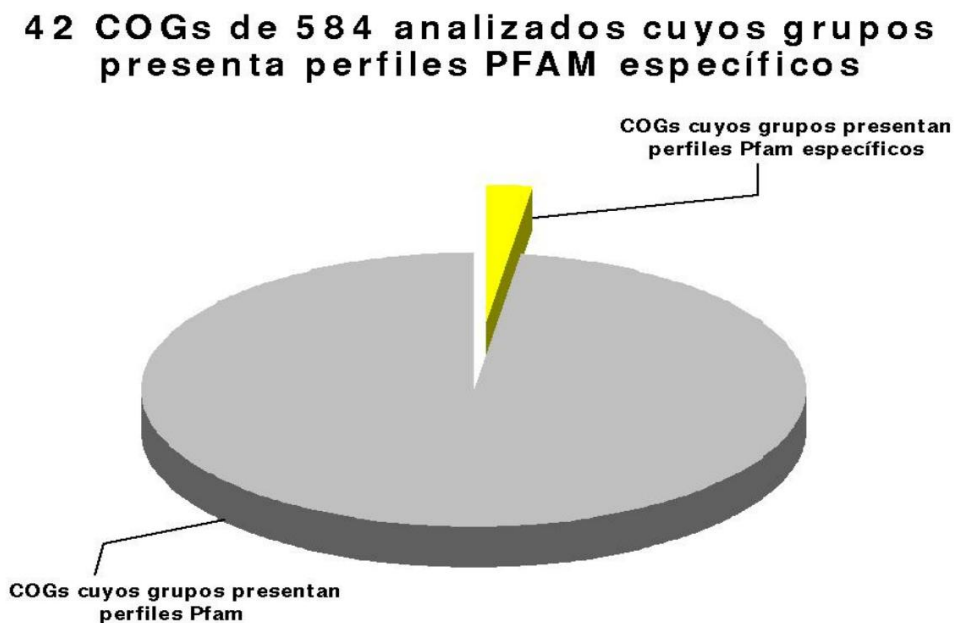


Figura 23: De los 584 COGs finales, 542 (92.8 %) COGs presentaron perfiles Pfam repetidos para sus grupos y 42 (7.19 %) COGs presentaron perfiles HMM diferentes para al menos uno de sus grupos.

La mayoría de los 42 COGs (ver apéndice 5, sección 13) cuyos grupos presentan perfiles HMM específicos fueron COGs con dos grupos (MOGs) a su interior. Mientras más grupos hay en el interior de los COGs como es el caso de los COGs que agrupan a las permeasas (49 subgrupos) y las deshidrogenasas

(32 subgrupos), mayor es la dificultad para poder asignar de manera clara perfiles HMM específicos a cada grupo (MOG). El COG0262 (Dihydrofolate reductase) es un ejemplo, en el cual se encontraron perfiles Pfam diferentes para cada grupo. Este COG fue dividido en dos grupos, para el grupo 1, en todas las secuencias se encontró el perfil Pfam PF00186, el cual corresponde a la dihidrofolato reductasa (DHFR), mientras que en el grupo 2, únicamente una secuencia presentó dicho perfil (Figura 24). La DHFR lleva a cabo la reducción de hidrofolato a tetrahidrofolato, un paso esencial en la biosíntesis *de novo* de la glicina y las purinas. En el grupo 2, todas las secuencias presentaron el perfil PF01872, el cual se corresponde a el dominio carboxilo RibD (Figura 24). Este dominio está envuelto en el segundo y tercer paso de la biosíntesis de riboflavina, y esta relacionado con DHFR; mientras que en el grupo 1, el perfil PF01872, no se presento en ninguna secuencia.

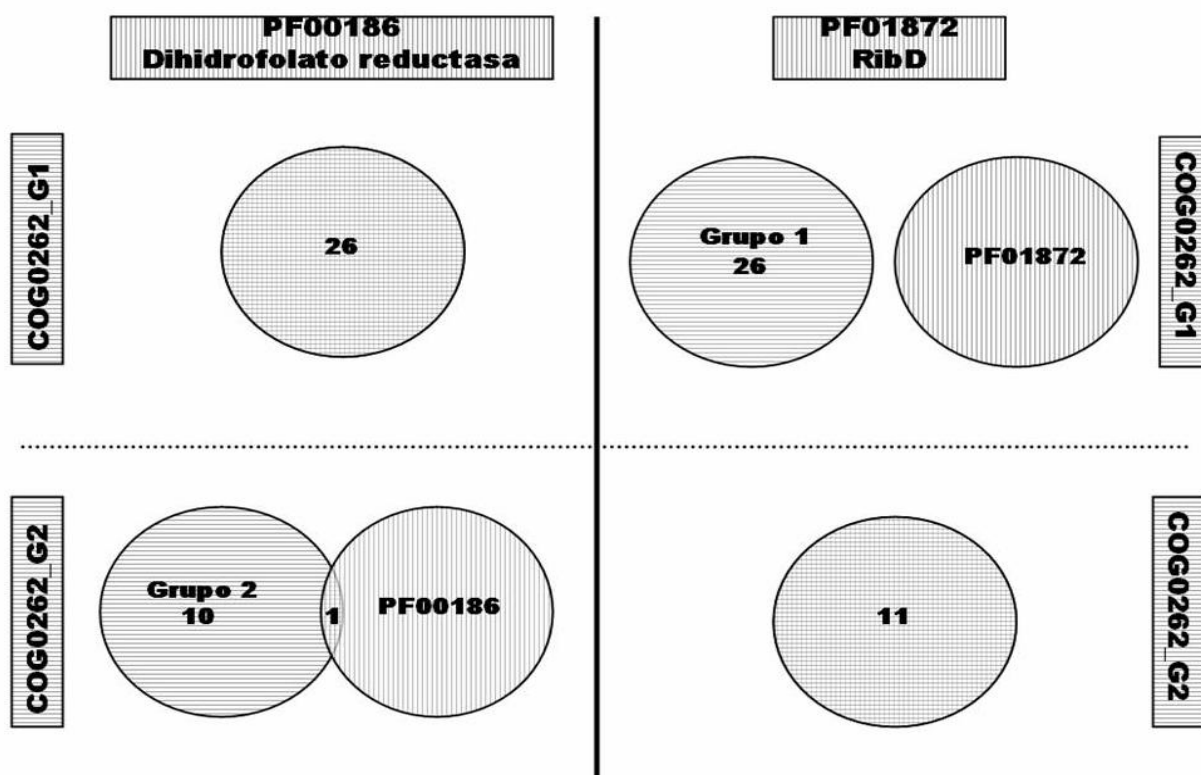


Figura 24: El perfil Pfam PF00186 se presentó en las 26 secuencias que conforman el grupo 1 del COG0262, mientras que en las secuencias del grupo 2, sólo una secuencia de 11 presentó el perfil PF00186. El perfil Pfam PF01872 se presentó en las 11 secuencias que conforman al grupo 2 del COG0262, mientras que ninguna de las 26 secuencias del grupo 1 tuvo este perfil. Los círculos con líneas verticales representan a los perfiles Pfam, mientras que los círculos con líneas horizontales representan a los grupos.

5.2. Comparación contra la base de datos PROSITE

La base de datos PROSITE contiene una colección de firmas biológicas significativas, las cuales son descritas como patrones o perfiles [33]. PROSITE utiliza dos tipos de firmas o descriptores para identificar regiones conservadas (motivos) o perfiles generalizados en las proteínas, cada firma en PROSITE se encuentra documentada [72]. Muchos motivos conservados entre las proteínas y biológicamente importantes por ser sitios con propiedades de pegado o sitios catalíticos son recogidos en PROSITE. Los motivos y perfiles (matrices de peso-posición) en PROSITE se encuentran almacenados para ser utilizados en la búsqueda de proteínas que pertenezcan a una familia documentada en PROSITE, si existe [31].

PROSITE ha sido utilizado en la confirmación de la función de un gene en *Candida albicans* homólogo de la prolina-tRNA, el cual presenta el motivo de las aminoacil-tRNA sintetasas clase II documentada en PROSITE [71], en la anotación de proteínas de proyectos de secuenciación en humanos [57], en la identificación de la localización de proteínas subcelulares de eucariotes [88], en la asignación de funciones del cluster de genes *ibeA*, el cual se encuentra en una islas de patogenicidad de *E. coli* enteropatógena [32], entre otros.

Los motivos y perfiles contenidos en PROSITE fueron utilizados para la búsqueda de firmas distintas en los grupos generados al interior de los COGs. 435 COGs de los 584 COGs finales divididos, presentaron firmas PROSITE en al menos uno de sus grupos. De estos 435 COGs, 398 presentaron firmas PROSITE repetidas para los grupos formados en ellos, mientras que los 37 COGs restantes presentaron firmas PROSITE específicas (diferentes) para al menos alguno de los grupos obtenidos dentro de cada COG. Estos 37 COGs representan un 6.33 % del total de los 584 COGs finales, para los cuales al menos uno de sus grupos tiene firmas PROSITE específicas (Fig. 25).

Al igual que en la búsqueda de perfiles diferentes para cada grupo (MOG) de los COGs finales con Pfam, las firmas PROSITE aparecen en su mayoría en aquellos COGs que fueron divididos en dos grupos. Dificultandose la asignación clara de firmas PROSITE para COGs con más de dos grupos (MOGs) en su interior. Un ejemplo de COG anotado con PROSITE es el COG1304 (L-lactate dehydrogenase (FMN-dependent) and related alpha-hydroxy acid dehydrogenases). El COG1304 fue dividido en dos

37 COGs de 584 analizados cuyos grupos presentan firmas PROSITE específicas

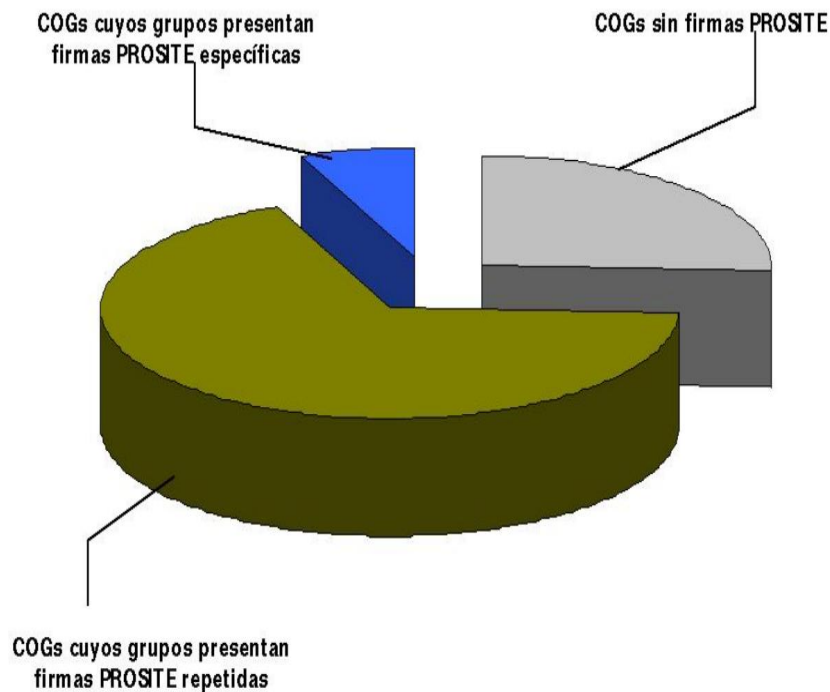


Figura 25: De los 584 COGs finales, 149 (25.51 %) COGs no presentaron firma PROSITE alguna, 398 (68.15 %) COGs presentaron firmas PROSITE repetidas en sus grupos y 37 (6.33 %) COGs presentaron firmas PROSITE diferentes para cada grupo.

grupos, en el grupo 1, 17 de 21 (80.95 %) secuencias presentan la firma PS00557 (FMN-dependent alpha-hydroxy acid dehydrogenases active site) (Figura 26), la cual se encuentra presente en proteínas estructuralmente relacionadas como la lactato deshidrogenasa (cataliza la conversión de lactato a piruvato), glicolato oxidasa, entre otras. Ninguna secuencia del grupo dos presento la firma PS00557 (Figura 26).

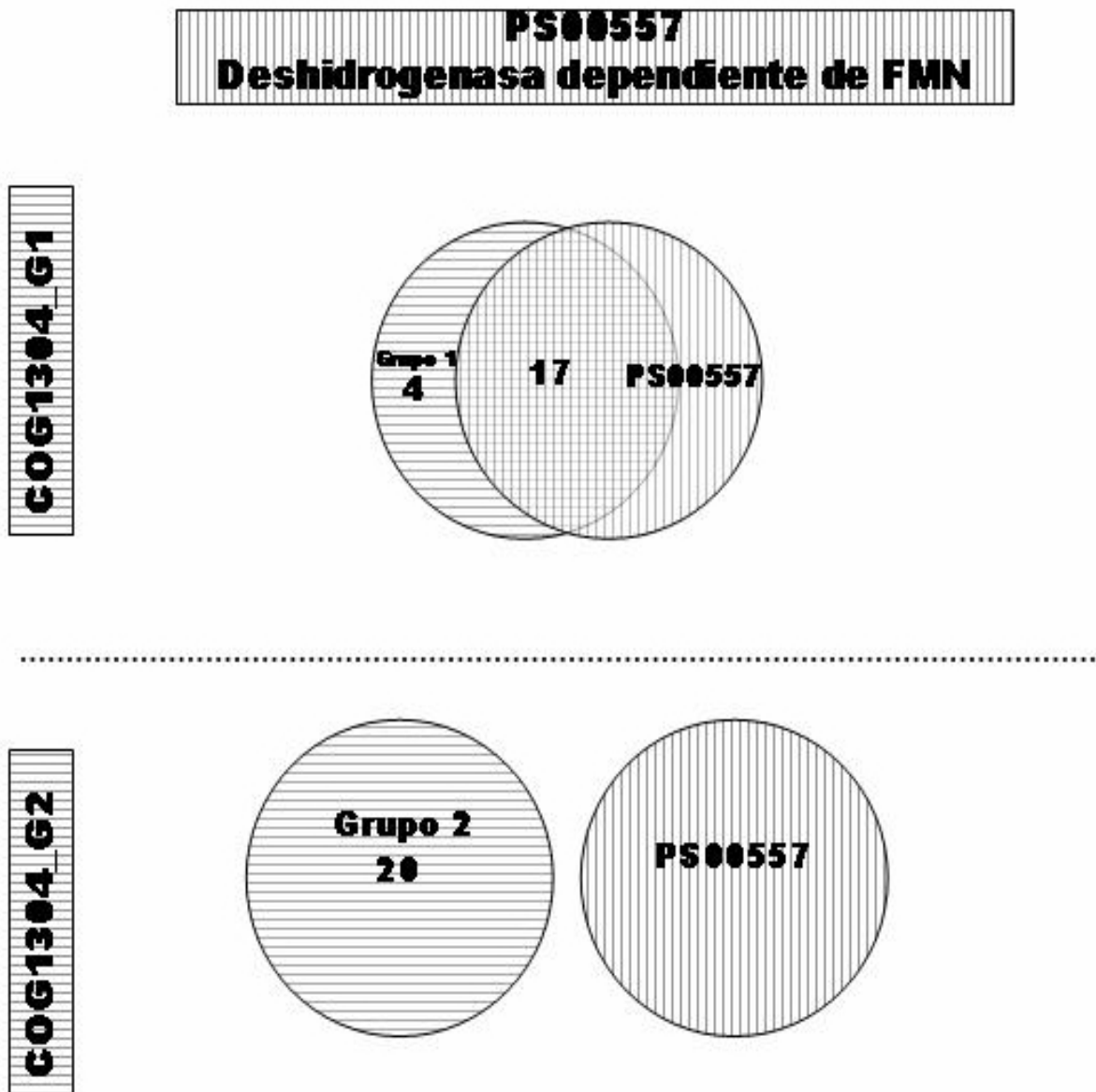


Figura 26: La firma PROSITE PS00557 se presentó en 17 de las 21 secuencias que conforman el grupo 1 del COG1304, mientras que en las secuencias del grupo 2, ninguna de las 20 secuencias del grupo tuvo este perfil. Los círculos con líneas verticales representan a la firma PROSITE, mientras que los círculos con líneas horizontales representan a los grupos.

5.3. Documentación manual de COGs

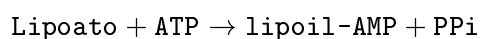
Se tomaron 4 COGs para los cuales se buscó información bibliográfica que pudiera documentar el significado biológico de las divisiones obtenidas en los COGs analizados. Estos estudios de caso se

presentan a continuación

5.3.1. COG0095. Lipoate-protein ligase A

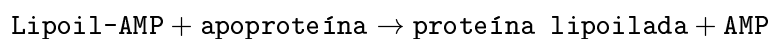
El ácido lipoico es un grupo prostético de la subunidad acetiltransferasa (E2) del piruvato y del α -cetoglutarato entre otros, se une específicamente a un residuo de lisina en la proteína vía una unión amida entre el grupo amino- ϵ del residuo de lisina y el grupo carboxil del ácido lipoico.

La unión del ácido lipoico a las proteínas ocurre en una reacción de dos pasos en la cual el intermediario lipoil-AMP es formado a partir del ácido lipoico y ATP (Reacción 1).



Reacción 1

En el segundo paso el lipoil-AMP es transferido al dominio lipoil de la apoproteína, produciendo la proteína con el grupo lipoil unido y AMP (Reacción 2) [23].



Reacción 2

Se han reportado dos rutas en *Escherichia coli* por las cuales se realiza la lipoilación, una de ellas llevada a cabo por la proteína LplA (lipoate-protein ligase) cuyos pasos son descritos en las reacciones 1 y 2, en donde LplA activa a un ácido lipoico exógeno; en la otra ruta interviene la proteína LipB la cual usa a la proteína transportadora acil-lipoil (lipoyl-acyl carrier lipoyl-ACP) como el donador del grupo lipoil ([53], [40]). La biotín-proteína ligasa (biotinyl-protein ligase BPL) lleva a cabo la activación de biotina con ATP formando biotín-AMP para posteriormente transferir el grupo biotín a la proteína transportadora biotín-carboxil (Biotin carboxyl carrier protein) en una serie de reacciones similares a las que lleva a cabo LplA [44].

Las enzimas LplA, LipB y BPL presentan muy poca similitud a nivel de secuencia conservándose únicamente un residuo de lisina en las tres proteínas. LplA y LipB presentan también una región homóloga al módulo catalítico de la enzima BPL. En *E. coli* es conocido que los residuos en el “loop”

que conectan las láminas $\beta 2$ y $\beta 3$ de BPL contribuyen a la unión del sustrato (biotina y ATP) y/o a la catálisis, los cuales están conservados en diferentes BPLs. Patrones de conservación de los residuos aparecen también en los loops equivalentes de LplA y LipB, lo cual hace inferir que las tres son proteínas relacionadas evolutivamente [68], la comparación de las secuencias y el árbol evolutivo de las tres muestra que LplA y LipB están evolutivamente más cercanas entre sí (Figura 27).

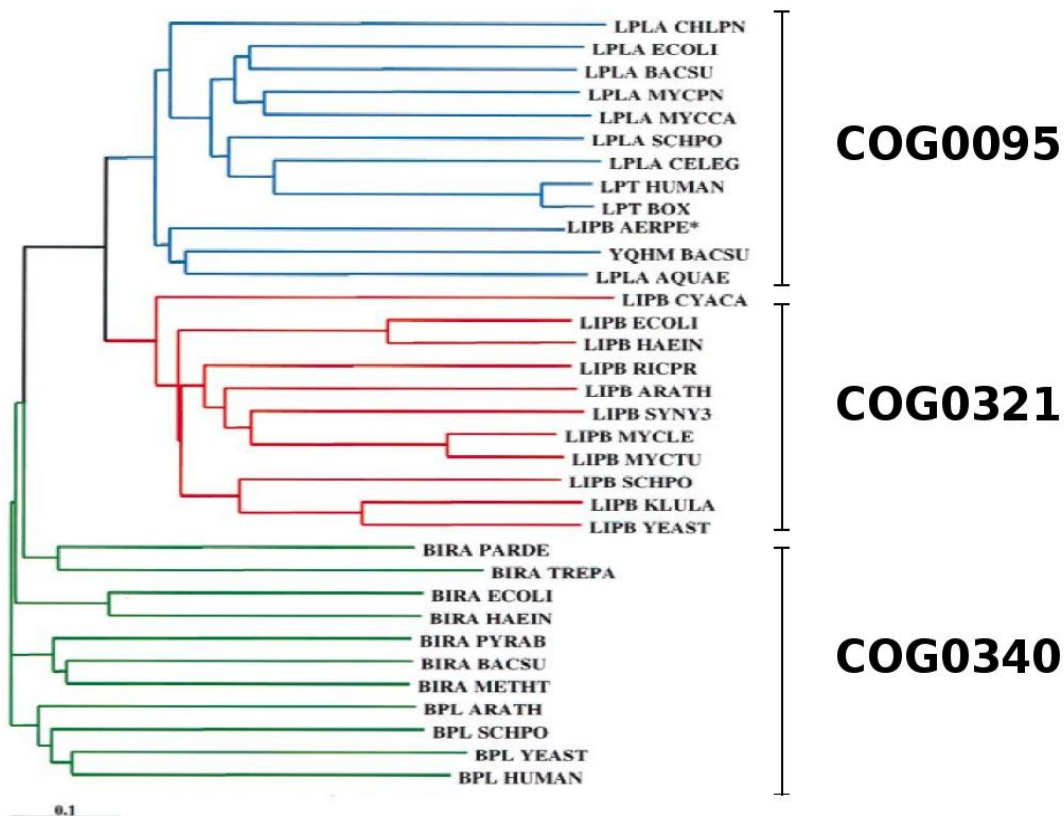


Figura 27: Árbol evolutivo de las proteínas LplA, LipB y BPL. LplA en azul, LipB en rojo y BPL en verde. Al lado de cada rama se muestra el COG al que pertenecen. Figura modificada de [68].

El alineamiento de las secuencias de BPL, LplA y LipB nos permite observar el residuo de lisina conservado en las tres familias de proteínas el cual se muestra de color amarillo sobre un fondo negro en la figura 28.

El patrón de residuos GRGR del loop que une las láminas $\beta 2$ y $\beta 3$ de BPL no se conservan en las otras dos proteínas, en ellas los residuos conservados entre las láminas equivalentes son RRxxGGGxxxxD y RGGxxTxHxPGQ para LplA y LipB respectivamente. Los residuos conservados encontrados por Reche

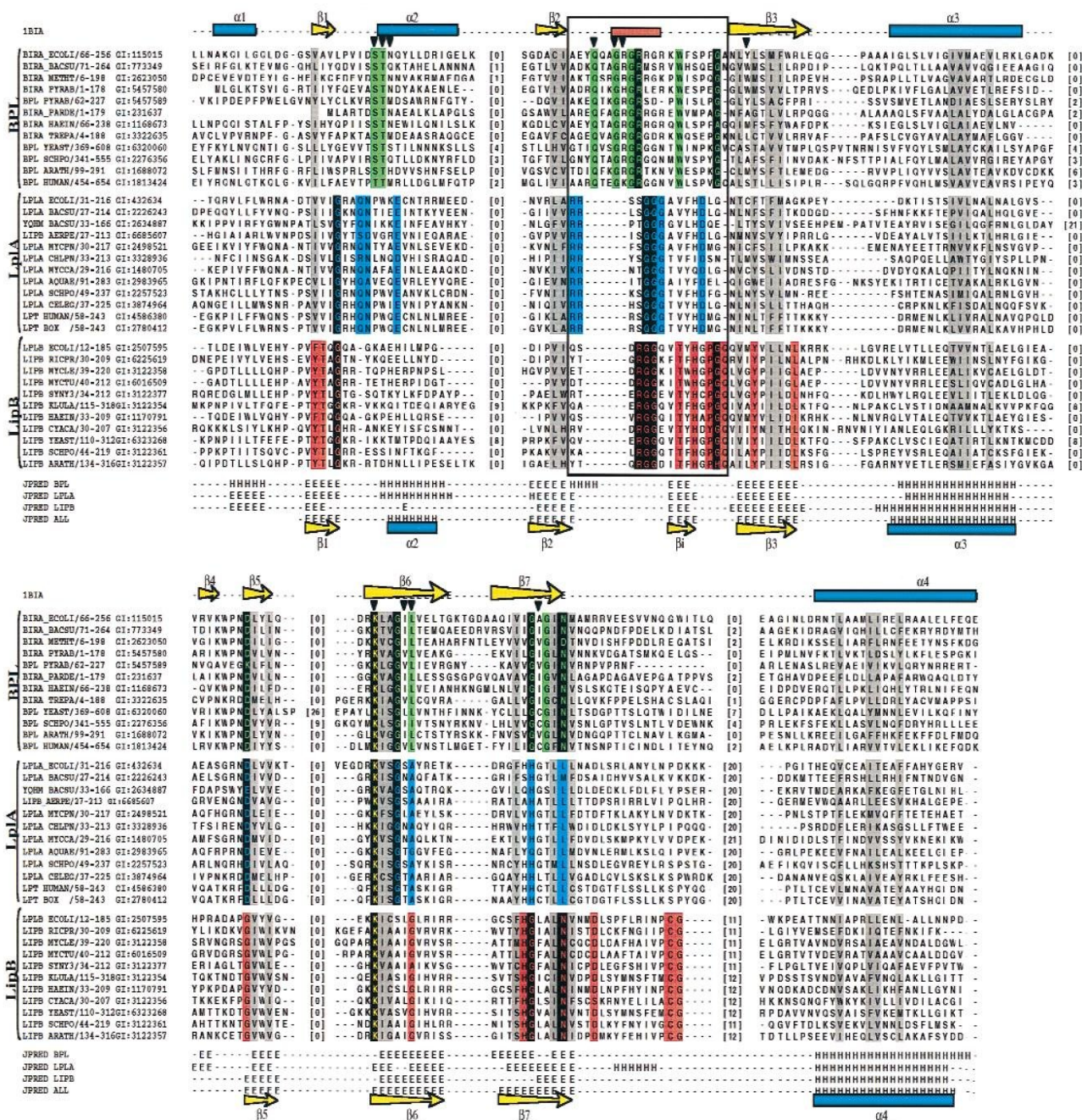


Figura 28: Alineación de las secuencias de las proteínas BPL, LplA y LipB, los residuos conservados a cada familia se muestran de verde, azul y rojo respectivamente. Figura tomada de [68].

[68] son reportados también por Kim *et al.* [40] quienes los nombran como motivo I RRxxGGGxV(F/Y)HD, motivo II KxxGxA donde se encuentra la lisina conservada en las tres familias de proteínas y el motivo III Hxx(L/M)LxxxLxxhL, estos tres motivos intervienen en la formación del sitio de unión del lipoil-AMP.

El COG0095 fue dividido en dos grupos, el grupo 1 con 18 elementos y el grupo 2 con 10 elementos. Los resultados de la búsqueda de los motivos reportados [68] [40] para la proteína LplA en las secuencias “semillas” de los grupos del COG0095 puede observarse en el cuadro 8.

Cuadro 8: Número de secuencias dentro de los grupos del COG0095 que presentan los motivos reportados para las proteínas LplA.

Motivo	Grupo 1	Grupo 2
RRxxGGGxV(F/Y)HD	17/18 (94.44 %)	0
KxxGxA	18/18 (100 %)	10/10 (100 %)
Hxx(L/M)LxxxLxxhL	13/18 (72.22 %)	0

Las proteínas recuperadas por MAST utilizando las matrices MEME de cada grupo fueron objeto de búsqueda de los motivos conservados de LplA, los resultados de esta búsqueda se observan en el cuadro 9. El número de secuencias recuperadas por las matrices del grupo 1 es de 18, mientras que para el grupo 2 es de 12.

Cuadro 9: Número de secuencias recuperadas por MAST y número de secuencias que contienen los motivos reportados para las proteínas LplA

Motivo	Grupo 1	Grupo 2
RRxxGGGxV(F/Y)HD	8/18 (44.44 %)	1/12 (8.33 %)
KhxGxA	12/18 (66.66 %)	6/12 (50 %)
Hxx(L/M)LxxxLxxhL	7/18 (38.88 %)	0

Únicamente las secuencias que conforman el grupo 1 del COG0095 así como las secuencias recuperadas por las matrices de este grupo presentan los tres motivos reportados para las proteínas LplA, el porcentaje de elementos que presentan los 3 motivos en las secuencias “semillas” y en las secuencias recuperadas por las matrices de este grupo se presentan en el cuadro 10. Las secuencias “semillas” del grupo 2 así como las secuencias recuperadas por MAST con las matrices del grupo 2 presentan de manera amplia el motivo II el cual contiene a la lisina conservada entre las tres familias de proteínas

(BPL, LplA y LipB), los porcentajes de las secuencias que presentan los motivos tanto en las secuencias “semillas” como en las recuperadas por las matrices de este grupo se presentan en el cuadro 10.

Cuadro 10: Porcentaje de elementos que presentan los motivos descritos para LplA [68], [40] tanto en las secuencias “semillas” como en las recuperadas por cada grupo del COG0095. R-Grupo: recuperados por MEME-MAST.

Motivo	Grupo 1	R-Grupo 1	Grupo 2	R-Grupo 2
RRxxGGGxV(F/Y)HD	94.44 %	44.44 %	0 %	8.33 %
KhxGxA	100 %	66.66 %	100 %	50 %
Hxx(L/M)LxxxLxxhL	72.22 %	38.88 %	0 %	0 %

5.3.2. COG0558. Phosphatidylglycerophosphate synthase PgsA

La síntesis de fosfatidil-glicerol (PG) y cardiolipina (CL) utiliza CDP-diacilglicerol (CDP-DAG), un intermediario central del metabolismo de los fosfolípidos. La reacción limitante en la síntesis de PG/CL es la transformación del CDP-DAG a fosfatidil-glicerofosfato, la cual es catalizada por la fosfatidil-glicerolfosfato sintasa (PG-P), en *Escherichia coli* la proteína PgsA es la encargada de llevar a cabo dicha reacción (Figura 29) [13].

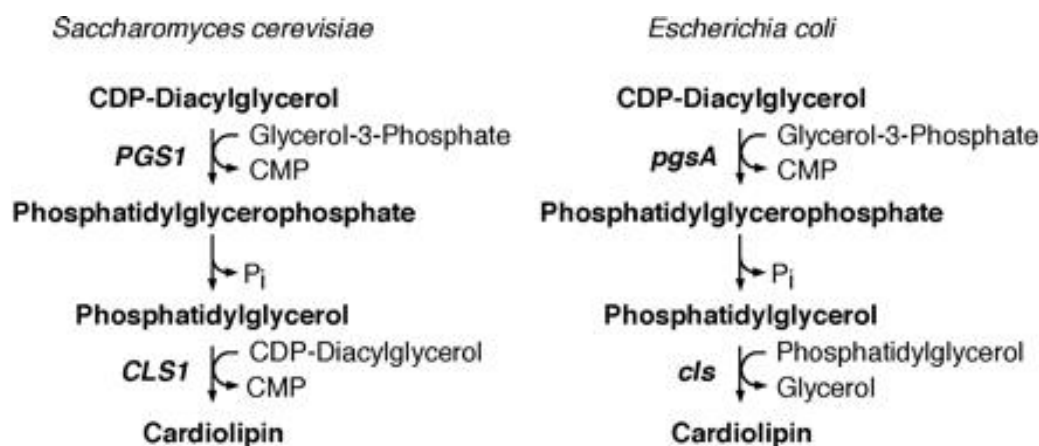


Figura 29: Ruta biosintética de la cardiolipina en *Escherichia coli* y *Saccharomyces cerevisiae*. Figura tomada de [13].

La PG-P tiene una similitud significativa con un motivo común a las enzimas que unen CDP-alcoholes y a las fosfatidiltransferasas. PgsA forma parte de una familia de proteínas que utilizan CDP-DAG como precursor para diferentes compuestos, entre las que se encuentran la fosfatidilcolina sintasa (Pcs), fosfatidilserina sintasa (Pss) y la fosfatidilinositol sintasa (Pis) (Figura 30) [74].

El alineamiento hecho por Sohlenkamp *et al.* [74] de las secuencias de aminoácidos de Pcs, Pss, PgsA y Pis (Figura 31) muestra en cuadros negros los motivos conservados en todas las CDP-alcohol fosfatidiltransferasas y en cuadros grises los residuos conservados dentro de cada familia. El motivo $DxxDGxxARx(8)GxxxDxxxDx$ se presenta en las proteínas, Pss, PgsA y Pis, la diferencia entre ellas radica en el último aminoácido, el cual no se conserva para las Pss, mientras que para las PgsA el último aminoácido lo constituye una lisina (K), y para las Pis una arginina (R).

El COG0558 fue dividido en dos grupos, el grupo 1 con 33 elementos y el grupo 2 con 15. La búsqueda

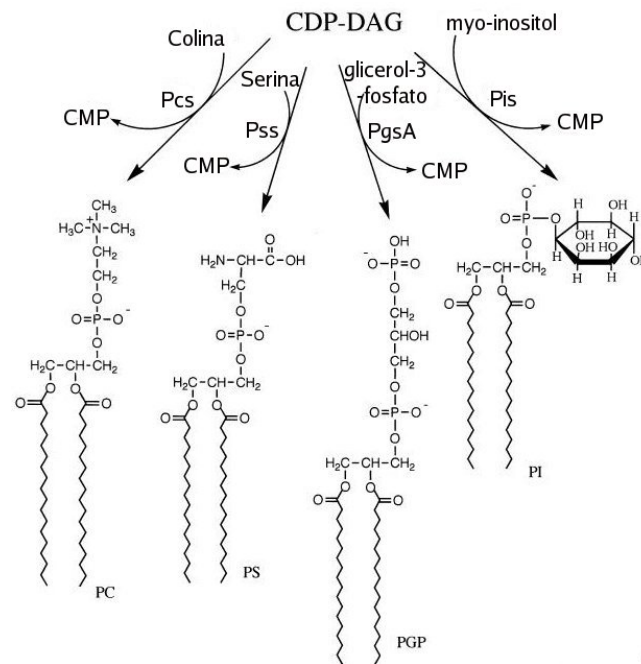


Figura 30: Enzimas que utilizan CDP-DAG como producto inicial de sus respectivas rutas. Pcs: fosfatidilcolina sintasa, Pss: fosfatidilserina sintasa, PgsA: fosfatidilglicerol-3-fosfato sintasa, Pis: fosfatidilinositol sintasa. Figura tomada de [74].

de los motivos reportados para las fosfatidiltransferasas en los grupos del COG0558 se muestran en el cuadro 11.

Cuadro 11: Número de secuencias en los grupos del COG0558 que presentan los motivos descritos para PgsA y Pis [74].

Proteína	Motivo	Grupo 1	Grupo 2
PgsA	DxxDGxxARx(8)GxxxDxxxDK	29/33 (87.87 %)	0
Pis	DxxDGxxARx(8)GxxxDxxxDR	1/33 (3.33 %)	12/15 (80 %)

Las secuencias que conforman al grupo 1 presentan en un 87 % el motivo DxxDGxxARx(8)GxxxDxxxDK, reportado por Sohlenkamp *et al.* como propio de la PgsA, mientras que en el grupo 2 ninguna secuencia presentó dicho motivo. El motivo DxxDGxxARx(8)GxxxDxxxDR de las proteínas Pis, se presenta en un 80 % en las secuencias del grupo 2, mientras que en las secuencias del grupo 1 se presentó en un 3.33 % (Cuadro 11). Las secuencias recuperadas por MAST usando las matrices MEME de los grupos del

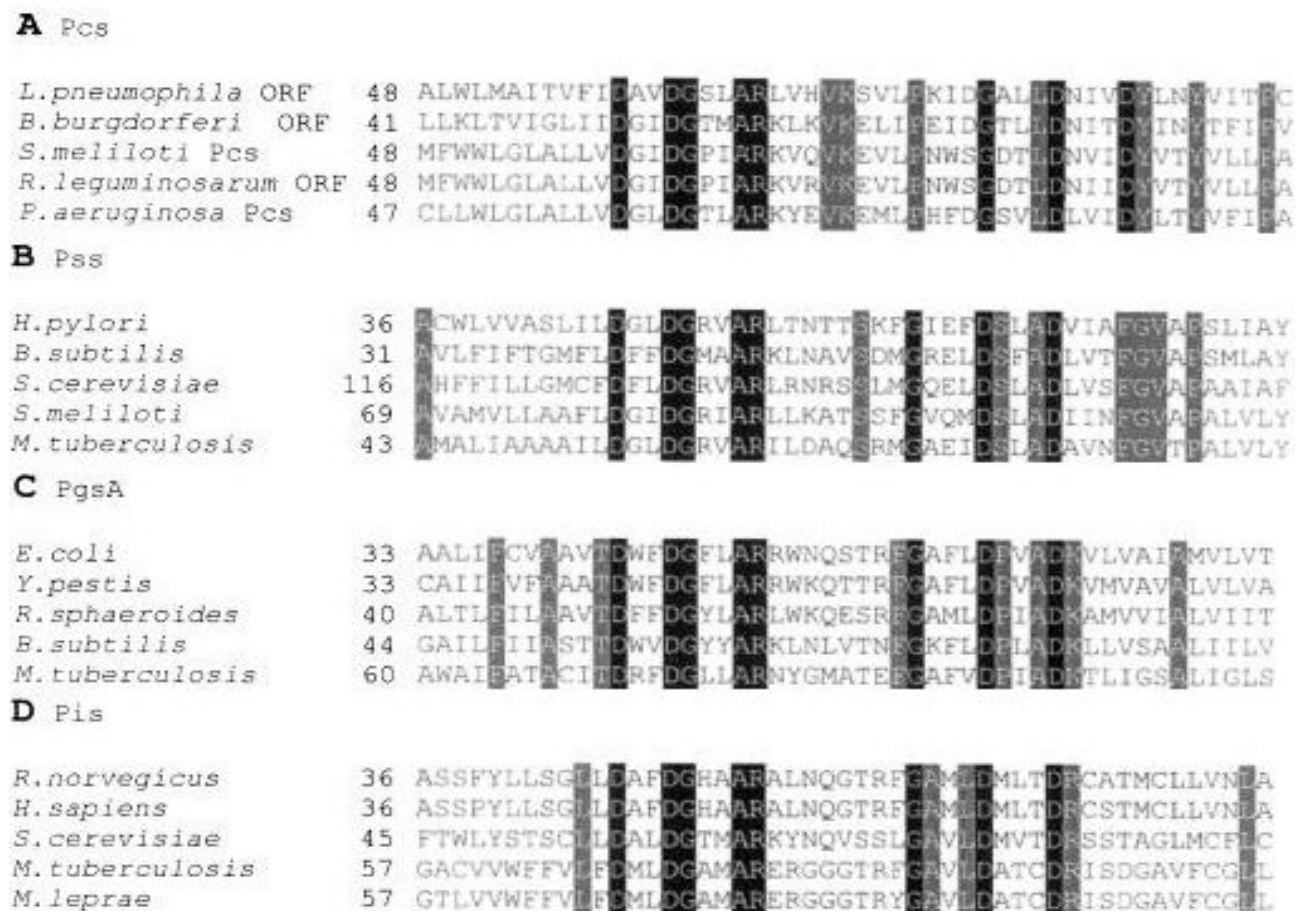


Figura 31: Alineamiento de los aminoácidos de las CDP-alcohol fosfatidiltransferasas, en cuadros negros se muestran los residuos conservados en todas las CDP-alcohol fosfatidiltransferasas, y en cuadros grises se muestran los residuos conservados dentro de cada familia. Figura tomada de [74].

COG0558, fueron analizadas para la búsqueda de los motivos de las proteínas PgsA y Pis. Las matrices del grupo 1 usadas por MAST recuperaron a 31 secuencias, mientras que las matrices del grupo 2 recuperaron 22 secuencias. En el grupo de secuencias recuperadas por el grupo 1, en 22 de ellas (71 %) se presenta el motivo reportado para la familia de proteínas PgsA, en las secuencias recuperadas por las matrices del grupo 2 sólo una (4.54 %) presentó el motivo de la proteína PgsA. El motivo de la proteína Pis se presentó en 4 de las 31 secuencias recuperadas por las matrices del grupo 1, mientras que en las secuencias recuperadas por las matrices del grupo 2, 11 (50 %) secuencias presentaron dicho motivo (Cuadro 12).

Cuadro 12: Número de secuencias recuperadas por MAST y número de secuencias que contienen los motivos descritos para PgsA y Pis [74].

Proteína	Motivo	Grupo 1	Grupo 2
PgsA	DxxDGxxARx(8)GxxxDxxxDK	22/31 (71.0 %)	1/22 (4.54 %)
Pis	DxxDGxxARx(8)GxxxDxxxDR	4/31 (12.90 %)	11/22 (50 %)

5.3.3. COG1393. Arsenate reductase and related proteins, glutaredoxin family

El sistema *ars* bacteriano confiere resistencia al arsenico. Esto se lleva a cabo por un sistema de bombeo el cual lo expulsa del interior de la célula al exterior. El transporte de arsenato al interior de las células como un oxianión es llevado a cabo por el sistema de transporte de membrana de fosfato del sistema Pit y Pst de *E. coli* (Figura 32). Una vez dentro de la célula el arsenato As(V) es reducido a arsenito As(III), lo cual es llevado a cabo por la arsenato reductasa ArsC. El sistema de bombeo de arsenito al exterior de la célula consiste de un complejo ATP-asa de dos componentes formado por los productos de los genes *arsB* y *arsA* [56], [77].

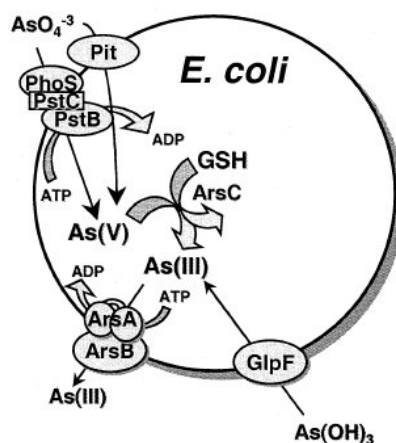


Figura 32: Transporte y reducción de arsenato en *E. coli*. PstB, PstC, PhoS son los tres componentes del sistema Pst, que junto con Pit introducen arsenato al interior de la célula. ArsB acoplada de ArsA lleva a cabo la expulsión de arsenito del interior de la célula. Mientras que ArsC reduce el arsenato a arsenito. GSH, glutatión. Figura tomada de [56].

La proteína ArsC se encuentra representada por tres clados [56] (Figura 33), dos de ellos en bacterias

y el tercero en levaduras, los cuales no comparten una ancestría común. El primer clado bacteriano referido como el grupo de glutaredoxin/ glutati6n (Grx/GSH), tiene como prototipo a la ArsC del pl6smido R773 de *E. coli*. El segundo grupo bacteriano llamado clado de tioredoxin (Trx), tiene al pl6smido p1258 de *Staphylococcus aureus* como prototipo. El tercer clado esta representado por la arsenato reductasa ACR2p de la levadura *Saccharomyces cerevisiae* [70].

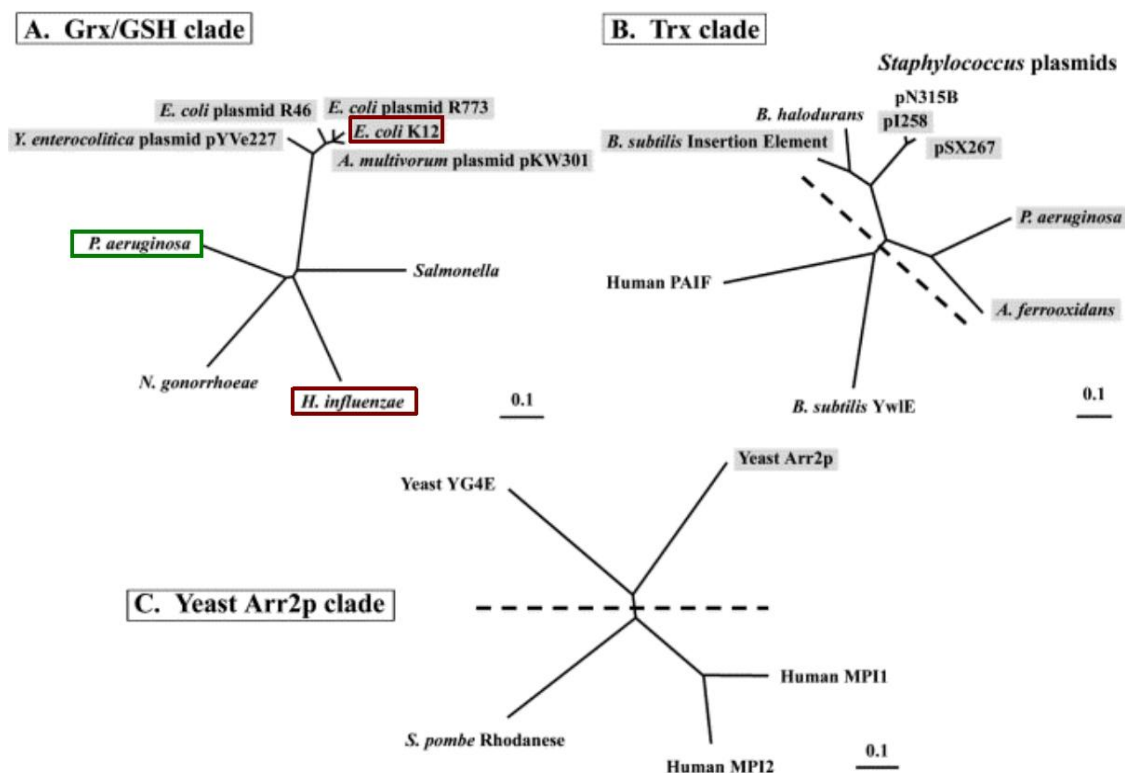


Figura 33: Tres familias de secuencias de arsenato reductasas relacionadas evolutivamente. **A.** Clado de Grx/GSH. **B.** Clado Trx. **C.** Clado de levaduras. Figura tomada de [56]. Encerrada en un rect6ngulo verde se encuentra la prote6na PAO1 (GI:15596147) de *Pseudomonas aeruginosa* que se encuentra en las secuencias semillas del grupo 2. Encerradas en un rect6ngulo rojo se encuentran la prote6na de *E. coli* (GI:16131375) y la prote6na de *Haemophilus influenzae* (GI:16272198) recuperadas por las matrices MEME-MAST del grupo 2.

Las arsenato reductasas bacterianas son peque1as enzimas monom6ricas citoplasm6ticas de 131 residuos de amino6cidos en *S. aureus*, y de 141 residuos en *E. coli*. La estructura terciaria de la ArsC del pl6smido R773 no tiene relaci6n alguna con la estructura terciaria de la ArsC de *S. aureus*, lo cual viene a robustecer la idea de or6genes independientes (convergencia) para los clados Grx/GSH y Trx de las arsenato reductasas. Una de las diferencias entre las arsenato reductasas bacterianas R773 y

p1258 es el uso de Grx/GSH o Trx respectivamente, en la reducción de arsenato a arsenito [56]. Para el clado bacteriano de ArsC que utiliza Trx en la reducción, se han reportado los motivos CX_5R y $CX_{(6/7)C}$ [54], mientras que para el clado que utiliza Grx/GSH se ha reportado el motivo HX_3CX_3R el cual es el sitio activo de la ArsC [76].

Spx es una proteína reguladora global de la transcripción, la cual puede regular de manera positiva o negativa. La forma en que la modulación se lleva a cabo por Spx, parece ser única entre los factores transcripcionales de procariotes, ya que no exhibe una actividad o dominio de unión al DNA, sino que la modulación la lleva uniéndose a la subunidad α de la RNA polimerasa (RNAP)[59], [89]. Debido a esta unión con la subunidad α de la RNAP, a Spx se le ha llamado factor “anti-alfa” [62]. Spx se encuentra altamente conservada entre un gran número de bacterias Gram-positivas [58], siendo una proteína monomérica [62]. En la regulación negativa, Spx se une a la Tyr263 en el extremo C de la subunidad α de la RNAP, con lo cual las interacciones entre los activadores transcripcionales y la RNAP es bloqueada (Figura 34).

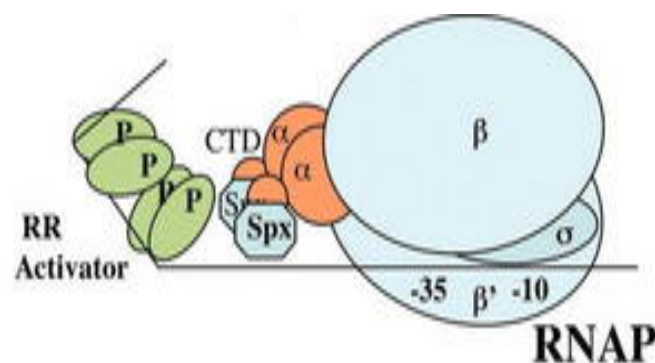


Figura 34: Bajo estrés de disulfuro Spx se acumula y se pega al dominio C de la subunidad α de la RNAP y bloquea la interacción activador-RNAP. Figura tomada de [89].

Ejemplos donde la actividad de los activadores transcripcionales se ve reducida por la interacción Spx-RNAP, es la modulación negativa que ejerce sobre el operón *srf*, el cual activa la transcripción de los genes requeridos para el desarrollo genético de la competencia celular [58], y ResD, el cual induce a los genes involucrados en la adaptación a niveles bajos de oxígeno [89]. La regulación positiva que ejerce Spx se da principalmente en genes que intervienen en el mantenimiento de la homeostasis del tiol. En condiciones tiol-oxidativas, Spx activa directamente la transcripción de los genes *trxA* y *trxB*,

los cuales codifican para tioredoxin y tioredoxin reductasa, respectivamente [58], [89]. La activación se da mediante la interacción de la subunidad α de la RANP con la forma activa de Spx, la cual tiene una unión disulfuro intramolecular en el motivo CXXC, el cual se encuentra en el extremo N [64], [59]. El motivo CXXC está altamente conservado entre Spx y sus homólogos [62], y es un probable sitio del sensado del estado redox intracelular, donde los efectos del estrés oxidativo pueden modular la actividad y/o estabilidad de Spx [58]. Una búsqueda con CD-BLAST reveló que la proteína Spx de *B. subtilis* es miembro de la familia de las arsenato reductasas ArsC, y exhibe una homología significativa con la estructura secundaria de la ArsC del plásmido R773 de *E. coli* [89] (Figura 35). Además del motivo CXXC, Zuber [89] identifica otras regiones con identidad de secuencia entre los ortólogos de Spx (Figura 35).

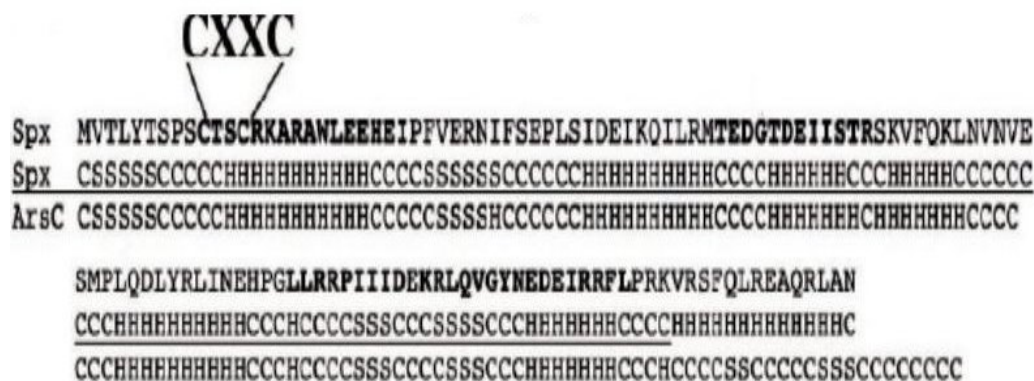


Figura 35: Secuencias de Spx, se resalta el motivo **CXXC**. Las regiones con identidad de secuencia entre los ortólogos de Spx se muestran en **negritas**. La segunda y tercera secuencia son las comparaciones de estructura secundaria (**H**: α -hélices, **S**: laminas- β , **C**: espiral) entre Spx y ArsC. Figura tomada de [89].

La sobreposición de la estructura de Spx de *B. subtilis* resuelta cristalográficamente por Newberry *et. al.* [62] y de la estructura de la ArsC de *E. coli*, resulta por Martin *et. al.* [52], muestra que son homólogos estructurales (Figura 36).

El COG1393 fue dividido en tres grupos. El grupo 1 esta constituido con 15 secuencias, el grupo 2 con 10 secuencias y el grupo 3 con 7 secuencias. La búsqueda de los motivos reportados para las Arsc del clado Grx/GSH [54], [76] y para las proteínas Spx [89] en las secuencias semillas del COG1393 se observan en el cuadro 13.

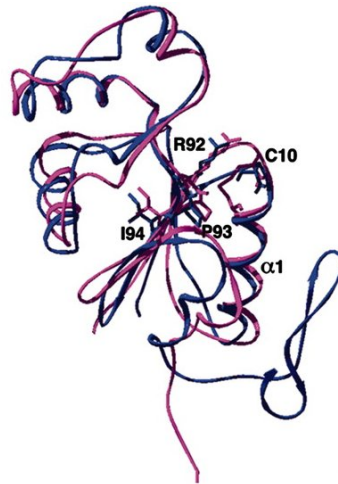


Figura 36: Sobreposición de las estructuras de la proteína Spx y ArsC de *B. subtilis* y *E. coli*, respectivamente. Spx se muestra en rosa y ArsC en azul. Figura tomada de [62].

Cuadro 13: Número de secuencias en los grupos del COG1393 que presentan los motivos descritos para ArsC [54], [76] y Spx [89].

Proteína	Motivo	Grupo 1	Grupo 2	Grupo 3
ArsC	Hx ₃ Cx ₃ R	0	9/10 (90 %)	0
Spx	CxxC	7/15 (46.6 %)	1/10 (10 %)	7/7 (100 %)
Spx modificado de [89]	E(D/N)G(T/F)xx(I/L)(I/F)STR	0	0	5/7 (71.42 %)
Spx modificado de [89]	(L/I)(L/I)RRP(I/L)Ixxx(K/N)	0	0	7/7 (100 %)
ArsC clado Trx [54]	Cx ₅ R	0	0	0
ArsC clado Trx [54]	Cx _(6/7) C	1/15 (6.66 %)	0	0

El motivo reportado para la ArsC del clado Grx/GSH se presenta únicamente en el grupo 2, cuyas secuencias están anotadas como arsenato reductasas por el GenBank. Los 3 motivos reportados para la proteína Spx se encuentran solamente en el grupo 3 (cuyas secuencias están anotadas como reguladores transcripcionales por el GenBank), en donde aparecen en más del 50 % de las secuencias; el grupo 1 y 2 presentaron solamente el motivo CxxX, con una aparición del 46.6 % y 10 %, respectivamente. Para los otros motivos de la Spx el grupo 1 y grupo 2 no los presentaron. La búsqueda de los motivos

pertenecientes a la ArsC del clado Trx arrojo como resultado que una secuencia del grupo 1 (6.66 %) la presenta, estando ausente en los grupos restantes.

Las secuencias recuperadas por MAST usando las matrices MEME de cada grupo fueron objeto de búsqueda de los motivos de ArsC y Spx. Las matrices del grupo 1 recuperaron 15 secuencias, las del grupo 2, 16 secuencias y las del grupo 3, 12 secuencias (Cuadro 14).

Cuadro 14: Número de secuencias recuperadas por MAST y número de secuencias que contienen los motivos de ArsC y Spx.

Proteína	Motivo	Grupo 1	Grupo 2	Grupo 3
ArsC	Hx ₃ Cx ₃ R	0	15/16 (93.75 %)	0
Spx	CxxC	2/15 (13.33 %)	0	9/12 (75 %)
Spx modificado de [89]	E(D/N)G(T/F)xx(I/L)(I/F)STR	0	0	2/12 (16.66 %)
Spx modificado de [89]	(L/I)(L/I)RRP(I/L)Ixxx(K/N)	0	0	4/12 (3.33 %)
ArsC clado Trx [54]	Cx ₅ R	0	0	0
ArsC clado Trx [54]	Cx _(6/7) C	0	0	0

Las secuencias recuperadas por las matrices del grupo 1 son las únicas en presentar el motivo ArsC del clado Grx/GSH. De los tres motivos reportados para Spx el motivo CxxC se presenta en el grupo 1 en un 13.33 %, mientras que en el grupo 3 aparece un 75 %; los otros dos motivos de Spx únicamente aparecen en el grupo 3 con un porcentaje menor al 17 %. El motivo de la ArsC del clado Trx no aparece en ninguna secuencia recuperada por las matrices de los grupos. La comparación de los porcentajes de aparición de los motivos en las secuencias semillas y en las secuencias recuperadas puede verse en el cuadro 15.

Cuadro 15: Porcentaje de elementos que presentan los motivos de ArsC y Spx, tanto en las secuencias “semillas” como en las recuperadas por las matrices de cada grupo. R-Grupo: recuperados por MEME-MAST.

Motivo	Grupo 1	R-Grupo 1	Grupo 2	R-Grupo 2	Grupo 3	R-Grupo 3
Hx ₃ Cx ₃ R	0	0	90 %	93.75 %	0	0
CxxC	46.6 %	13.33 %	10 %	0	100 %	75 %
E(D/N)G(T/F)xx(I/L)(I/F)STR	0	0	0	0	71.42 %	16.66 %
(L/I)(L/I)RRP(I/L)Ixxx(K/N)	0	0	0	0	100 %	3.33 %
Cx ₅ R	0	0	0	0	0	0
Cx _(6/7) C	6.6 %	0	0	0	0	0

5.3.4. COG0147. Anthranilate/para-aminobenzoate synthases component I

El producto final de la ruta del shikimato es el corismato (Figura 37), el cual participa en la producción de diferentes biomoléculas como lo son los aminoácidos aromáticos (triptófano, fenilalanina y tirosina), el folato, la menaquinona y la ubiquinona, los sideróforos, el cloranfenicol, las vitaminas K y E, la coenzima Q, las plastoquinonas entre otros metabolitos [29].

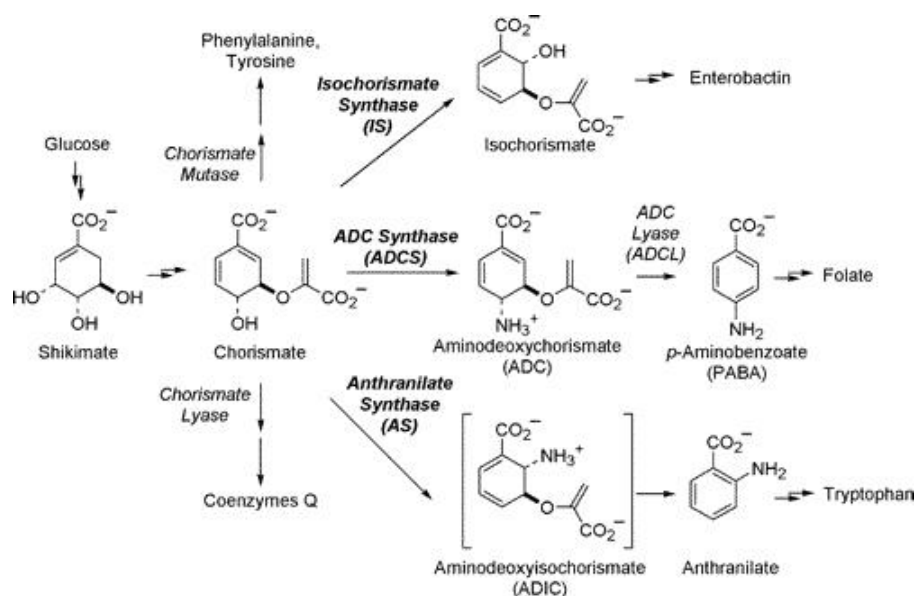


Figura 37: Esquema de la ruta de biosíntesis del shikimato, folato y triptófano. Figura tomada de [29].

La biosíntesis de triptófano se lleva a partir de la condensación del fosfoenolpiruvato (PEP) y la eritrosa 4-fosfato (E4P), los cuales son procesados a corismato. Posteriormente el corismato es convertido a antranilato [35], reacción llevada a cabo por el complejo multidominio antranilato sintasa (AS), con lo cual se inicia la ruta de biosíntesis del triptófano (Figura 38).

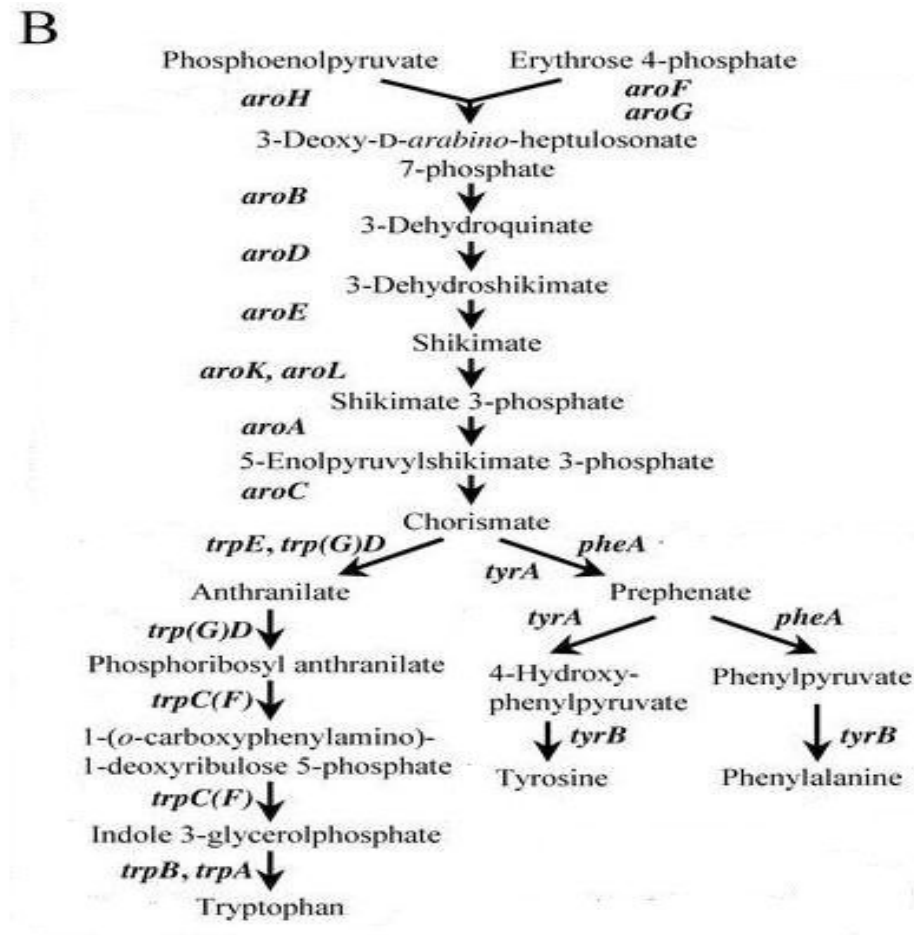


Figura 38: Esquema de la ruta de biosíntesis del triptófano en *E. coli*. Figura modificada de [35].

La AS se encuentra en las bacterias como un heterotetrámero, el cual es producto de los genes *trpE* y *trpG*. El gene *trpG* codifica para una amidotransferasa, mientras que el producto de *trpE* es una enzima bifuncional. TrpE contiene dos dominios independientes que separados físicamente pueden mostrar actividad de ADIC sintasa y ADIC liasa.

Se han reportado dos familias de proteínas parálogos de TrpE, las cuales llevan a cabo reacciones similares. Una de estas familias está involucrada en la biosíntesis de fenazina, el cual es un metabo-

lito secundario con actividad antibiótica importante. PhzE usa corismato como sustrato, el cual es convertido en ADIC, el intermediario no difusible en la síntesis de antranilato. La segunda familia de proteínas parálogos de TrpE es PabB, la cual participa en la biosíntesis del 4-amino-4-deoxycorismato (ADC), precursor del metabolito primario ácido para-aminobenzoico, (PABA) el cual forma parte de la ruta de biosíntesis del ácido fólico (Figura 39).

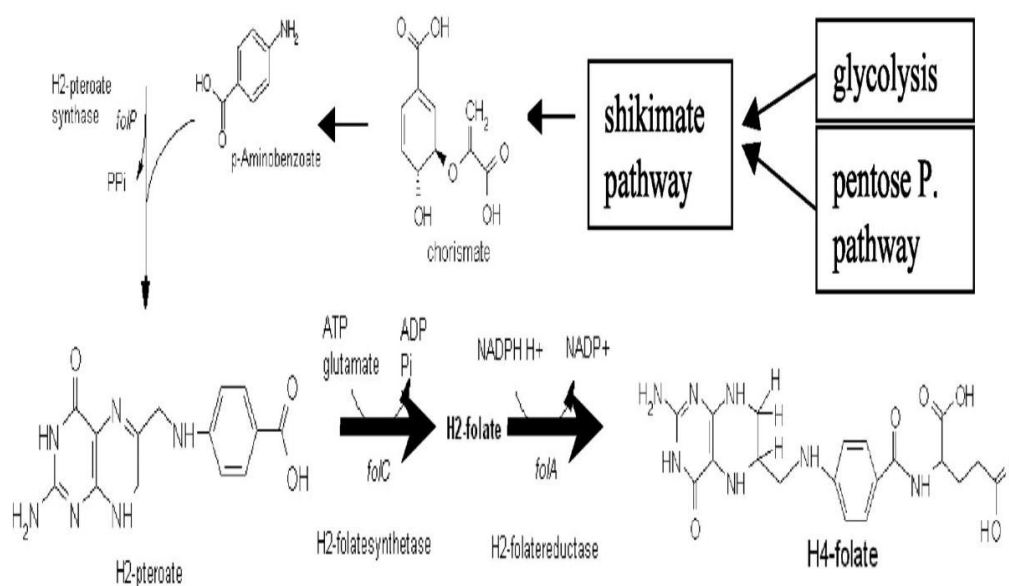
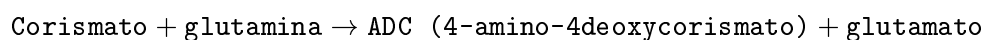


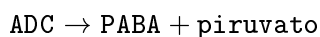
Figura 39: Ruta de biosíntesis del folato. Figura modificada de [78].

PABA es generado en dos reacciones [37],

Reacción 1



Reacción 2



En *E. coli*, los productos de los tres genes *pabA*, *pabB* y *pabC* han mostrado que pueden llevar a cabo estas dos reacciones [37]. La para-aminobenzoato sintasa encargada de la conversión de corismato a ADC se encuentra codificada por los genes *pabB* y *pabA*, los cuales son parálogos de *trpE* y *trpG* [37],

respectivamente. En muchas bacterias *pabA* y *pabB* se encuentran separados, pero también se pueden encontrar fusionadas como en los actinomicetos [37].

El COG0147 el cual agrupa a las proteínas TrpE y PabB fue dividido en dos grupos. El grupo 1 con 34 secuencia y el grupo 2 con 21. Al no encontrar reportes en la bibliografía sobre motivos específicos que diferencien a estas dos familias de proteínas, se llevo a cabo un análisis tomando en cuenta las anotaciones hechas por el GenBank para estas secuencias (Cuadro 16).

Cuadro 16: Número de secuencias con las anotaciones de antranilato sintasa o para-aminobenzoato sintasa que aparecen en los grupos del COG0147

Proteína	Anotación	Grupo 1	Grupo 2
TrpE	antranilato sintasa	30/34 (88.83 %)	3/21 (14.28 %)
PabB	para-aminobenzoato sintasa	1/34 (2.94 %)	12/21 (57.14 %)

Las secuencias recuperadas por MAST usando las matrices MEME de cada grupo, también fueron objeto del mismo análisis . Las matrices del grupo 1 recuperaron a 29 secuencias, mientras que las matrices del grupo recuperaron a 17 secuencias (Cuadro 17).

Cuadro 17: Número de secuencias recuperadas y con las anotaciones de antranilato sintasa o para-aminobenzoato sintas que aparecen en las secuencias recuperadas por las matrices MEME de cada grupo

Proteína	Anotación	Grupo 1	Grupo 2
TrpE	antranilato sintasa	22/29 (75.86 %)	1/17 (5.88 %)
PabB	para-aminobenzoato sintasa	0	13/17 (76.47 %)

La comparación de las anotaciones de antranilato sintasa y para-aminobenzoato sintasa que aparecen en las secuencias de los grupos “semilla” y de las secuencias recuperadas por MEME-MAST, muestran que en mas de un 75 % de las secuencias del grupo 1 y las correspondientes secuencias recuperadas por sus matrices, aparece la anotación de antranilato sintasa; mientras que en las secuencias “semillas” del grupo 2 y las recuperadas de este grupo, dicha anotación aparece en menos de un 15 %. La anotación de 3A para-aminobemzoato sintasa aparece en el grupo 1 y sus secuencias recuperadas menos de un

3 %, mientras que en grupo 2 y sus secuencia recuperadas, la anotación de para-aminobenzoato sintasa aparece en mas del 50 % (Cuadro 18).

Cuadro 18: Comparación del porcentaje de aparición de las anotaciones de antranilato sintasa o para-aminobenzoato sintas que aparecen en las secuencias “semillas” y recuperadas por las matrices MEME de cada grupo del COG0147. R-Grupo:recuperado or MEME-MAST

Proteína	Anotación	Grupo 1	R-Grupo 1	Grupo 2	R-Grupo 2
TrpE	antranilato sintasa	88.83 %	75.86 %	14.28 %	5.88 %
PabB	para-aminobenzoato sintasa	2.94 %	0	57.14 %	76.47 %

6. Discusión

La base de datos de proteínas ortólogas COGs es un repositorio de información generado y mantenido por el National Center for Biotechnology Information (NCBI). Dicha base de datos ha sido empleada para realizar la asignación de funciones biológicas de manera extensa en los proyectos de secuenciación genómica; sin embargo la confiabilidad de las asignaciones realizadas por esta herramienta se ve disminuida en la medida en que la presencia de proteínas parálogas dentro de los COGs va en aumento. Los grupos obtenidos en los 584 COGs con presencia de parálogos, tratan de definir a subfamilias de proteínas, que aunque se encuentran agrupadas dentro del mismo COG, biológicamente tienen funciones diferenciables.

La documentación manual que se realizó de los grupos (MOGs) de cuatro COGs (COG0095, COG0558, COG1393, COG0147) de los 584 COGs finales, mostró que dichos grupos definen a familias de proteínas que tienen funciones biológicas distintas. Estos cuatro COGs representan una cobertura del 0.68 % con respecto de los 584 COGs analizados. El número de COGs documentados de esta manera es un proceso lento, ya que involucra la búsqueda de las referencias bibliográficas necesarias que puedan documentar y validar la separación de las proteínas en grupos diferentes, los cuales representan a subfamilias con funciones biológicas distintas. El incremento en el número de COGs documentados y validados, cuyos grupos representarán a familias de proteínas funcionalmente distintas, se realizó utilizando las bases de datos Pfam y PROSITE. En el análisis realizado con PROSITE de los 584 COGs finales divididos, en 37 COGs se encontraron firmas específicas para alguno de los grupos contenidos en ellos, lo cual representa un 6.33 % (Figura 25) de cobertura. En el análisis realizado con Pfam la cobertura de COGs para los cuales se encontraron perfiles HMM distintos para alguno de los grupos contenidos en ellos, fue de 42 COGs, lo cual representa un 7.19 % (Figura 23). Los motivos hallados por PROSITE y Pfam tienen una frecuencia de aparición igual o mayor al 50 % en un grupo y menor al 50 % en el otro. Para los grupos de 8 de los 584 COGs analizados, se encontraron firmas PROSITE y perfiles HMM específicos que documentan a los mismos grupos dentro del COG. Estos perfiles y firmas coinciden en documentar a los mismos dominios o proteínas. Estos 8 COGs donde sus grupos son documentados tanto por Pfam como por PROSITE representan un 1.3 % del total de los 584

COGs analizados (Figura 40).

8 COGs de 584 analizados que presentan perfiles PFAM y firmas PROSITE para los mismos grupos.

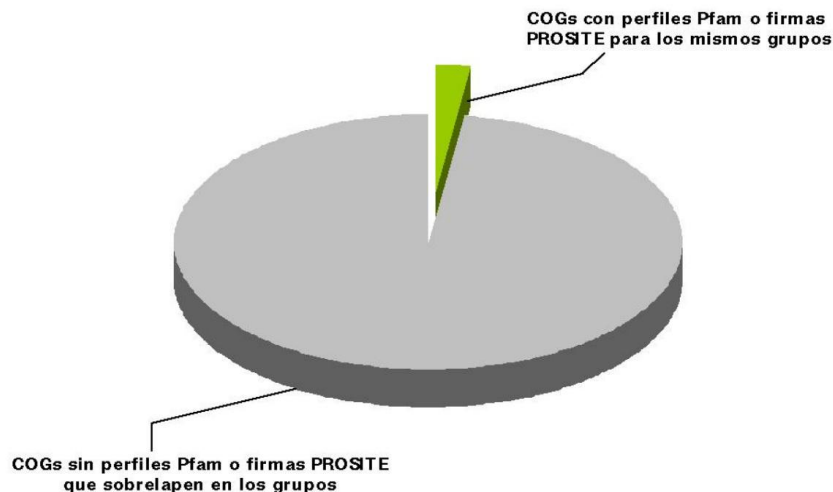


Figura 40: Para 8 COGs de los 584 analizados, sus grupos coincidieron en estar anotados por Pfam y PROSITE. Estos 8 COGs representa un 1.3 % de los 584 COGs finales.

Un ejemplo de COG documentado por PROSITE y Pfam es el COG0262, en el cual tanto PROSITE como Pfam encuentran en el grupo 1 la firma PS00075 y el perfil PF00186, respectivamente (Figura 41). Tanto la firma PS00075 y como el perfil PF00186 se corresponden a la familia de proteínas de la dihidrofolato reductasa, la cual participa en la biosíntesis *de novo* de la glicina y las purinas.

La suma de COGs anotados por ambas bases de datos, para los cuales se encontraron perfiles HMM o firmas PROSITE específicas para alguno de sus grupos fue de 71 COGs, los cuales representan una cobertura del 12.15 %, con respecto a los 584 COGs totales (Figura 42).

El número de COGs donde sus grupos fueron documentados manualmente más los COGs anotados por PFAM y PROSITE es de 75, lo que representa un 12.84 % de cobertura.

El número de COGs cuyos grupos fueron documentados tiene un incremento mientras mayor sea el número de dominios o familias anotadas en la base de datos contra la cual se compara. La documentación manual de los COGs parte de una búsqueda de información dentro de un universo amplio de referencias, dicho proceso involucra la lectura y selección de las referencias que documenten apropiadamente cada familia de proteínas. Las bases de datos PFAM y PROSITE tienen registrados y

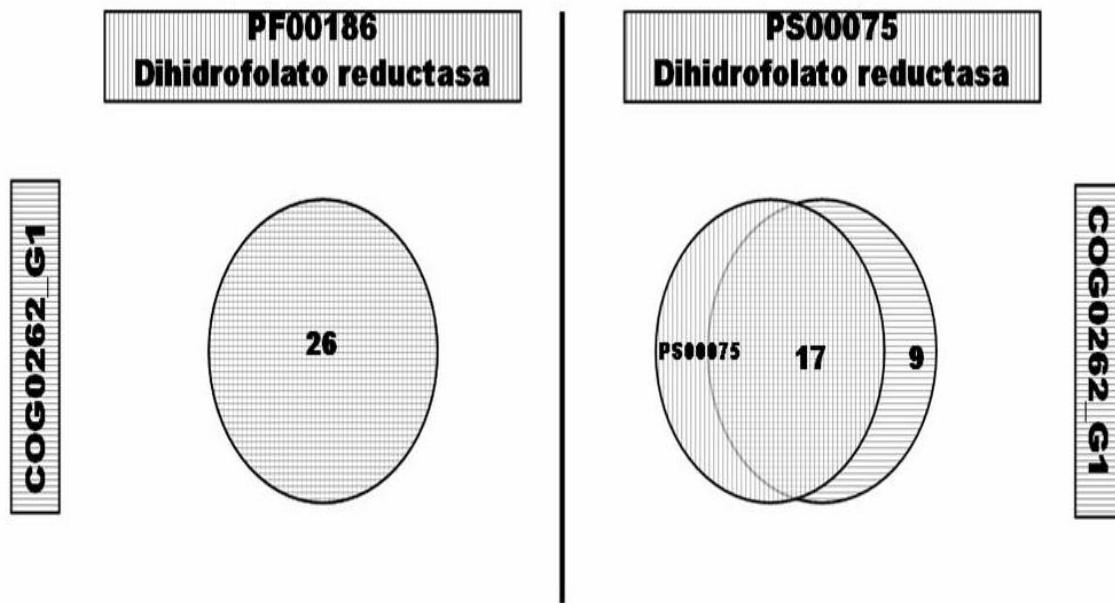


Figura 41: El perfil Pfam PF00186 se presentó en las 26 secuencias que conforman el grupo 1 del COG0262, mientras que la firma PROSITE PS00075 se presentó en 17 de las 26 secuencias del grupo 1. Tanto el perfil PF00186 como la firma PS00075 documentan a la misma familia de proteínas, la dihidrofolato reductasa. Los círculos con líneas verticales representan al perfil Pfam y a la firma PROSITE, mientras que los círculos con líneas horizontales representan a los grupos.

documentados motivos o perfiles de familias de proteínas, además de software que facilita la búsqueda de proteínas relacionadas. Esta diferencia en la búsqueda de información se hace patente al pasar de un 0.68 % de COGs cuyos grupos fueron documentados manualmente a un 6.33 % en la cobertura de COGs cuyos grupos fueron documentados utilizando PROSITE, y utilizando Pfam obtenemos un 7.19 %. Cuando hacemos la comparación de los COGs cuyos grupos fueron anotados manualmente contra los anotados por Pfam y PROSITE juntas, pasamos de un 0.68 % a un 12.15 % (Figura 43)

La diferencia observada al comparar el porcentaje de COGs cuyos grupos fueron documentados por PROSITE o Pfam se debe al tamaño de la base de datos y al número de familias que se encuentran almacenadas en ellas. PROSITE tiene almacenada 2,038 entradas de firmas y perfiles de dominios o familias de proteínas, mientras que Pfam contiene 8,957 perfiles HMM de familias o dominios de proteínas. El incremento en el número de COGs cubiertos depende del número de familias caracterizadas



Figura 42: 71 COGs fueron anotados en total por las 2 bases de datos PFAM y PROSITE, lo cual representa un 12.15 % (A). Mientras que para los 513 (B) (87.84 %) COGs restantes no se encontro perfil HMM o firma PROSITE alguna.

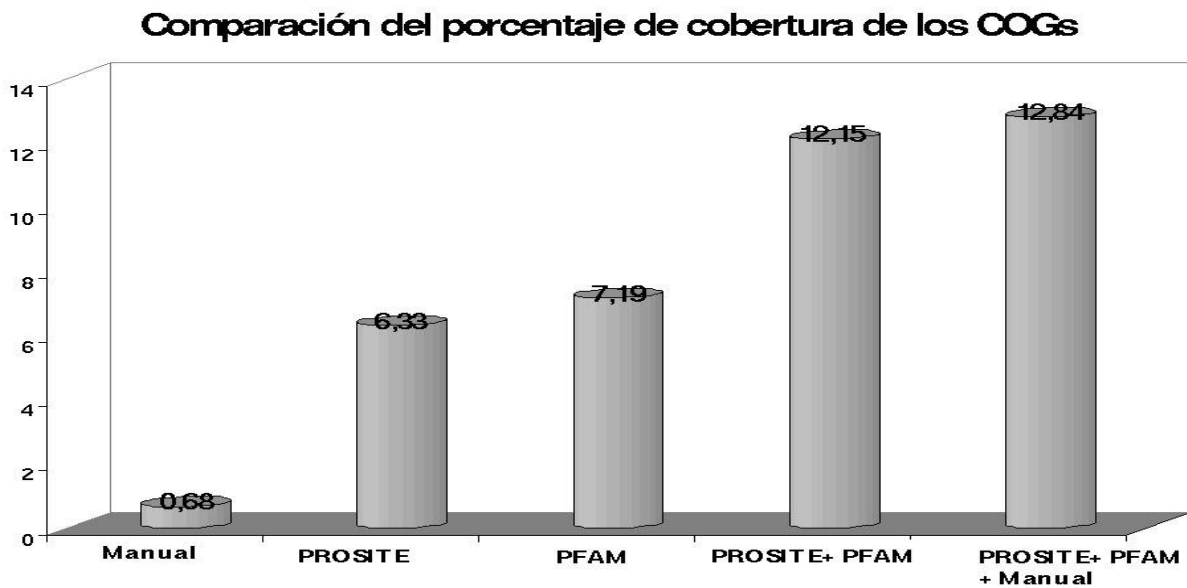


Figura 43: Diferencia en el porcentaje de cobertura de los COGs.

y almacenadas en las bases de datos en las cuales se busca.

El alto número de COGs (509) cuyos grupos no fueron documentados de manera específica, puede deberse a que no contamos con la información necesaria que caracterize a las diferentes familias de

proteínas. Los 4 COGs documentados manualmente no fueron cubiertos por Pfam ni por PROSITE, lo cual nos indica que existe poca información sobre el universo de las proteínas, sus correspondientes subfamilias y sus funciones. El COG0147 dividido en dos grupos, el de las antranilato sintasas y las para-aminobenzoato sintasas, no presento perfiles HMM que los diferenciara; ambos grupos presentaron el perfil HMM correspondiente a la antranilato sintasa. Sin embargo como fue documentado manualmente (ver Sección 5.3) estos dos grupos representan a subfamilias de proteínas que participan en rutas biosintéticas distintas.

Los miembros de una familia de proteínas tienen generalmente una función bioquímica en común, pero frecuentemente, uno o más subgrupos evolucionan para adquirir una función ligeramente diferente. Una subfamilia puede ser vista como un grupo de proteínas, las cuales son el resultado de una línea particular de evolución [1]. La detección de estos ligeros cambios funcionales dentro de las familias de proteínas es importante para lograr una anotación funcional más correcta. La exploración de las relaciones de bidireccionalidad entre las secuencias de las proteínas de una familia, y el uso de estas relaciones bidireccionales para la formación de grupos (subfamilias) utilizando el programa MCL, logró dividir a los 584 COGs finales con presencia de parálogos, en 2,044 grupos (MOGs). 75 COGs de los 584 analizados, presentaron para al menos alguno de sus grupos perfiles Pfam o firmas PROSITE específicas, que los caracterizan como subfamilias dentro de los COGs. Las familias de proteínas ortólogas representadas por estos COGs (con presencia de parálogos), así como sus subfamilias, son agrupadas de manera conjunta en la base de datos de los COGs, por lo que un mayor conocimiento de estas subfamilias involucra primero, el conocimiento biológico de la presencia de subgrupos dentro de la familia de proteínas determinada, y segundo, un método que nos permita definir a los grupos existentes dentro de la familia, generalmente dicho método utiliza las distancias filogenéticas entre las proteínas. El enfoque utilizado en este trabajo permite realizar la exploración de una familia de proteínas para la búsqueda de subgrupos, si los hay, sin tener conocimiento biológico alguno sobre las subfamilias existentes dentro del grupo, además de no recurrir a la utilización de cladogramas para la separación de las subfamilias, los cuales son demandantes en tiempo durante sus análisis

La utilización de nuestra metodología para la división de familias de proteínas con presencia de parálogos en subgrupos, puede ser utilizada en otras bases de datos como Pfam. Con el objetivo de evaluar la potencial utilidad de dicha metodología en la subdivisión de subgrupos Pfam, la familia de proteínas de las antranilato sintasa y para-aminobenzoato sintasa que se encuentran agrupadas en el perfil PF04715, fueron analizadas con nuestro enfoque. La división que hizo MCL explorando las relaciones de bidireccionalidad que se dan entre las proteínas que conforman dicho perfil, fue de dos grupos al usar una granularidad (I) de 5.0. En el grupo 1 conformado por 112 proteínas, 98 (87.5%) de ellas presentaron la anotación de antranilato sintasa. Para el grupo 2 conformado de 77 secuencias, 45 (58.44%) de ellas presentaron la anotación de para-aminobenzoato sintasa. La exploración del contexto genómico de algunas secuencias que conforman estos grupos, muestra claramente la participación en rutas biosintéticas diferentes para cada subfamilia encontrada por nuestro método (ver Figuras 44 y 45). Otro ejemplo de división en el Pfam con nuestro enfoque, es el de la familia de las CDP-alcohol fosfatidiltransferasas. El perfil Pfam que representa a dicha familia es el PF01066, el cual después de aplicar nuestra metodología usando una granularidad de $I=2$, se dividió en 10 grupos. De los 10 grupos en que fueron divididas las fosfatidiltransferasas, cuatro de ellos agrupan a las subfamilias representadas por las proteínas fosfatidilglicerolfosfato sintasa (PgsA), fosfatidilserina sintasa (Pss), fosfatidilinositol sintasa (Pis) y fosfatidilcolina sintasa (Pcs), los cuales son el grupo 1, grupo 2, grupo 3 y grupo 6 respectivamente. El grupo 4 obtenido con nuestra división agrupa a proteínas que en su mayoría pertenecen a organismo eucariotes, como hongos, nemátodos y moscas. Los motivos reportados por Sohlenkamp *et al.* [74] (ver Sección 5.3.2) para las proteínas PgsA, Pss, Pis, Pcs se encuentran presentes en las secuencias de los respectivos grupos en que fueron separadas estas familias (ver Figuras 46, 47, 48, 49 y 50).

La división que pueda hacerse de una familia de proteínas en subfamilias utilizando nuestro enfoque, depende del método utilizado para la formación del grupo, así como del número de elementos con que dicho grupo cuente. Esto se observa en los diferentes valores de granularidad (I) empleados para la división de la familia de proteínas que agrupa a las antranilato sintasa y las para-aminobenzoato sintasa. Para la división de esta familia en la base de datos COG se utilizó un valor de 2.0, con el cual

se obtuvieron dos grupos, los cuales representan a cada subfamilia por separado, mientras que para la base de datos Pfam, el valor de granularidad (I) que nos dividió a esta familia en dos grupos fue de 5.0. EL valor de granularidad (I) de 5.0 usado con Pfam representa una división más fina del grupo realizada por MCL, lo que pudiera deberse a una sobrerrepresentación de secuencias de la proteína antranilato sintasa en el grupo Pfam, lo que obliga a usar un valor mayor para poder encontrar y dividir al grupo de secuencias de para-aminobenzoato sintasas que estan subrepresentadas en el universo en el cual se busca. Sin embargo el valor de granularidad (I) de 2.0 empleado para la división de los COGs en este trabajo, al ser utilizado en la familia de proteínas de las CDP-alcohol fosfatidiltransferasas (PF01066) del Pfam, nos separó a la subfamilia de la PgsA y de la Pis en grupos diferentes, como lo encontrado en la división del COG0558 (ver Sección 5.3.2). Además de estas dos subfamilias (PgsA y Pis) encontradas al dividir el Pfam PF01066, se encontraron las subfamilias representadas por las proteínas Pss y Pcs, así como una quinta subfamilia representada en su mayoría por secuencias de organismos eucariotes. El hallazgo de 4 subfamilias más en la base de datos Pfam, las cuales no fueron encontradas en el COG0558, se debe a el mayor número de secuencias representadas en la familia del Pfam que contiene a las fosfatidiltransferasas.

7. Conclusión

La identificación de subgrupos (subfamilias) dentro de las familias de proteínas, es una tarea necesaria para poder hacer una anotación específica en los estudios genómicos que utilizan bases de datos como los COGs y Pfam. Nuestro enfoque para la división de grupos con presencia de parálogos representa una alternativa para lograr dicho objetivo. La correcta división hecha por nuestro método de familias con parálogos representadas tanto en los COGs como en el Pfam, muestran el alcance de los BBTs para la división de dichos grupos. Las familias de proteínas ortólogas son utilizadas ampliamente en estudios genómicos como anotaciones de genomas, o bien en estudios filogenéticos, donde se emplean en la construcción de árboles evolutivos. La correcta separación de las familias de proteínas existentes, en subgrupos lograda con nuestro enfoque, ayuda a poder mejorar las anotaciones funcionales llevadas a cabo en los estudios genómicos, en los análisis y reconstrucciones evolutivas de los genes y sus especies o en la búsqueda de señales de regulación.

8. Perspectivas

La utilización de MCL para la subdivisión de las familias de proteínas en subgrupos, debe ser evaluada de manera más específica en cuanto a los valores de granularidad (I) utilizados, con el objetivo de lograr un valor o intervalo de valores adecuado a la mayoría de las familias analizadas. La evaluación del contexto genómico, o bien en el caso de enzimas, su pertenencia en vías metabólicas específicas descritas en bases de datos como el KEGG, pueden ser una herramienta útil que nos permita lograr dicho objetivo.

Para la documentación de las subfamilias (MOGs) encontradas en los COGs, se pondrán de manera pública los datos obtenidos en el Internet, de tal forma que puedan ser consultados y documentados por los propios usuarios, habiendo de esta manera una retroalimentación entre los MOGs y sus usuarios.

Realización de estudios filogenéticos como método de división de las familias con parálogos, para su comparación con nuestros resultados.

9. Apéndice 1: Cuadro COGs eliminados en 1er. filtro

Cuadro con los COGs eliminados en el primer filtro (ver 4.2).

Cuadro 19: COGs eliminados en el 1er. filtro

COG	Descripción
COG1106	Predicted ATPases
COG1306	Uncharacterized conserved prot
COG1334	Uncharacterized flagellar prot
COG1367	Uncharacterized protein predic
COG1417	Uncharacterized conserved prot
COG1421	Uncharacterized protein predic
COG1431	Uncharacterized protein contai
COG1531	Uncharacterized protein conser
COG1532	Predicted RNA-binding protein
COG1547	Uncharacterized conserved prot
COG1563	Predicted subunit of the Multi
COG1567	Uncharacterized protein predic
COG1617	Uncharacterized conserved prot
COG1627	Uncharacterized protein conser
COG1659	Uncharacterized protein, linoc
COG1662	Transposase and inactivated de
COG1667	Uncharacterized protein conser
COG1688	Uncharacterized protein predic
COG1710	Uncharacterized protein conser
COG1711	Uncharacterized protein conser
COG1719	Predicted hydrocarbon binding

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG1747	Uncharacterized N-terminal dom
COG1769	Uncharacterized protein predic
COG1771	Uncharacterized protein conser
COG1783	Phage terminase large subunit
COG1880	CO dehydrogenase/acetyl-CoA sy
COG1895	Uncharacterized conserved prot
COG1978	Uncharacterized protein conser
COG1991	Uncharacterized conserved prot
COG1996	DNA-directed RNA polymerase, s
COG2104	Sulfur transfer protein involv
COG2143	Thioredoxin-related protein
COG2164	Uncharacterized conserved prot
COG2168	Uncharacterized conserved prot
COG2247	Putative cell wall-binding dom
COG2253	Uncharacterized conserved prot
COG2254	Predicted HD superfamily hydro
COG2306	Uncharacterized conserved prot
COG2314	Predicted membrane protein
COG2316	Predicted hydrolase (HD superf
COG2331	Uncharacterized protein conser
COG2348	Uncharacterized protein involv
COG2380	Uncharacterized protein conser
COG2402	Predicted nucleic acid-binding

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG2404	Predicted phosphohydrolase (DH
COG2410	Uncharacterized conserved prot
COG2427	Uncharacterized conserved prot
COG2442	Uncharacterized conserved prot
COG2456	Uncharacterized conserved prot
COG2512	Uncharacterized membrane-assoc
COG2522	Predicted transcriptional regu
COG2732	Barstar, RNase (barnase) inhib
COG2747	Negative regulator of flagelli
COG2770	FOG: HAMP domain
COG2810	Predicted type IV restriction
COG2880	Uncharacterized protein conser
COG2881	Uncharacterized protein conser
COG2886	Uncharacterized small protein
COG2892	Uncharacterized protein conser
COG2906	Bacterioferritin-associated fe
COG2946	Putative phage replication pro
COG2952	Uncharacterized protein conser
COG3014	Uncharacterized protein conser
COG3015	Uncharacterized lipoprotein NI
COG3016	Uncharacterized iron-regulated
COG3018	Uncharacterized protein conser
COG3031	Type II secretory pathway, com

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3042	Putative hemolysin
COG3050	DNA polymerase III, psi subuni
COG3059	Predicted membrane protein
COG3061	Cell envelope opacity-associat
COG3062	Uncharacterized protein involv
COG3087	Cell division protein
COG3112	Uncharacterized protein conser
COG3113	Predicted NTP binding protein
COG3114	Heme exporter protein D
COG3122	Uncharacterized protein conser
COG3130	Ribosome modulation factor
COG3137	Putative salt-induced outer me
COG3140	Uncharacterized protein conser
COG3141	Uncharacterized protein conser
COG3149	Type II secretory pathway, com
COG3162	Predicted membrane protein
COG3187	Heat shock protein
COG3197	Uncharacterized protein, possi
COG3202	ATP/ADP translocase
COG3219	Uncharacterized protein conser
COG3222	Uncharacterized protein conser
COG3225	ABC-type uncharacterized trans
COG3233	Predicted deacetylase

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3242	Uncharacterized protein conser
COG3262	Ni,Fe-hydrogenase III componen
COG3266	Uncharacterized protein conser
COG3267	Type II secretory pathway, com
COG3270	Uncharacterized conserved prot
COG3295	Uncharacterized protein conser
COG3296	Uncharacterized protein conser
COG3297	Type II secretory pathway, com
COG3302	DMSO reductase anchor subunit
COG3305	Predicted membrane protein
COG3308	Predicted membrane protein
COG3309	Uncharacterized virulence-asso
COG3318	Predicted metal-binding protei
COG3337	Uncharacterized protein predic
COG3349	Uncharacterized conserved prot
COG3350	Uncharacterized conserved prot
COG3352	Putative archaeal flagellar pr
COG3353	Putative archaeal flagellar pr
COG3361	Uncharacterized conserved prot
COG3364	Zn-ribbon containing protein
COG3366	Uncharacterized protein conser
COG3368	Predicted permease
COG3369	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3370	Uncharacterized protein conser
COG3375	Uncharacterized conserved prot
COG3393	Predicted acetyltransferase
COG3398	Uncharacterized protein conser
COG3399	Uncharacterized protein conser
COG3401	Fibronectin type 3 domain-cont
COG3403	Uncharacterized conserved prot
COG3410	Uncharacterized conserved prot
COG3415	Transposase and inactivated de
COG3417	Collagen-binding surface adhes
COG3418	Flagellar biosynthesis/type II
COG3430	Uncharacterized protein conser
COG3431	Predicted membrane protein
COG3432	Predicted transcriptional regu
COG3433	Aryl carrier domain
COG3434	Predicted signal transduction
COG3440	Predicted restriction endonucl
COG3450	Predicted enzyme of the cupin
COG3461	Uncharacterized conserved prot
COG3462	Predicted membrane protein
COG3463	Predicted membrane protein
COG3465	Uncharacterized conserved prot
COG3466	Putative transposon-encoded pr

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3528	Uncharacterized protein conser
COG3529	Predicted nucleic-acid-binding
COG3548	Predicted integral membrane pr
COG3551	Uncharacterized protein conser
COG3582	Predicted nucleic acid binding
COG3584	Uncharacterized protein conser
COG3585	Molybdopterin-binding protein
COG3597	Uncharacterized protein/domain
COG3598	RecA-family ATPase
COG3612	Uncharacterized protein conser
COG3621	Patatin
COG3629	DNA-binding transcriptional ac
COG3631	Ketosteroid isomerase-related
COG3646	Uncharacterized phage-encoded
COG3650	Predicted membrane protein
COG3656	Predicted periplasmic protein
COG3659	Carbohydrate-selective porin
COG3664	Beta-xylosidase
COG3671	Predicted membrane protein
COG3675	Predicted lipase
COG3676	Transposase and inactivated de
COG3677	Transposase and inactivated de
COG3678	P pilus assembly/Cpx signaling

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3680	Uncharacterized protein conser
COG3691	Uncharacterized protein conser
COG3692	Uncharacterized protein conser
COG3702	Type IV secretory pathway, Vir
COG3725	Membrane protein required for
COG3729	General stress protein
COG3742	Uncharacterized protein conser
COG3743	Uncharacterized conserved prot
COG3746	Phosphate-selective porin
COG3747	Phage terminase, small subunit
COG3751	Predicted proline hydroxylase
COG3754	Lipopolysaccharide biosynthesi
COG3756	Uncharacterized protein conser
COG3765	Chain length determinant prote
COG3767	Uncharacterized low-complexity
COG3769	Predicted hydrolase (HAD super
COG3776	Predicted membrane protein
COG3778	Uncharacterized protein conser
COG3779	Uncharacterized protein conser
COG3780	DNA endonuclease related to in
COG3782	Uncharacterized protein conser
COG3783	Soluble cytochrome b562
COG3784	Uncharacterized protein conser

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3787	Uncharacterized protein conser
COG3788	Uncharacterized relative of gl
COG3790	Predicted membrane protein
COG3792	Uncharacterized protein conser
COG3794	Plastocyanin
COG3798	Uncharacterized protein conser
COG3801	Uncharacterized protein conser
COG3804	Uncharacterized conserved prot
COG3805	Aromatic ring-cleaving dioxyge
COG3815	Predicted membrane protein
COG3831	Uncharacterized conserved prot
COG3838	Type IV secretory pathway, Vir
COG3843	Type IV secretory pathway, Vir
COG3847	Flp pilus assembly protein, pi
COG3858	Predicted glycosyl hydrolase
COG3863	Uncharacterized distant relati
COG3864	Uncharacterized protein conser
COG3874	Uncharacterized conserved prot
COG3875	Uncharacterized conserved prot
COG3877	Uncharacterized protein conser
COG3879	Uncharacterized protein conser
COG3881	Uncharacterized protein conser
COG3884	Acyl-ACP thioesterase

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG3890	Phosphomevalonate kinase
COG3896	Chloramphenicol 3-O-phosphotra
COG3903	Predicted ATPase
COG3904	Predicted periplasmic protein
COG3905	Predicted transcriptional regu
COG3909	Cytochrome c556
COG3919	Predicted ATP-grasp enzyme
COG3923	Primosomal replication protein
COG3924	Predicted membrane protein
COG3925	N-terminal domain of the phosp
COG3934	Endo-beta-mannanase
COG3936	Protein involved in polysaccha
COG3937	Uncharacterized conserved prot
COG3942	Surface antigen
COG3947	Response regulator containing
COG3964	Predicted amidohydrolase
COG3976	Uncharacterized protein conser
COG4001	Predicted metal-binding protei
COG4003	Uncharacterized protein conser
COG4004	Uncharacterized protein conser
COG4006	Uncharacterized protein conser
COG4009	Uncharacterized protein conser
COG4010	Uncharacterized protein conser

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4013	Uncharacterized protein conser
COG4014	Uncharacterized protein conser
COG4023	Preprotein translocase subunit
COG4024	Uncharacterized protein conser
COG4025	Predicted membrane protein
COG4027	Uncharacterized protein conser
COG4035	Predicted membrane protein
COG4038	Predicted membrane protein
COG4039	Predicted membrane protein
COG4040	Predicted membrane protein
COG4042	Predicted membrane protein
COG4043	Uncharacterized conserved prot
COG4066	Uncharacterized protein conser
COG4067	Uncharacterized protein conser
COG4068	Uncharacterized protein contai
COG4077	Uncharacterized protein conser
COG4084	Uncharacterized protein conser
COG4085	Predicted RNA-binding protein,
COG4092	Predicted glycosyltransferase
COG4094	Predicted membrane protein
COG4104	Uncharacterized conserved prot
COG4113	Predicted nucleic acid-binding
COG4114	Uncharacterized Fe-S protein

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4118	Antitoxin of toxin-antitoxin s
COG4127	Uncharacterized conserved prot
COG4128	Zonula occludens toxin
COG4138	ABC-type cobalamin transport s
COG4139	ABC-type cobalamin transport s
COG4185	Uncharacterized protein conser
COG4186	Predicted phosphoesterase or p
COG4190	Predicted transcriptional regu
COG4193	Beta- N-acetylglucosaminidase
COG4195	Phage-related replication prot
COG4197	Uncharacterized protein conser
COG4200	Uncharacterized protein conser
COG4218	Tetrahydromethanopterin S-meth
COG4220	Phage DNA packaging protein, N
COG4227	Antirestriction protein
COG4228	Mu-like prophage DNA circulati
COG4237	Hydrogenase 4 membrane compone
COG4238	Murein lipoprotein
COG4241	Predicted membrane protein
COG4242	Cyanophycinase and related exo
COG4243	Predicted membrane protein
COG4244	Predicted membrane protein
COG4245	Uncharacterized protein encode

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4249	Uncharacterized protein contai
COG4250	Predicted sensor protein/domai
COG4253	Uncharacterized protein conser
COG4255	Uncharacterized protein conser
COG4256	Hemin uptake protein
COG4258	Predicted exporter
COG4260	Putative virion core protein (
COG4261	Predicted acyltransferase
COG4273	Uncharacterized conserved prot
COG4276	Uncharacterized conserved prot
COG4278	Uncharacterized conserved prot
COG4280	Predicted membrane protein
COG4289	Uncharacterized protein conser
COG4293	Uncharacterized protein conser
COG4296	Uncharacterized protein conser
COG4298	Uncharacterized protein conser
COG4304	Uncharacterized protein conser
COG4308	Limonene-1,2-epoxide hydrolase
COG4316	Uncharacterized protein conser
COG4320	Uncharacterized protein conser
COG4322	Uncharacterized protein conser
COG4323	Predicted membrane protein
COG4327	Predicted membrane protein

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4333	Uncharacterized protein conser
COG4334	Uncharacterized protein conser
COG4338	Uncharacterized protein conser
COG4342	Uncharacterized protein conser
COG4343	Uncharacterized protein conser
COG4352	Ribosomal protein L13E
COG4359	Uncharacterized conserved prot
COG4367	Uncharacterized protein conser
COG4370	Uncharacterized protein conser
COG4374	Uncharacterized protein conser
COG4378	Uncharacterized protein conser
COG4379	Mu-like prophage tail protein
COG4381	Mu-like prophage protein gp46
COG4382	Mu-like prophage protein gp16
COG4384	Mu-like prophage protein gp45
COG4390	Uncharacterized protein conser
COG4391	Uncharacterized protein conser
COG4409	Neuraminidase (sialidase)
COG4416	Mu-like prophage protein Com
COG4421	Capsular polysaccharide biosyn
COG4422	Bacteriophage protein gp37
COG4423	Uncharacterized protein conser
COG4446	Uncharacterized protein conser

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4449	Predicted protease of the Abi
COG4459	Periplasmic nitrate reductase
COG4460	Uncharacterized protein conser
COG4471	Uncharacterized protein conser
COG4518	Mu-like prophage FluMu protein
COG4519	Uncharacterized protein conser
COG4538	Uncharacterized conserved prot
COG4545	Glutaredoxin-related protein
COG4568	Transcriptional antiterminator
COG4572	Putative cation transport regu
COG4574	Serine protease inhibitor ecot
COG4575	Uncharacterized conserved prot
COG4578	Glucitol operon activator
COG4619	ABC-type uncharacterized trans
COG4627	Uncharacterized protein conser
COG4632	Exopolysaccharide biosynthesis
COG4633	Uncharacterized protein conser
COG4634	Uncharacterized protein conser
COG4636	Uncharacterized protein conser
COG4639	Predicted kinase
COG4641	Uncharacterized protein conser
COG4648	Predicted membrane protein
COG4654	Cytochrome c551/c552

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4682	Predicted membrane protein
COG4683	Uncharacterized protein conser
COG4684	Predicted membrane protein
COG4688	Uncharacterized protein conser
COG4691	Plasmid stability protein
COG4692	Predicted neuraminidase (siali
COG4694	Uncharacterized protein conser
COG4696	Uncharacterized protein conser
COG4697	Uncharacterized protein conser
COG4699	Uncharacterized protein conser
COG4701	Uncharacterized protein conser
COG4703	Uncharacterized protein conser
COG4705	Uncharacterized membrane-ancho
COG4707	Uncharacterized protein conser
COG4710	Predicted DNA-binding protein
COG4711	Predicted membrane protein
COG4712	Uncharacterized protein conser
COG4713	Predicted membrane protein
COG4715	Uncharacterized conserved prot
COG4718	Phage-related protein
COG4719	Uncharacterized protein conser
COG4720	Predicted membrane protein
COG4724	Endo-beta-N-acetylglucosaminid

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4725	Transcriptional activator, ade
COG4726	Tfp pilus assembly protein Pil
COG4728	Uncharacterized protein conser
COG4729	Uncharacterized conserved prot
COG4734	Antirestriction protein
COG4735	Uncharacterized protein conser
COG4736	Cbb3-type cytochrome oxidase,
COG4738	Predicted transcriptional regu
COG4739	Uncharacterized protein contai
COG4742	Predicted transcriptional regu
COG4744	Uncharacterized conserved prot
COG4745	Predicted membrane-bound manno
COG4746	Uncharacterized protein conser
COG4754	Uncharacterized conserved prot
COG4759	Uncharacterized protein conser
COG4762	Uncharacterized protein conser
COG4763	Predicted membrane protein
COG4767	Glycopeptide antibiotics resis
COG4792	Type III secretory pathway, co
COG4797	Predicted regulatory domain of
COG4798	Predicted methyltransferase
COG4805	Uncharacterized protein conser
COG4817	Uncharacterized protein conser

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4821	Uncharacterized protein contai
COG4827	Predicted transporter
COG4828	Predicted membrane protein
COG4829	Muconolactone delta-isomerase
COG4835	Uncharacterized protein conser
COG4839	Protein required for the initi
COG4840	Uncharacterized protein conser
COG4845	Chloramphenicol O-acetyltransf
COG4847	Uncharacterized protein conser
COG4854	Predicted membrane protein
COG4858	Uncharacterized membrane-bound
COG4859	Uncharacterized protein conser
COG4869	Propanediol utilization protei
COG4871	Uncharacterized protein conser
COG4877	Uncharacterized protein conser
COG4882	Predicted aminopeptidase, Iap
COG4885	Uncharacterized protein conser
COG4888	Uncharacterized Zn ribbon-cont
COG4890	Predicted outer membrane lipop
COG4891	Uncharacterized conserved prot
COG4892	Predicted heme/steroid binding
COG4893	Uncharacterized protein conser
COG4894	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4895	Uncharacterized conserved prot
COG4896	Uncharacterized protein conser
COG4897	Uncharacterized protein conser
COG4898	Uncharacterized protein conser
COG4899	Uncharacterized protein conser
COG4900	Predicted metallopeptidase
COG4903	Genetic competence transcripti
COG4904	Uncharacterized protein conser
COG4907	Predicted membrane protein
COG4911	Uncharacterized conserved prot
COG4916	Uncharacterized protein contai
COG4918	Uncharacterized protein conser
COG4919	Ribosomal protein S30
COG4920	Predicted membrane protein
COG4921	Uncharacterized protein conser
COG4925	Uncharacterized conserved prot
COG4929	Uncharacterized membrane-ancho
COG4933	Uncharacterized conserved prot
COG4934	Predicted protease
COG4935	Regulatory P domain of the sub
COG4937	Predicted regulatory domain of
COG4938	Uncharacterized conserved prot
COG4939	Major membrane immunogen, memb

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG4940	Competence protein ComGF
COG4950	Uncharacterized protein conser
COG4966	Tfp pilus assembly protein Pil
COG4967	Tfp pilus assembly protein Pil
COG4978	Transcriptional regulator, eff
COG4980	Gas vesicle protein
COG4990	Uncharacterized protein conser
COG4991	Uncharacterized protein with a
COG4994	Uncharacterized protein conser
COG4995	Uncharacterized protein conser
COG4996	Predicted phosphatase
COG4999	Uncharacterized domain of BarA
COG5003	Mu-like prophage protein gp37
COG5004	P2-like prophage tail protein
COG5005	Mu-like prophage protein gpG
COG5015	Uncharacterized conserved prot
COG5017	Uncharacterized conserved prot
COG5018	Inhibitor of the KinA pathway
COG5019	Septin family protein
COG5024	Cyclin
COG5025	Transcription factor of the Fo
COG5030	Clathrin adaptor complex, smal
COG5031	Uncharacterized protein involv

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5034	Chromatin remodeling protein,
COG5035	Cell cycle control protein
COG5036	SPX domain-containing protein
COG5037	Gluconate transport-inducing p
COG5040	14-3-3 family protein
COG5045	Ribosomal protein S10E
COG5046	Protein involved in Mod5 prote
COG5048	FOG: Zn-finger
COG5051	Ribosomal protein L36E
COG5053	Translation initiation factor
COG5055	Recombination DNA repair prote
COG5056	Acyl-CoA cholesterol acyltrans
COG5058	Protein transporter of the TRA
COG5061	Oxidoreductin, endoplasmic ret
COG5062	Uncharacterized membrane prote
COG5063	CCCH-type Zn-finger protein
COG5064	Karyopherin (importin) alpha
COG5065	Protein involved in inorganic
COG5066	VAMP-associated protein involv
COG5067	Protein kinase essential for t
COG5069	Ca ²⁺ -binding actin-bundling pr
COG5072	Serine/threonine kinase of the
COG5073	Vacuolar import and degradatio

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5074	t-SNARE complex subunit, synta
COG5077	Ubiquitin carboxyl-terminal hy
COG5079	Nuclear protein export factor
COG5082	Arginine methyltransferase-int
COG5085	Predicted membrane protein
COG5086	Uncharacterized conserved prot
COG5087	Uncharacterized conserved prot
COG5090	Transcription initiation facto
COG5091	Suppressor of G2 allele of skp
COG5093	Uncharacterized conserved prot
COG5094	Transcription initiation facto
COG5097	RNA polymerase II transcriptio
COG5101	Importin beta-related nuclear
COG5102	Membrane protein involved in E
COG5104	Splicing factor
COG5106	Uncharacterized conserved prot
COG5109	Uncharacterized conserved prot
COG5112	U1-like Zn-finger-containing p
COG5118	Transcription initiation facto
COG5119	Uncharacterized protein, conta
COG5120	Membrane protein involved in G
COG5122	Transport protein particle (TR
COG5123	Transcription initiation facto

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5124	Protein predicted to be involv
COG5125	Uncharacterized conserved prot
COG5127	Vacuolar H ⁺ -ATPase V1 sector,
COG5128	Transport protein particle (TR
COG5130	Prenylated rab acceptor 1 and
COG5131	Ubiquitin-like protein
COG5134	Uncharacterized conserved prot
COG5135	Uncharacterized conserved prot
COG5136	U1 snRNP-specific protein C
COG5138	Uncharacterized conserved prot
COG5139	Uncharacterized conserved prot
COG5142	Oxidation resistance protein
COG5146	Pantothenate kinase, acetyl-Co
COG5148	26S proteasome regulatory comp
COG5149	Transcription initiation facto
COG5152	Uncharacterized conserved prot
COG5156	Anaphase-promoting complex (AP
COG5157	RNA polymerase II assessor y fa
COG5162	Transcription initiation facto
COG5164	Transcription elongation facto
COG5173	Exocyst complex subunit SEC6
COG5174	Transcription initiation facto
COG5177	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5188	Splicing factor 3a, subunit 3
COG5189	Putative transcriptional repre
COG5192	GTP-binding protein required f
COG5195	Uncharacterized conserved prot
COG5198	Protein tyrosine phosphatase-l
COG5200	U1 snRNP component, mediates U
COG5201	SCF ubiquitin ligase, SKP1 com
COG5204	Transcription elongation facto
COG5207	Isopeptidase T
COG5218	Chromosome condensation comple
COG5223	Uncharacterized conserved prot
COG5224	CCAAT-binding factor, subunit
COG5225	Uncharacterized protein involv
COG5227	Ubiquitin-like protein (senti
COG5229	Chromosome condensation comple
COG5230	Uncharacterized conserved prot
COG5231	Vacuolar H ⁺ -ATPase V1 sector,
COG5232	Preprotein translocase subunit
COG5233	Peripheral Golgi membrane prot
COG5234	Beta-tubulin folding cofactor
COG5235	Single-stranded DNA-binding re
COG5236	Uncharacterized conserved prot
COG5242	RNA polymerase II transcriptio

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5243	HRD ubiquitin ligase complex,
COG5244	Dynactin complex subunit invol
COG5245	Dynein, heavy chain
COG5246	Splicing factor 3a, subunit 2
COG5247	Class 2 transcription represso
COG5248	Transcription initiation facto
COG5250	RNA polymerase II, fourth larg
COG5251	Transcription initiation facto
COG5252	Uncharacterized conserved prot
COG5254	Predicted membrane protein
COG5259	RSC chromatin remodeling compl
COG5260	DNA polymerase sigma
COG5262	Histone H2A
COG5263	FOG: Glucan-binding domain (YG
COG5266	ABC-type Co ²⁺ transport system
COG5268	Type IV secretory pathway, Trb
COG5274	Cytochrome b involved in lipid
COG5275	BRCT domain type II
COG5276	Uncharacterized conserved prot
COG5279	Uncharacterized protein involv
COG5282	Uncharacterized conserved prot
COG5285	Protein involved in biosynthes
COG5290	IkappaB kinase complex, IKAP c

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5291	Predicted membrane protein
COG5293	Uncharacterized protein conser
COG5294	Uncharacterized protein conser
COG5296	Transcription factor involved
COG5297	Cellobiohydrolase A (1,4-beta-
COG5298	Uncharacterized protein conser
COG5302	Post-segregation antitoxin (cc
COG5304	Uncharacterized protein conser
COG5305	Predicted membrane protein
COG5306	Uncharacterized conserved prot
COG5308	Nuclear pore complex subunit
COG5309	Exo-beta-1,3-glucanase
COG5324	Uncharacterized conserved prot
COG5325	t-SNARE complex subunit, synta
COG5331	Uncharacterized protein conser
COG5333	Cdk activating kinase (CAK)/RN
COG5337	Spore coat assembly protein
COG5339	Uncharacterized protein conser
COG5340	Predicted transcriptional regu
COG5341	Uncharacterized protein conser
COG5346	Predicted membrane protein
COG5354	Uncharacterized protein, conta
COG5362	Phage-related terminase

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5366	Protein involved in propagatio
COG5369	Uncharacterized conserved prot
COG5371	Golgi nucleoside diphosphatase
COG5374	Uncharacterized conserved prot
COG5377	Phage-related protein, predict
COG5378	Predicted nucleotide-binding p
COG5379	S-adenosylmethionine:diacygly
COG5380	Lipase chaperone
COG5381	Uncharacterized protein conser
COG5384	U3 small nucleolar ribonucleop
COG5391	Phox homology (PX) domain prot
COG5393	Predicted membrane protein
COG5398	Heme oxygenase
COG5399	Uncharacterized protein conser
COG5401	Spore germination protein
COG5403	Uncharacterized conserved prot
COG5404	SOS-response cell division inh
COG5408	SPX domain-containing protein
COG5409	EXS domain-containing protein
COG5410	Uncharacterized protein conser
COG5411	Phosphatidylinositol 5-phospha
COG5413	Uncharacterized integral membr
COG5414	TATA-binding protein-associate

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5415	Predicted integral membrane me
COG5416	Uncharacterized integral membr
COG5417	Uncharacterized small protein
COG5418	Predicted secreted protein
COG5419	Uncharacterized conserved prot
COG5420	Uncharacterized conserved smal
COG5421	Transposase
COG5422	RhoGEF, Guanine nucleotide exc
COG5423	Predicted metal-binding protei
COG5424	Pyrroloquinoline quinone (Coen
COG5426	Uncharacterized membrane prote
COG5427	Uncharacterized membrane prote
COG5428	Uncharacterized conserved smal
COG5431	Uncharacterized metal-binding
COG5432	RING-finger-containing E3 ubiq
COG5439	Uncharacterized conserved prot
COG5444	Uncharacterized conserved prot
COG5450	Transcription regulator of the
COG5456	Predicted integral membrane pr
COG5459	Predicted rRNA methylase
COG5460	Uncharacterized conserved prot
COG5463	Predicted integral membrane pr
COG5464	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5466	Predicted small metal-binding
COG5469	Predicted metal-binding protei
COG5471	Uncharacterized conserved prot
COG5472	Predicted small integral membr
COG5473	Predicted integral membrane pr
COG5474	Uncharacterized conserved prot
COG5475	Uncharacterized small protein
COG5479	Uncharacterized protein potent
COG5483	Uncharacterized conserved prot
COG5484	Uncharacterized conserved prot
COG5486	Predicted metal-binding integr
COG5489	Uncharacterized conserved prot
COG5490	Uncharacterized conserved prot
COG5491	Conserved protein implicated i
COG5492	Bacterial surface proteins con
COG5494	Predicted thioredoxin/glutared
COG5495	Uncharacterized conserved prot
COG5497	Predicted secreted protein
COG5498	Predicted glycosyl hydrolase
COG5501	Predicted secreted protein
COG5504	Predicted Zn-dependent proteas
COG5505	Predicted integral membrane pr
COG5506	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5507	Uncharacterized conserved prot
COG5509	Uncharacterized small protein
COG5510	Predicted small secreted prote
COG5511	Bacteriophage capsid protein
COG5512	Zn-ribbon-containing, possibly
COG5513	Predicted secreted protein
COG5514	Uncharacterized conserved prot
COG5516	Conserved protein containing a
COG5517	Small subunit of phenylpropion
COG5518	Bacteriophage capsid portal pr
COG5522	Predicted integral membrane pr
COG5523	Predicted integral membrane pr
COG5524	Bacteriorhodopsin
COG5525	Bacteriophage tail assembly pr
COG5527	Protein involved in initiation
COG5529	Pyocin large subunit
COG5532	Uncharacterized conserved prot
COG5534	Plasmid replication initiator
COG5535	DNA repair protein RAD4
COG5536	Protein prenyltransferase, alp
COG5537	Cohesin
COG5538	Preprotein translocase subunit
COG5539	Predicted cysteine protease (O

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5540	RING-finger-containing ubiquit
COG5541	Vesicle coat complex COPI, zet
COG5542	Predicted integral membrane pr
COG5543	Uncharacterized conserved prot
COG5544	Predicted periplasmic lipoprot
COG5547	Small integral membrane protei
COG5548	Small integral membrane protei
COG5549	Predicted Zn-dependent proteas
COG5550	Predicted aspartyl protease
COG5551	Uncharacterized conserved prot
COG5553	Predicted metal-dependent enzy
COG5554	Nitrogen fixation protein
COG5555	Cytolysin, a secreted calcineu
COG5557	Polysulphide reductase
COG5558	Transposase
COG5559	Uncharacterized conserved smal
COG5560	Ubiquitin C-terminal hydrolase
COG5561	Predicted metal-binding protei
COG5562	Phage envelope protein
COG5564	Predicted TIM-barrel enzyme, p
COG5566	Uncharacterized conserved prot
COG5567	Predicted small periplasmic li
COG5569	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5573	Predicted nucleic-acid-binding
COG5574	RING-finger-containing E3 ubiq
COG5575	Origin recognition complex, su
COG5576	Homeodomain-containing transcr
COG5577	Spore coat protein
COG5578	Predicted integral membrane pr
COG5580	Activator of HSP90 ATPase
COG5581	Predicted glycosyltransferase
COG5582	Uncharacterized conserved prot
COG5583	Uncharacterized small protein
COG5584	Predicted small secreted prote
COG5585	NAD ⁺ -asparagine ADP-ribosyltr
COG5586	Uncharacterized conserved prot
COG5588	Uncharacterized conserved prot
COG5589	Uncharacterized conserved prot
COG5590	Uncharacterized conserved prot
COG5591	Uncharacterized conserved prot
COG5592	Uncharacterized conserved prot
COG5593	Nucleic-acid-binding protein p
COG5595	Zn-ribbon-containing, possibly
COG5596	Mitochondrial import inner mem
COG5597	Alpha-N-acetylglucosamine tran
COG5598	Trimethylamine:corrinoid methy

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5599	Protein tyrosine phosphatase
COG5600	Transcription-associated recom
COG5601	General negative regulator of
COG5603	Subunit of TRAPP, an ER-Golgi
COG5604	Uncharacterized conserved prot
COG5606	Uncharacterized conserved smal
COG5608	Conserved secreted protein
COG5609	Uncharacterized conserved prot
COG5611	Predicted nucleic-acid-binding
COG5612	Predicted integral membrane pr
COG5613	Uncharacterized conserved prot
COG5614	Bacteriophage head-tail adapto
COG5615	Predicted integral membrane pr
COG5616	Predicted integral membrane pr
COG5619	Uncharacterized conserved prot
COG5620	Uncharacterized conserved prot
COG5622	Protein required for attachmen
COG5623	Pre-mRNA cleavage and polyaden
COG5624	Transcription initiation facto
COG5625	Predicted transcription regula
COG5626	Uncharacterized conserved smal
COG5627	DNA repair protein MMS21
COG5628	Predicted acetyltransferase

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5629	Predicted metal-binding protei
COG5630	Acetylglutamate synthase
COG5632	N-acetylmuramoyl-L-alanine ami
COG5633	Predicted periplasmic lipoprot
COG5634	Uncharacterized conserved prot
COG5635	Predicted NTPase (NACHT family
COG5636	Uncharacterized conserved prot
COG5637	Predicted integral membrane pr
COG5638	Uncharacterized conserved prot
COG5639	Uncharacterized conserved smal
COG5640	Secreted trypsin-like serine p
COG5641	GATA Zn-finger-containing tran
COG5642	Uncharacterized conserved prot
COG5643	Protein containing a metal-bin
COG5645	Predicted periplasmic lipoprot
COG5646	Uncharacterized conserved prot
COG5647	Cullin, a subunit of E3 ubiqui
COG5648	Chromatin-associated proteins
COG5650	Predicted integral membrane pr
COG5651	PPE-repeat proteins
COG5652	Predicted integral membrane pr
COG5653	Protein involved in cellulose
COG5654	Uncharacterized conserved prot

Continúa ...

Cuadro 19: COGs eliminados (continuación)

COG	Descripción
COG5655	Plasmid rolling circle replica
COG5657	CAS/CSE protein involved in ch
COG5658	Predicted integral membrane pr
COG5659	FOG: Transposase
COG5660	Predicted integral membrane pr
COG5663	Uncharacterized conserved prot
COG5665	CCR4-NOT transcriptional regul

Fin tabla

10. Apendice 2: Cuadro COGs eliminados en 2do. filtro

Cuadro con los COGs eliminados en el segundo filtro (ver 4.3).

Cuadro 20: COGs eliminados en el 1er. filtro

COG	Descripción
COG0699	Predicted GTPases (dynamin-rel
COG1083	CMP-N-acetylneuraminic acid sy
COG1124	ABC-type dipeptide/oligopeptid
COG1221	Transcriptional regulators con
COG1261	Flagellar basal body P-ring bi
COG1416	Uncharacterized conserved prot
COG1424	Pimeloyl-CoA synthetase
COG1441	O-succinylbenzoate synthase
COG1442	Lipopolysaccharide biosynthesi
COG1462	Uncharacterized protein involv
COG1630	Uncharacterized protein conser
COG1633	Uncharacterized conserved prot
COG1689	Uncharacterized protein conser
COG1700	Uncharacterized conserved prot
COG1753	Uncharacterized conserved prot
COG1790	Uncharacterized protein conser
COG1822	Predicted archaeal membrane pr
COG1848	Predicted nucleic acid-binding
COG1851	Uncharacterized conserved prot
COG1969	Ni,Fe-hydrogenase I cytochrome
COG2000	Predicted Fe-S protein

Continúa . . .

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG2018	Uncharacterized distant relati
COG2140	Thermophilic glucose-6-phospha
COG2161	Antitoxin of toxin-antitoxin s
COG2229	Predicted GTPase
COG2389	Uncharacterized metal-binding
COG2443	Preprotein translocase subunit
COG2445	Uncharacterized conserved prot
COG2452	Predicted site-specific integr
COG2457	Uncharacterized conserved prot
COG2503	Predicted secreted acid phosph
COG2604	Uncharacterized protein conser
COG2829	Outer membrane phospholipase A
COG2830	Uncharacterized protein conser
COG2833	Uncharacterized protein conser
COG2841	Uncharacterized protein conser
COG2842	Uncharacterized ATPase, putati
COG2845	Uncharacterized protein conser
COG2849	Uncharacterized protein conser
COG2900	Uncharacterized protein conser
COG2916	DNA-binding protein H-NS
COG2926	Uncharacterized protein conser
COG2965	Primosomal replication protein
COG2973	Trp operon repressor

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG2980	Rare lipoprotein B
COG2991	Uncharacterized protein conser
COG2992	Uncharacterized FlgJ-related p
COG2994	ACP:hemolysin acyltransferase
COG2999	Glutaredoxin 2
COG3002	Uncharacterized protein conser
COG3006	Uncharacterized protein involv
COG3009	Uncharacterized protein conser
COG3036	Uncharacterized protein conser
COG3054	Predicted transcriptional regu
COG3055	Uncharacterized protein conser
COG3056	Uncharacterized lipoprotein
COG3057	Negative regulator of replicat
COG3060	Transcriptional regulator of m
COG3064	Membrane protein involved in c
COG3066	DNA mismatch repair protein
COG3068	Uncharacterized protein conser
COG3069	C4-dicarboxylate transporter
COG3073	Negative regulator of sigma E
COG3074	Uncharacterized protein conser
COG3076	Uncharacterized protein conser
COG3078	Uncharacterized protein conser
COG3079	Uncharacterized protein conser

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG3082	Uncharacterized protein conser
COG3083	Predicted hydrolase of alkaline
COG3084	Uncharacterized protein conser
COG3085	Uncharacterized protein conser
COG3086	Positive regulator of sigma E
COG3089	Uncharacterized protein conser
COG3092	Uncharacterized protein conser
COG3094	Uncharacterized protein conser
COG3095	Uncharacterized protein involv
COG3096	Uncharacterized protein involv
COG3097	Uncharacterized protein conser
COG3099	Uncharacterized protein conser
COG3100	Uncharacterized protein conser
COG3101	Uncharacterized protein conser
COG3102	Uncharacterized protein conser
COG3105	Uncharacterized protein conser
COG3109	Activator of osmoprotectant tr
COG3110	Uncharacterized protein conser
COG3111	Uncharacterized conserved prot
COG3116	Cell division protein
COG3117	Uncharacterized protein conser
COG3120	Uncharacterized protein conser
COG3139	Uncharacterized protein conser

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG3148	Uncharacterized conserved prot
COG3155	Uncharacterized protein involv
COG3165	Uncharacterized protein conser
COG3171	Uncharacterized protein conser
COG3172	Predicted ATPase/kinase involv
COG3184	Uncharacterized protein conser
COG3190	Flagellar biogenesis protein
COG3196	Uncharacterized protein conser
COG3198	Uncharacterized protein conser
COG3203	Outer membrane protein (porin)
COG3230	Heme oxygenase
COG3235	Predicted membrane protein
COG3236	Uncharacterized protein conser
COG3241	Azurin
COG3245	Cytochrome c5
COG3248	Nucleoside-binding outer membr
COG3251	Uncharacterized protein conser
COG3271	Predicted double-glycine pepti
COG3274	Uncharacterized protein conser
COG3298	Predicted 3'-5' exonuclease re
COG3299	Uncharacterized homolog of pha
COG3303	Formate-dependent nitrite redu
COG3306	Glycosyltransferase involved i

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG3311	Predicted transcriptional regu
COG3317	Uncharacterized lipoprotein
COG3334	Uncharacterized conserved prot
COG3379	Uncharacterized conserved prot
COG3381	Uncharacterized component of a
COG3392	Adenine-specific DNA methylase
COG3400	Uncharacterized protein conser
COG3405	Endoglucanase Y
COG3409	Putative peptidoglycan-binding
COG3411	Ferredoxin
COG3419	Tfp pilus assembly protein, ti
COG3421	Uncharacterized protein conser
COG3422	Uncharacterized conserved prot
COG3423	Predicted transcriptional regu
COG3439	Uncharacterized conserved prot
COG3445	Acid-induced glycy radical en
COG3456	Uncharacterized conserved prot
COG3457	Predicted amino acid racemase
COG3460	Uncharacterized enzyme of phen
COG3468	Type V secretory pathway, adhe
COG3470	Uncharacterized protein probab
COG3471	Predicted periplasmic/secreted
COG3477	Predicted periplasmic/secreted

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG3499	Phage protein U
COG3512	Uncharacterized protein conser
COG3513	Uncharacterized protein conser
COG3514	Uncharacterized protein conser
COG3518	Uncharacterized protein conser
COG3521	Uncharacterized protein conser
COG3524	Capsule polysaccharide export
COG3553	Uncharacterized protein conser
COG3561	Phage anti-repressor protein
COG3562	Capsule polysaccharide export
COG3563	Capsule polysaccharide export
COG3565	Predicted dioxygenase of extra
COG3587	Restriction endonuclease
COG3615	Uncharacterized protein/domain
COG3617	Prophage antirepressor
COG3618	Predicted metal-dependent hydr
COG3630	Na ⁺ -transporting methylmalonyl
COG3637	Opacity protein and related su
COG3645	Uncharacterized phage-encoded
COG3647	Predicted membrane protein
COG3649	Uncharacterized protein predic
COG3654	Prophage maintenance system ki
COG3660	Predicted nucleoside-diphospha

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG3683	ABC-type uncharacterized trans
COG3700	Acid phosphatase (class B)
COG3712	Fe ²⁺ -dicitrate sensor, membran
COG3713	Outer membrane protein V
COG3719	Ribonuclease I
COG3721	Putative heme iron utilization
COG3722	Transcriptional regulator
COG3726	Uncharacterized membrane prote
COG3728	Phage terminase, small subunit
COG3731	Phosphotransferase system sorb
COG3740	Phage head maturation protease
COG3744	Uncharacterized protein conser
COG3771	Predicted membrane protein
COG3774	Mannosyltransferase OCH1 and r
COG3777	Uncharacterized conserved prot
COG3809	Uncharacterized protein conser
COG3822	ABC-type sugar transport syste
COG3851	Signal transduction histidine
COG3871	Uncharacterized stress protein
COG3899	Predicted ATPase
COG3907	PAP2 (acid phosphatase) superf
COG3908	Uncharacterized protein conser
COG3911	Predicted ATPase

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG3913	Uncharacterized protein conser
COG3914	Predicted O-linked N-acetylglu
COG3915	Uncharacterized protein conser
COG3916	N-acyl-L-homoserine lactone sy
COG3918	Predicted membrane protein
COG3926	Putative secretion activating
COG3931	Predicted N-formylglutamate am
COG3933	Transcriptional antiterminator
COG3940	Predicted beta-xylosidase
COG3941	Mu-like prophage protein
COG3943	Virulence protein
COG3950	Predicted ATP-binding protein
COG3952	Predicted membrane protein
COG3953	SLT domain proteins
COG3955	Exopolysaccharide biosynthesis
COG3972	Superfamily I DNA and RNA heli
COG3978	Acetolactate synthase (isozyme
COG3979	Uncharacterized protein contai
COG4007	Predicted dehydrogenase relate
COG4015	Predicted dinucleotide-utilizi
COG4016	Uncharacterized protein conser
COG4017	Uncharacterized protein conser
COG4018	Uncharacterized protein conser

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4019	Uncharacterized protein conser
COG4022	Uncharacterized protein conser
COG4026	Uncharacterized protein contai
COG4030	Uncharacterized protein conser
COG4033	Uncharacterized protein conser
COG4036	Predicted membrane protein
COG4037	Predicted membrane protein
COG4041	Predicted membrane protein
COG4051	Uncharacterized protein conser
COG4053	Uncharacterized protein conser
COG4055	Methyl coenzyme M reductase, s
COG4057	Methyl coenzyme M reductase, g
COG4058	Methyl coenzyme M reductase, a
COG4062	Tetrahydromethanopterin S-meth
COG4069	Uncharacterized protein conser
COG4071	Uncharacterized protein conser
COG4073	Uncharacterized protein conser
COG4074	H ₂ -forming N ₅ ,N ₁₀ -methylenetet
COG4075	Uncharacterized conserved prot
COG4076	Predicted RNA methylase
COG4078	Predicted membrane protein
COG4081	Uncharacterized protein conser
COG4087	Soluble P-type ATPase

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4093	Uncharacterized protein conser
COG4099	Predicted peptidase
COG4103	Uncharacterized protein conser
COG4111	Uncharacterized conserved prot
COG4112	Predicted phosphoesterase (Mut
COG4124	Beta-mannanase
COG4135	ABC-type uncharacterized trans
COG4146	Predicted symporter
COG4158	Predicted ABC-type sugar trans
COG4161	ABC-type arginine transport sy
COG4167	ABC-type antimicrobial peptide
COG4168	ABC-type antimicrobial peptide
COG4170	ABC-type antimicrobial peptide
COG4171	ABC-type antimicrobial peptide
COG4192	Signal transduction histidine
COG4196	Uncharacterized protein conser
COG4219	Antirepressor regulating drug
COG4248	Uncharacterized protein with p
COG4259	Uncharacterized protein conser
COG4268	McrBC 5-methylcytosine restric
COG4270	Predicted membrane protein
COG4310	Uncharacterized protein conser
COG4315	Uncharacterized protein conser

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4321	Uncharacterized protein relate
COG4332	Uncharacterized protein conser
COG4335	DNA alkylation repair enzyme
COG4373	Mu-like prophage FluMu protein
COG4380	Uncharacterized protein conser
COG4383	Mu-like prophage protein gp29
COG4386	Mu-like prophage tail sheath p
COG4387	Mu-like prophage protein gp36
COG4388	Mu-like prophage I protein
COG4389	Site-specific recombinase
COG4392	Predicted membrane protein
COG4393	Predicted membrane protein
COG4396	Mu-like prophage host-nuclease
COG4397	Mu-like prophage major head su
COG4408	Uncharacterized protein conser
COG4413	Urea transporter
COG4454	Uncharacterized copper-binding
COG4455	Protein of avirulence locus in
COG4457	Uncharacterized protein conser
COG4458	Uncharacterized protein conser
COG4461	Uncharacterized protein conser
COG4487	Uncharacterized protein conser
COG4495	Uncharacterized protein conser

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4502	Uncharacterized protein conser
COG4506	Uncharacterized protein conser
COG4509	Uncharacterized protein conser
COG4512	Membrane protein putatively in
COG4517	Uncharacterized protein conser
COG4520	Surface antigen
COG4529	Uncharacterized protein conser
COG4530	Uncharacterized protein conser
COG4540	Phage P2 baseplate assembly pr
COG4541	Predicted membrane protein
COG4544	Uncharacterized conserved prot
COG4548	Nitric oxide reductase activat
COG4549	Uncharacterized protein conser
COG4564	Signal transduction histidine
COG4570	Holliday junction resolvase
COG4571	Outer membrane protease
COG4590	ABC-type uncharacterized trans
COG4594	ABC-type Fe ³⁺ -citrate transpor
COG4628	Uncharacterized conserved prot
COG4640	Predicted membrane protein
COG4646	DNA methylase
COG4652	Uncharacterized protein conser
COG4690	Dipeptidase

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4693	Oxidoreductase (NAD-binding),
COG4706	Predicted 3-hydroxylacyl-(acyl
COG4709	Predicted membrane protein
COG4714	Uncharacterized membrane-ancho
COG4716	Myosin-crossreactive antigen
COG4723	Phage-related protein, tail co
COG4727	Uncharacterized protein conser
COG4731	Uncharacterized protein conser
COG4732	Predicted membrane protein
COG4733	Phage-related protein, tail co
COG4737	Uncharacterized protein conser
COG4740	Predicted metalloprotease
COG4743	Predicted membrane protein
COG4747	ACT domain-containing protein
COG4750	CTP:phosphocholine cytidyltr
COG4752	Uncharacterized protein conser
COG4757	Predicted alpha/beta hydrolase
COG4771	Outer membrane receptor for fe
COG4772	Outer membrane receptor for Fe
COG4773	Outer membrane receptor for fe
COG4776	Exoribonuclease II
COG4781	Membrane domain of membrane-an
COG4785	Lipoprotein NlpI, contains TPR

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4802	Ferredoxin-thioredoxin reducta
COG4808	Uncharacterized protein conser
COG4810	Ethanolamine utilization prote
COG4811	Predicted membrane protein
COG4812	Ethanolamine utilization cobal
COG4814	Uncharacterized protein with a
COG4815	Uncharacterized protein conser
COG4823	Abortive infection bacterioph
COG4831	Uncharacterized conserved prot
COG4837	Uncharacterized protein conser
COG4841	Uncharacterized protein conser
COG4842	Uncharacterized protein conser
COG4843	Uncharacterized protein conser
COG4855	Uncharacterized protein conser
COG4863	Uncharacterized protein conser
COG4873	Uncharacterized protein conser
COG4875	Uncharacterized protein conser
COG4876	Uncharacterized protein conser
COG4883	Uncharacterized protein conser
COG4884	Uncharacterized protein conser
COG4902	Uncharacterized protein conser
COG4906	Predicted membrane protein
COG4909	Propanediol dehydratase, large

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG4910	Propanediol dehydratase, small
COG4913	Uncharacterized protein conser
COG4915	5-bromo-4-chloroindolyl phosph
COG4922	Uncharacterized protein conser
COG4924	Uncharacterized protein conser
COG4927	Predicted choloylglycine hydro
COG4930	Predicted ATP-dependent Lon-ty
COG4936	Predicted sensor domain
COG4943	Predicted signal transduction
COG4945	Membrane-anchored protein pred
COG4947	Uncharacterized protein conser
COG4954	Uncharacterized protein conser
COG4955	Uncharacterized protein conser
COG4957	Predicted transcriptional regu
COG4959	Type IV secretory pathway, pro
COG4970	Tfp pilus assembly protein Fim
COG4975	Putative glucose uptake permea
COG4984	Predicted membrane protein
COG4985	ABC-type phosphate transport s
COG5001	Predicted signal transduction
COG5108	Mitochondrial DNA-directed RNA
COG5212	Low-affinity cAMP phosphodiect
COG5301	Phage-related tail fibre prote

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG5307	SEC7 domain proteins
COG5314	Conjugal transfer/entry exclus
COG5323	Uncharacterized conserved prot
COG5330	Uncharacterized protein conser
COG5342	Invasion protein B, involved i
COG5345	Uncharacterized protein conser
COG5350	Predicted protein tyrosine pho
COG5351	Uncharacterized protein conser
COG5386	Cell surface protein
COG5387	Chaperone required for the ass
COG5388	Uncharacterized protein conser
COG5389	Uncharacterized protein conser
COG5395	Predicted membrane protein
COG5412	Phage-related protein
COG5425	Usg protein, probable subunit
COG5430	Uncharacterized secreted prote
COG5433	Transposase
COG5435	Uncharacterized conserved prot
COG5436	Predicted integral membrane pr
COG5437	Predicted secreted protein
COG5440	Uncharacterized conserved prot
COG5441	Uncharacterized conserved prot
COG5442	Flagellar biosynthesis regulat

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG5446	Predicted integral membrane pr
COG5447	Uncharacterized conserved prot
COG5448	Uncharacterized conserved prot
COG5449	Uncharacterized conserved prot
COG5451	Predicted secreted protein
COG5453	Uncharacterized conserved prot
COG5454	Predicted secreted protein
COG5455	Predicted integral membrane pr
COG5457	Uncharacterized conserved smal
COG5462	Predicted secreted (periplasmi
COG5468	Predicted secreted (periplasmi
COG5476	Uncharacterized conserved prot
COG5480	Predicted integral membrane pr
COG5481	Uncharacterized conserved smal
COG5488	Integral membrane protein
COG5502	Uncharacterized conserved prot
COG5503	Uncharacterized conserved smal
COG5515	Uncharacterized conserved smal
COG5519	Superfamily II helicase and in
COG5520	O-Glycosyl hydrolase
COG5521	Predicted integral membrane pr
COG5546	Small integral membrane protei
COG5556	Uncharacterized conserved prot

Continúa ...

Cuadro 20: COGs eliminados (continuación)

COG	Descripción
COG5568	Uncharacterized small protein
COG5570	Uncharacterized small protein
COG5571	Autotransporter protein or dom
COG5587	Uncharacterized conserved prot
COG5610	Predicted hydrolase (HAD super
COG5617	Predicted integral membrane pr
COG5618	Predicted periplasmic lipoprot
COG5621	Predicted secreted hydrolase
COG5631	Predicted transcription regula
COG5661	Predicted secreted Zn-dependen
COG5664	Predicted secreted Zn-dependen

Fin tabla

11. Apendice 3: Cuadro COGs eliminados en 3er. filtro

Cuadro con los COGs eliminados en el tercer filtro (ver 4.4 y 4.5).

Cuadro 21: COGs eliminados en el 3er. filtro

COG	Descripción
COG0001	Glutamate-1-semialdehyde amino
COG0002	Acetylglutamate semialdehyde d
COG0003	Oxyanion-translocating ATPase
COG0004	Ammonia permease
COG0005	Purine nucleoside phosphorylas
COG0006	Xaa-Pro aminopeptidase
COG0007	Uroporphyrinogen-III methylase
COG0009	Putative translation factor (S
COG0011	Uncharacterized conserved prot
COG0012	Predicted GTPase, probable tra
COG0013	Alanyl-tRNA synthetase
COG0014	Gamma-glutamyl phosphate reduc
COG0015	Adenylosuccinate lyase
COG0016	Phenylalanyl-tRNA synthetase a
COG0017	Aspartyl/asparaginyl-tRNA synt
COG0018	Arginyl-tRNA synthetase
COG0020	Undecaprenyl pyrophosphate syn
COG0021	Transketolase
COG0022	Pyruvate/2-oxoglutarate dehydr
COG0023	Translation initiation factor
COG0024	Methionine aminopeptidase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0026	Phosphoribosylaminoimidazole c
COG0027	Formate-dependent phosphoribos
COG0029	Aspartate oxidase
COG0030	Dimethyladenosine transferase
COG0033	Phosphoglucomutase
COG0034	Glutamine phosphoribosylpyroph
COG0035	Uracil phosphoribosyltransfera
COG0036	Pentose-5-phosphate-3-epimeras
COG0038	Chloride channel protein EriC
COG0039	Malate/lactate dehydrogenases
COG0040	ATP phosphoribosyltransferase
COG0041	Phosphoribosylcarboxyaminoimid
COG0043	3-polyprenyl-4-hydroxybenzoate
COG0045	Succinyl-CoA synthetase, beta
COG0046	Phosphoribosylformylglycinamid
COG0047	Phosphoribosylformylglycinamid
COG0048	Ribosomal protein S12
COG0049	Ribosomal protein S7
COG0050	GTPases - translation elongati
COG0051	Ribosomal protein S10
COG0053	Predicted Co/Zn/Cd cation tran
COG0054	Riboflavin synthase beta-chain
COG0055	F0F1-type ATP synthase, beta s

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0056	F0F1-type ATP synthase, alpha
COG0058	Glucan phosphorylase
COG0059	KetoL-acid reductoisomerase
COG0060	Isoleucyl-tRNA synthetase
COG0061	Predicted sugar kinase
COG0062	Uncharacterized conserved prot
COG0063	Predicted sugar kinase
COG0064	Asp-tRNA ^{Asn} /Glu-tRNA ^{Gln} amidot
COG0065	3-isopropylmalate dehydratase
COG0066	3-isopropylmalate dehydratase
COG0068	Hydrogenase maturation factor
COG0069	Glutamate synthase domain 2
COG0072	Phenylalanyl-tRNA synthetase b
COG0074	Succinyl-CoA synthetase, alpha
COG0075	Serine-pyruvate aminotransfera
COG0076	Glutamate decarboxylase and re
COG0077	Prephenate dehydratase
COG0078	Ornithine carbamoyltransferase
COG0080	Ribosomal protein L11
COG0081	Ribosomal protein L1
COG0082	Chorismate synthase
COG0083	Homoserine kinase
COG0084	Mg-dependent DNase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0087	Ribosomal protein L3
COG0090	Ribosomal protein L2
COG0092	Ribosomal protein S3
COG0093	Ribosomal protein L14
COG0094	Ribosomal protein L5
COG0096	Ribosomal protein S8
COG0097	Ribosomal protein L6P/L9E
COG0098	Ribosomal protein S5
COG0099	Ribosomal protein S13
COG0100	Ribosomal protein S11
COG0101	Pseudouridylate synthase
COG0103	Ribosomal protein S9
COG0104	Adenylosuccinate synthase
COG0105	Nucleoside diphosphate kinase
COG0106	Phosphoribosylformimino-5-amin
COG0107	Imidazoleglycerol-phosphate sy
COG0108	3,4-dihydroxy-2-butanone 4-pho
COG0109	Polyprenyltransferase (cytochr
COG0112	Glycine/serine hydroxymethyltr
COG0113	Delta-aminolevulinic acid dehy
COG0114	Fumarase
COG0116	Predicted N6-adenine-specific
COG0117	Pyrimidine deaminase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0118	Glutamine amidotransferase
COG0120	Ribose 5-phosphate isomerase
COG0123	Deacetylases, including yeast
COG0124	Histidyl-tRNA synthetase
COG0125	Thymidylate kinase
COG0126	3-phosphoglycerate kinase
COG0127	Xanthosine triphosphate pyroph
COG0128	5-enolpyruvylshikimate-3-phosp
COG0130	Pseudouridine synthase
COG0131	Imidazoleglycerol-phosphate de
COG0132	Dethiobiotin synthetase
COG0133	Tryptophan synthase beta chain
COG0134	Indole-3-glycerol phosphate sy
COG0135	Phosphoribosylanthranilate iso
COG0136	Aspartate-semialdehyde dehydro
COG0137	Argininosuccinate synthase
COG0138	AICAR transformylase/IMP cyclo
COG0139	Phosphoribosyl-AMP cyclohydrol
COG0140	Phosphoribosyl-ATP pyrophospho
COG0141	Histidinol dehydrogenase
COG0143	Methionyl-tRNA synthetase
COG0146	N-methylhydantoinase B/acetone
COG0148	Enolase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0150	Phosphoribosylaminoimidazole (
COG0151	Phosphoribosylamine-glycine li
COG0152	Phosphoribosylaminoimidazolesu
COG0153	Galactokinase
COG0155	Sulfite reductase, beta subuni
COG0156	7-keto-8-aminopelargonate synt
COG0157	Nicotinate-nucleotide pyrophos
COG0158	Fructose-1,6-bisphosphatase
COG0159	Tryptophan synthase alpha chai
COG0162	Tyrosyl-tRNA synthetase
COG0163	3-polyprenyl-4-hydroxybenzoate
COG0164	Ribonuclease HII
COG0165	Argininosuccinate lyase
COG0166	Glucose-6-phosphate isomerase
COG0167	Dihydroorotate dehydrogenase
COG0168	Trk-type K ⁺ transport systems,
COG0169	Shikimate 5-dehydrogenase
COG0170	Dolichol kinase
COG0171	NAD synthase
COG0172	Seryl-tRNA synthetase
COG0173	Aspartyl-tRNA synthetase
COG0176	Transaldolase
COG0177	Predicted EndoIII-related endo

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0178	Excinuclease ATPase subunit
COG0180	Tryptophanyl-tRNA synthetase
COG0181	Porphobilinogen deaminase
COG0182	Predicted translation initiati
COG0185	Ribosomal protein S19
COG0190	5,10-methylene-tetrahydrofolat
COG0191	Fructose/tagatose bisphosphate
COG0192	S-adenosylmethionine synthetas
COG0193	Peptidyl-tRNA hydrolase
COG0194	Guanylate kinase
COG0196	FAD synthase
COG0201	Preprotein translocase subunit
COG0203	Ribosomal protein L17
COG0205	6-phosphofructokinase
COG0206	Cell division GTPase
COG0208	Ribonucleotide reductase, beta
COG0209	Ribonucleotide reductase, alph
COG0210	Superfamily I DNA and RNA heli
COG0211	Ribosomal protein L27
COG0213	Thymidine phosphorylase
COG0214	Pyridoxine biosynthesis enzyme
COG0215	Cysteinyl-tRNA synthetase
COG0216	Protein chain release factor A

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0217	Uncharacterized conserved prot
COG0218	Predicted GTPase
COG0219	Predicted rRNA methylase (SpoU
COG0220	Predicted S-adenosylmethionine
COG0221	Inorganic pyrophosphatase
COG0222	Ribosomal protein L7/L12
COG0223	Methionyl-tRNA formyltransfera
COG0224	F0F1-type ATP synthase, gamma
COG0225	Peptide methionine sulfoxide r
COG0226	ABC-type phosphate transport s
COG0228	Ribosomal protein S16
COG0229	Conserved domain frequently as
COG0230	Ribosomal protein L34
COG0232	dGTP triphosphohydrolase
COG0233	Ribosome recycling factor
COG0234	Co-chaperonin GroES (HSP10)
COG0236	Acyl carrier protein
COG0238	Ribosomal protein S18
COG0239	Integral membrane protein poss
COG0240	Glycerol-3-phosphate dehydroge
COG0241	Histidinol phosphatase and rel
COG0242	N-formylmethionyl-tRNA deformy
COG0245	2C-methyl-D-erythritol 2,4-cyc

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0248	Exopolyphosphatase
COG0249	Mismatch repair ATPase (MutS f
COG0252	L-asparaginase/archaeal Glu-tR
COG0253	Diaminopimelate epimerase
COG0259	Pyridoxamine-phosphate oxidase
COG0260	Leucyl aminopeptidase
COG0261	Ribosomal protein L21
COG0263	Glutamate 5-kinase
COG0264	Translation elongation factor
COG0265	Trypsin-like serine proteases,
COG0266	Formamidopyrimidine-DNA glycos
COG0268	Ribosomal protein S20
COG0269	3-hexulose-6-phosphate synthas
COG0270	Site-specific DNA methylase
COG0272	NAD-dependent DNA ligase (cont
COG0274	Deoxyribose-phosphate aldolase
COG0275	Predicted S-adenosylmethionine
COG0276	Protoheme ferro-lyase (ferroch
COG0278	Glutaredoxin-related protein
COG0279	Phosphoheptose isomerase
COG0281	Malic enzyme
COG0282	Acetate kinase
COG0283	Cytidylate kinase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0285	Folypolyglutamate synthase
COG0286	Type I restriction-modificatio
COG0289	Dihydrodipicolinate reductase
COG0290	Translation initiation factor
COG0291	Ribosomal protein L35
COG0292	Ribosomal protein L20
COG0293	23S rRNA methylase
COG0295	Cytidine deaminase
COG0297	Glycogen synthase
COG0298	Hydrogenase maturation factor
COG0299	Folate-dependent phosphoribosy
COG0301	Thiamine biosynthesis ATP pyro
COG0302	GTP cyclohydrolase I
COG0303	Molybdopterin biosynthesis enz
COG0305	Replicative DNA helicase
COG0306	Phosphate/sulphate permeases
COG0307	Riboflavin synthase alpha chai
COG0308	Aminopeptidase N
COG0309	Hydrogenase maturation factor
COG0311	Predicted glutamine amidotrans
COG0313	Predicted methyltransferases
COG0314	Molybdopterin converting facto
COG0315	Molybdenum cofactor biosynthes

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0319	Predicted metal-dependent hydr
COG0320	Lipoate synthase
COG0321	Lipoate-protein ligase B
COG0322	Nuclease subunit of the excinu
COG0323	DNA mismatch repair enzyme (pr
COG0324	tRNA delta(2)-isopentenylpyrop
COG0325	Predicted enzyme with a TIM-ba
COG0326	Molecular chaperone, HSP90 fam
COG0327	Uncharacterized conserved prot
COG0332	3-oxoacyl-[acyl-carrier-protei
COG0335	Ribosomal protein L19
COG0336	tRNA-(guanine-N1)-methyltransf
COG0337	3-dehydroquinase synthetase
COG0338	Site-specific DNA methylase
COG0339	Zn-dependent oligopeptidases
COG0340	Biotin-(acetyl-CoA carboxylase
COG0341	Preprotein translocase subunit
COG0342	Preprotein translocase subunit
COG0343	Queuine/archaeosine tRNA-ribos
COG0344	Predicted membrane protein
COG0345	Pyrroline-5-carboxylate reduct
COG0347	Nitrogen regulatory protein PI
COG0350	Methylated DNA-protein cystein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0351	Hydroxymethylpyrimidine/phosph
COG0352	Thiamine monophosphate synthas
COG0353	Recombinational DNA repair pro
COG0354	Predicted aminomethyltransfera
COG0355	F0F1-type ATP synthase, epsilo
COG0356	F0F1-type ATP synthase, subuni
COG0357	Predicted S-adenosylmethionine
COG0359	Ribosomal protein L9
COG0360	Ribosomal protein S6
COG0362	6-phosphogluconate dehidrogena
COG0364	Glucose-6-phosphate 1-dehidrog
COG0367	Asparagine synthase (glutamine
COG0368	Cobalamin-5-phosphate synthase
COG0369	Sulfite reductase, alpha subun
COG0370	Fe ²⁺ transport system protein
COG0371	Glycerol dehidrogenase and rel
COG0372	Citrate synthase
COG0373	Glutamyl-tRNA reductase
COG0374	Ni,Fe-hydrogenase I large subu
COG0375	Zn finger protein HypA/HybF (p
COG0376	Catalase (peroxidase I)
COG0377	NADH:ubiquinone oxidoreductase
COG0378	Ni ²⁺ -binding GTPase involved i

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0379	Quinolinate synthase
COG0380	Trehalose-6-phosphate synthase
COG0381	UDP-N-acetylglucosamine 2-epim
COG0384	Predicted epimerase, PhzC/PhzF
COG0386	Glutathione peroxidase
COG0387	Ca ²⁺ /H ⁺ antiporter
COG0390	ABC-type uncharacterized trans
COG0393	Uncharacterized conserved prot
COG0396	ABC-type transport system invo
COG0397	Uncharacterized conserved prot
COG0398	Uncharacterized conserved prot
COG0401	Uncharacterized homolog of Blt
COG0403	Glycine cleavage system protei
COG0405	Gamma-glutamyltransferase
COG0407	Uroporphyrinogen-III decarboxy
COG0408	Coproporphyrinogen III oxidase
COG0409	Hydrogenase maturation factor
COG0413	Ketopantoate hydroxymethyltran
COG0414	Pantothenate synthetase
COG0415	Deoxyribodipyrimidine photolya
COG0416	Fatty acid/phospholipid biosyn
COG0418	Dihydroorotase
COG0419	ATPase involved in DNA repair

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0421	Spermidine synthase
COG0422	Thiamine biosynthesis protein
COG0423	Glycyl-tRNA synthetase (class
COG0424	Nucleotide-binding protein imp
COG0426	Uncharacterized flavoproteins
COG0427	Acetyl-CoA hydrolase
COG0428	Predicted divalent heavy-metal
COG0429	Predicted hydrolase of the alp
COG0430	RNA 3'-terminal phosphate cycl
COG0432	Uncharacterized conserved prot
COG0434	Predicted TIM-barrel enzyme
COG0435	Predicted glutathione S-transf
COG0437	Fe-S-cluster-containing hydrog
COG0439	Biotin carboxylase
COG0440	Acetolactate synthase, small (
COG0441	Threonyl-tRNA synthetase
COG0442	Prolyl-tRNA synthetase
COG0445	NAD/FAD-utilizing enzyme appar
COG0447	Dihydroxynaphthoic acid synt ha
COG0449	Glucosamine 6-phosphate synthe
COG0450	Peroxiredoxin
COG0452	Phosphopantothenoilcysteine sy
COG0455	ATPases involved in chromosome

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0456	Acetyltransferases
COG0458	Carbamoylphosphate synthase la
COG0460	Homoserine dehydrogenase
COG0461	Orotate phosphoribosyltransfer
COG0462	Phosphoribosylpyrophosphate sy
COG0465	ATP-dependent Zn proteases
COG0466	ATP-dependent Lon protease, ba
COG0467	RecA-superfamily ATPases impli
COG0468	RecA/RadA recombinase
COG0469	Pyruvate kinase
COG0473	Isocitrate/isopropylmalate deh
COG0474	Cation transport ATPase
COG0475	Kef-type K ⁺ transport systems,
COG0476	Dinucleotide-utilizing enzymes
COG0478	RIO-like serine/threonine prot
COG0479	Succinate dehydrogenase/fumara
COG0480	Translation elongation factors
COG0481	Membrane GTPase LepA
COG0482	Predicted tRNA(5-methylaminome
COG0484	DnaJ-class molecular chaperone
COG0486	Predicted GTPase
COG0490	Putative regulatory, ligand-bi
COG0496	Predicted acid phosphatase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0497	ATPase involved in DNA repair
COG0499	S-adenosylhomocysteine hydrola
COG0502	Biotin synthase and related en
COG0504	CTP synthase (UTP-ammonia lyas
COG0505	Carbamoylphosphate synthase sm
COG0506	Proline dehydrogenase
COG0507	ATP-dependent exoDNase (exonuc
COG0509	Glycine cleavage system H prot
COG0512	Anthranilate/para-aminobenzoat
COG0514	Superfamily II DNA helicase
COG0519	GMP synthase, PP-ATPase domain
COG0521	Molybdopterin biosynthesis enz
COG0525	Valyl-tRNA synthetase
COG0527	Aspartokinases
COG0528	Uridylate kinase
COG0529	Adenylylsulfate kinase and rel
COG0530	Ca ²⁺ /Na ⁺ antiporter
COG0532	Translation initiation factor
COG0533	Metal-dependent proteases with
COG0536	Predicted GTPase
COG0539	Ribosomal protein S1
COG0540	Aspartate carbamoyltransferase
COG0541	Signal recognition particle GT

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0544	FKBP-type peptidyl-prolyl cis-
COG0545	FKBP-type peptidyl-prolyl cis-
COG0548	Acetylglutamate kinase
COG0549	Carbamate kinase
COG0550	Topoisomerase IA
COG0551	Zn-finger domain associated wi
COG0552	Signal recognition particle GT
COG0554	Glycerol kinase
COG0555	ABC-type sulfate transport sys
COG0556	Helicase subunit of the DNA ex
COG0562	UDP-galactopyranose mutase
COG0565	rRNA methylase
COG0567	2-oxoglutarate dehydrogenase c
COG0571	dsRNA-specific ribonuclease
COG0572	Uridine kinase
COG0573	ABC-type phosphate transport s
COG0576	Molecular chaperone GrpE (heat
COG0578	Glycerol-3-phosphate dehydroge
COG0580	Glycerol uptake facilitator an
COG0581	ABC-type phosphate transport s
COG0585	Uncharacterized conserved prot
COG0587	DNA polymerase III, alpha subu
COG0588	Phosphoglycerate mutase 1

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0590	Cytosine/adenosine deaminases
COG0593	ATPase involved in DNA replica
COG0594	RNase P protein component
COG0595	Predicted hydrolase of the met
COG0597	Lipoprotein signal peptidase
COG0603	Predicted PP-loop superfamily
COG0605	Superoxide dismutase
COG0606	Predicted ATPase with chaperon
COG0610	Type I site-specific restricti
COG0611	Thiamine monophosphate kinase
COG0615	Cytidyltransferase
COG0617	tRNA nucleotidyltransferase/po
COG0620	Methionine synthase II (cobala
COG0623	Enoyl-[acyl-carrier-protein] r
COG0627	Predicted esterase
COG0629	Single-stranded DNA-binding pr
COG0631	Serine/threonine protein phosp
COG0632	Holliday junction resolvasome,
COG0633	Ferredoxin
COG0634	Hypoxanthine-guanine phosphori
COG0641	Arylsulfatase regulator (Fe-S
COG0643	Chemotaxis protein histidine k
COG0645	Predicted kinase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0646	Methionine synthase I (cobalam
COG0649	NADH:ubiquinone oxidoreductase
COG0650	Formate hydrogenlyase subunit
COG0652	Peptidyl-prolyl cis-trans isom
COG0653	Preprotein translocase subunit
COG0658	Predicted membrane metal-bindi
COG0661	Predicted unusual protein kina
COG0662	Mannose-6-phosphate isomerase
COG0666	FOG: Ankyrin repeat
COG0669	Phosphopantetheine adenylyltra
COG0670	Integral membrane protein, int
COG0672	High-affinity Fe ²⁺ /Pb ²⁺ permea
COG0676	Uncharacterized enzymes relate
COG0677	UDP-N-acetyl-D-mannosaminurona
COG0679	Predicted permeases
COG0680	Ni,Fe-hydrogenase maturation f
COG0682	Prolipoprotein diacylglycerylt
COG0684	Demethylmenaquinone methyltran
COG0686	Alanine dehydrogenase
COG0687	Spermidine/putrescine-binding
COG0689	RNase PH
COG0690	Preprotein translocase subunit
COG0691	tmRNA-binding protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0692	Uracil DNA glycosylase
COG0694	Thioredoxin-like proteins and
COG0695	Glutaredoxin and related prote
COG0696	Phosphoglyceromutase
COG0698	Ribose 5-phosphate isomerase R
COG0700	Uncharacterized membrane prote
COG0703	Shikimate kinase
COG0706	Preprotein translocase subunit
COG0707	UDP-N-acetylglucosamine:LPS N-
COG0708	Exonuclease III
COG0709	Selenophosphate synthase
COG0710	3-dehydroquinate dehydratase
COG0711	F0F1-type ATP synthase, subuni
COG0712	F0F1-type ATP synthase, delta
COG0713	NADH:ubiquinone oxidoreductase
COG0716	Flavodoxins
COG0718	Uncharacterized protein conser
COG0721	Asp-tRNA ^{Asn} /Glu-tRNA ^{Gln} amidot
COG0722	3-deoxy-D-arabino-heptulosonate
COG0724	RNA-binding proteins (RRM doma
COG0727	Predicted Fe-S-cluster oxidore
COG0728	Uncharacterized membrane prote
COG0729	Outer membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0733	Na ⁺ -dependent transporters of
COG0736	Phosphopantetheinyl transferas
COG0738	Fucose permease
COG0742	N ⁶ -adenine-specific methylase
COG0743	1-deoxy-D-xylulose 5-phosphate
COG0748	Putative heme iron utilization
COG0749	DNA polymerase I - 3'-5' exonu
COG0751	Glycyl-tRNA synthetase, beta s
COG0752	Glycyl-tRNA synthetase, alpha
COG0754	Glutathionylspermidine synthas
COG0755	ABC-type transport system invo
COG0756	dUTPase
COG0757	3-dehydroquinate dehydratase I
COG0758	Predicted Rossmann fold nucleo
COG0759	Uncharacterized conserved prot
COG0761	Penicillin tolerance protein
COG0763	Lipid A disaccharide synthetas
COG0767	ABC-type transport system invo
COG0770	UDP-N-acetylmuramyl pentapepti
COG0771	UDP-N-acetylmuramoylalanine-D-
COG0774	UDP-3-O-acyl-N-acetylglucosami
COG0777	Acetyl-CoA carboxylase beta su
COG0779	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0780	Enzyme related to GTP cyclohyd
COG0781	Transcription termination fact
COG0783	DNA-binding ferritin-like prot
COG0785	Cytochrome c biogenesis protei
COG0786	Na ⁺ /glutamate symporter
COG0787	Alanine racemase
COG0788	Formyltetrahydrofolate hydrola
COG0792	Predicted endonuclease distant
COG0796	Glutamate racemase
COG0797	Lipoproteins
COG0798	Arsenite efflux pump ACR3 and
COG0799	Uncharacterized homolog of pla
COG0800	2-keto-3-deoxy-6-phosphoglucon
COG0801	7,8-dihydro-6-hydroxymethylpte
COG0802	Predicted ATPase or kinase
COG0804	Urea amidohydrolase (urease) a
COG0805	Sec-independent protein secret
COG0806	RimM protein, required for 16S
COG0807	GTP cyclohydrolase II
COG0809	S-adenosylmethionine:tRNA-ribo
COG0810	Periplasmic protein TonB, link
COG0812	UDP-N-acetylmuramate dehydroge
COG0813	Purine-nucleoside phosphorylas

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0814	Amino acid permeases
COG0815	Apolipoprotein N-acyltransfera
COG0816	Predicted endonuclease involve
COG0817	Holliday junction resolvasome,
COG0818	Diacylglycerol kinase
COG0819	Putative transcription activat
COG0820	Predicted Fe-S-cluster redox e
COG0821	Enzyme involved in the deoxyxy
COG0822	NifU homolog involved in Fe-S
COG0823	Periplasmic component of the T
COG0825	Acetyl-CoA carboxylase alpha s
COG0827	Adenine-specific DNA methylase
COG0828	Ribosomal protein S21
COG0829	Urease accessory protein UreH
COG0830	Urease accessory protein UreF
COG0831	Urea amidohydrolase (urease) g
COG0832	Urea amidohydrolase (urease) b
COG0833	Amino acid transporters
COG0836	Mannose-1-phosphate guanylyltr
COG0837	Glucokinase
COG0838	NADH:ubiquinone oxidoreductase
COG0839	NADH:ubiquinone oxidoreductase
COG0846	NAD-dependent protein deacetyl

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG0847	DNA polymerase III, epsilon su
COG0849	Actin-like ATPase involved in
COG0851	Septum formation topological s
COG0852	NADH:ubiquinone oxidoreductase
COG0853	Aspartate 1-decarboxylase
COG0854	Pyridoxal phosphate biosynthes
COG0855	Polyphosphate kinase
COG0856	Orotate phosphoribosyltransfer
COG0857	BioD-like N-terminal domain of
COG0858	Ribosome-binding factor A
COG0860	N-acetylmuramoyl-L-alanine ami
COG0864	Predicted transcriptional regu
COG1001	Adenine deaminase
COG1002	Type II restriction enzyme, me
COG1003	Glycine cleavage system protei
COG1004	Predicted UDP-glucose 6-dehydr
COG1005	NADH:ubiquinone oxidoreductase
COG1006	Multisubunit Na ⁺ /H ⁺ antiporter
COG1007	NADH:ubiquinone oxidoreductase
COG1008	NADH:ubiquinone oxidoreductase
COG1009	NADH:ubiquinone oxidoreductase
COG1010	Precorrin-3B methylase
COG1015	Phosphopentomutase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1017	Hemoglobin-like flavoprotein
COG1019	Predicted nucleotidyltransfera
COG1021	Peptide arylation enzymes
COG1022	Long-chain acyl-CoA synthetase
COG1023	Predicted 6-phosphogluconate d
COG1025	Secreted/periplasmic Zn-depend
COG1026	Predicted Zn-dependent peptida
COG1027	Aspartate ammonia-lyase
COG1029	Formylmethanofuran dehydrogena
COG1030	Membrane-bound serine protease
COG1031	Uncharacterized Fe-S oxidoredu
COG1033	Predicted exporters of the RND
COG1034	NADH dehydrogenase/NADH:ubiqui
COG1035	Coenzyme F420-reducing hydroge
COG1036	Archaeal flavoproteins
COG1038	Pyruvate carboxylase
COG1039	Ribonuclease HIII
COG1040	Predicted amidophosphoribosylt
COG1041	Predicted DNA modification met
COG1042	Acyl-CoA synthetase (NDP formi
COG1043	Acyl-[acyl carrier protein]-U
COG1044	UDP-3-O-[3-hydroxymyristoyl] g
COG1045	Serine acetyltransferase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1047	FKBP-type peptidyl-prolyl cis-
COG1048	Aconitase A
COG1049	Aconitase B
COG1054	Predicted sulfurtransferase
COG1055	Na ⁺ /H ⁺ antiporter NhaD and rel
COG1056	Nicotinamide mononucleotide ad
COG1057	Nicotinic acid mononucleotide
COG1058	Predicted nucleotide-utilizing
COG1059	Thermostable 8-oxoguanine DNA
COG1062	Zn-dependent alcohol dehydroge
COG1066	Predicted ATP-dependent serine
COG1067	Predicted ATP-dependent protea
COG1069	Ribulose kinase
COG1071	Pyruvate/2-oxoglutarate dehydr
COG1072	Panthothenate kinase
COG1074	ATP-dependent exoDNAse (exonuc
COG1075	Predicted acetyltransferases a
COG1078	HD superfamily phosphohydrolas
COG1079	Uncharacterized ABC-type trans
COG1080	Phosphoenolpyruvate-protein ki
COG1084	Predicted GTPase
COG1085	Galactose-1-phosphate uridylyl
COG1086	Predicted nucleoside-diphospha

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1087	UDP-glucose 4-epimerase
COG1088	dTDP-D-glucose 4,6-dehydratase
COG1089	GDP-D-mannose dehydratase
COG1090	Predicted nucleoside-diphospha
COG1091	dTDP-4-dehydrorhamnose reducta
COG1092	Predicted SAM-dependent methyl
COG1093	Translation initiation factor
COG1094	Predicted RNA-binding protein
COG1095	DNA-directed RNA polymerase, s
COG1096	Predicted RNA-binding protein
COG1097	RNA-binding protein Rrp4 and r
COG1099	Predicted metal-dependent hydr
COG1101	ABC-type uncharacterized trans
COG1102	Cytidylate kinase
COG1103	Archaea-specific pyridoxal pho
COG1104	Cysteine sulfinat desulfina
COG1105	Fructose-1-phosphate kinase an
COG1107	Archaea-specific RecJ-like exo
COG1110	Reverse gyrase
COG1111	ERCC4-like helicases
COG1113	Gamma-aminobutyrate permease a
COG1114	Branched-chain amino acid perm
COG1117	ABC-type phosphate transport s

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1118	ABC-type sulfate/molybdate tra
COG1119	ABC-type molybdenum transport
COG1123	ATPase components of various A
COG1125	ABC-type proline/glycine betai
COG1127	ABC-type transport system invo
COG1133	ABC-type long-chain fatty acid
COG1134	ABC-type polysaccharide/polyol
COG1135	ABC-type metal ion transport s
COG1137	ABC-type (unclassified) transp
COG1138	Cytochrome c biogenesis factor
COG1139	Uncharacterized conserved prot
COG1140	Nitrate reductase beta subunit
COG1141	Ferredoxin
COG1142	Fe-S-cluster-containing hydrog
COG1143	Formate hydrogenlyase subunit
COG1144	Pyruvate:ferredoxin oxidoreduc
COG1146	Ferredoxin
COG1148	Heterodisulfide reductase, sub
COG1150	Heterodisulfide reductase, sub
COG1151	6Fe-6S prismatic cluster-contai
COG1152	CO dehydrogenase/acetyl-CoA sy
COG1153	Formylmethanofuran dehydrogena
COG1154	Deoxyxylulose-5-phosphate synt

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1155	Archaeal/vacuolar-type H ⁺ -ATPa
COG1156	Archaeal/vacuolar-type H ⁺ -ATPa
COG1157	Flagellar biosynthesis/type II
COG1158	Transcription termination fact
COG1159	GTPase
COG1160	Predicted GTPases
COG1162	Predicted GTPases
COG1163	Predicted GTPase
COG1164	Oligoendopeptidase F
COG1165	2-succinyl-6-hydroxy-2,4-cyclo
COG1166	Arginine decarboxylase (spermi
COG1168	Bifunctional PLP-dependent enz
COG1169	Isochorismate synthase
COG1171	Threonine dehydratase
COG1172	Ribose/xylose/arabinose/galact
COG1174	ABC-type proline/glycine betai
COG1176	ABC-type spermidine/putrescine
COG1177	ABC-type spermidine/putrescine
COG1179	Dinucleotide-utilizing enzymes
COG1182	Acyl carrier protein phosphodi
COG1183	Phosphatidylserine synthase
COG1184	Translation initiation factor
COG1185	Polyribonucleotide nucleotidyl

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1189	Predicted rRNA methylase
COG1190	Lysyl-tRNA synthetase (class I
COG1193	Mismatch repair ATPase (MutS f
COG1194	A/G-specific DNA glycosylase
COG1195	Recombinational DNA repair ATP
COG1196	Chromosome segregation ATPases
COG1197	Transcription-repair coupling
COG1198	Primosomal protein N ¹ (replica
COG1200	RecG-like helicase
COG1201	Lhr-like helicases
COG1202	Superfamily II helicase, archa
COG1203	Predicted helicases
COG1204	Superfamily II helicase
COG1206	NAD(FAD)-utilizing enzyme poss
COG1207	N-acetylglucosamine-1-phosphat
COG1209	dTDP-glucose pyrophosphorylase
COG1210	UDP-glucose pyrophosphorylase
COG1211	4-diphosphocytidyl-2-methyl-D-
COG1212	CMP-2-keto-3-deoxyoctulosonic
COG1213	Predicted sugar nucleotidyltra
COG1214	Inactive homolog of metal-depe
COG1217	Predicted membrane GTPase invo
COG1218	3'-Phosphoadenosine 5'-phospho

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1219	ATP-dependent protease Clp, AT
COG1220	ATP-dependent protease HslVU (
COG1222	ATP-dependent 26S proteasome r
COG1223	Predicted ATPase (AAA+ superfa
COG1224	DNA helicase TIP49, TBP-intera
COG1225	Peroxiredoxin
COG1227	Inorganic pyrophosphatase/exop
COG1229	Formylmethanofuran dehydrogena
COG1230	Co/Zn/Cd efflux system compone
COG1231	Monoamine oxidase
COG1232	Protoporphyrinogen oxidase
COG1237	Metal-dependent hydrolases of
COG1240	Mg-chelatase subunit ChlD
COG1241	Predicted ATPase involved in r
COG1242	Predicted Fe-S oxidoreductase
COG1243	Histone acetyltransferase
COG1244	Predicted Fe-S oxidoreductase
COG1245	Predicted ATPase, RNase L inhi
COG1246	N-acetylglutamate synthase and
COG1247	Sortase and related acyltransf
COG1251	NAD(P)H-nitrite reductase
COG1254	Acylphosphatases
COG1255	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1256	Flagellar hook-associated prot
COG1257	Hydroxymethylglutaryl-CoA redu
COG1258	Predicted pseudouridylate synt
COG1259	Uncharacterized conserved prot
COG1260	Myo-inositol-1-phosphate synth
COG1262	Uncharacterized conserved prot
COG1266	Predicted metal-dependent memb
COG1268	Uncharacterized conserved prot
COG1269	Archaeal/vacuolar-type H ⁺ -ATPa
COG1270	Cobalamin biosynthesis protein
COG1271	Cytochrome bd-type quinol oxid
COG1272	Predicted membrane protein, he
COG1273	Uncharacterized conserved prot
COG1274	Phosphoenolpyruvate carboxykin
COG1276	Putative copper export protein
COG1279	Lysine efflux permease
COG1281	Disulfide bond chaperones of t
COG1282	NAD/NADP transhydrogenase beta
COG1283	Na ⁺ /phosphate symporter
COG1285	Uncharacterized membrane prote
COG1287	Uncharacterized membrane prote
COG1288	Predicted membrane protein
COG1289	Predicted membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1290	Cytochrome b subunit of the bc
COG1291	Flagellar motor component
COG1292	Choline-glycine betaine transp
COG1293	Predicted RNA-binding protein
COG1294	Cytochrome bd-type quinol oxid
COG1295	Predicted membrane protein
COG1296	Predicted branched-chain amino
COG1297	Predicted membrane protein
COG1298	Flagellar biosynthesis pathway
COG1299	Phosphotransferase system, fru
COG1300	Uncharacterized membrane prote
COG1301	Na ⁺ /H ⁺ -dicarboxylate symporter
COG1303	Uncharacterized protein conser
COG1307	Uncharacterized protein conser
COG1308	Transcription factor homologou
COG1310	Predicted metal-dependent prot
COG1311	Archaeal DNA polymerase II, sm
COG1312	D-mannonate dehydratase
COG1313	Uncharacterized Fe-S protein P
COG1314	Preprotein translocase subunit
COG1315	Predicted polymerase, most pro
COG1316	Transcriptional regulator
COG1317	Flagellar biosynthesis/type II

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1318	Predicted transcriptional regu
COG1319	Aerobic-type carbon monoxide d
COG1320	Multisubunit Na ⁺ /H ⁺ antiporter
COG1321	Mn-dependent transcriptional r
COG1322	Uncharacterized protein conser
COG1323	Predicted nucleotidyltransfera
COG1324	Uncharacterized protein involv
COG1325	Predicted exosome subunit
COG1326	Uncharacterized archaeal Zn-fi
COG1327	Predicted transcriptional regu
COG1328	Oxygen-sensitive ribonucleosid
COG1329	Transcriptional regulators, si
COG1330	Exonuclease V gamma subunit
COG1331	Highly conserved protein conta
COG1332	Uncharacterized protein predic
COG1333	ResB protein required for cyto
COG1336	Uncharacterized protein predic
COG1337	Uncharacterized protein predic
COG1338	Flagellar biosynthesis pathway
COG1339	Transcriptional regulator of a
COG1340	Uncharacterized archaeal coile
COG1341	Predicted GTPase or GTP-bindin
COG1342	Predicted DNA-binding proteins

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1343	Uncharacterized protein predic
COG1345	Flagellar capping protein
COG1346	Putative effector of murein hy
COG1347	Na ⁺ -transporting NADH:ubiquino
COG1348	Nitrogenase subunit NifH (ATPa
COG1350	Predicted alternative tryptoph
COG1351	Predicted alternative thymidyl
COG1352	Methylase of chemotaxis methyl
COG1355	Predicted dioxygenase
COG1356	Uncharacterized protein conser
COG1359	Uncharacterized conserved prot
COG1361	S-layer domain
COG1362	Aspartyl aminopeptidase
COG1364	N-acetylglutamate synthase (N-
COG1365	Predicted ATPase (PP-loop supe
COG1366	Anti-anti-sigma regulatory fac
COG1368	Phosphoglycerol transferase an
COG1369	RNase P/RNase MRP subunit POP5
COG1370	Prefoldin, molecular chaperone
COG1371	Uncharacterized conserved prot
COG1372	Intein/homing endonuclease
COG1374	Protein involved in ribosomal
COG1377	Flagellar biosynthesis pathway

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1378	Predicted transcriptional regu
COG1379	Uncharacterized conserved prot
COG1380	Putative effector of murein hy
COG1382	Prefoldin, chaperonin cofactor
COG1383	Ribosomal protein S17E
COG1384	Lysyl-tRNA synthetase (class I
COG1385	Uncharacterized protein conser
COG1386	Predicted transcriptional regu
COG1387	Histidinol phosphatase and rel
COG1388	FOG: LysM repeat
COG1389	DNA topoisomerase VI, subunit
COG1390	Archaeal/vacuolar-type H ⁺ -ATPa
COG1391	Glutamine synthetase adenylyt
COG1394	Archaeal/vacuolar-type H ⁺ -ATPa
COG1395	Predicted transcriptional regu
COG1397	ADP-ribosylglycohydrolase
COG1398	Fatty-acid desaturase
COG1399	Predicted metal-binding, possi
COG1400	Signal recognition particle 19
COG1401	GTPase subunit of restriction
COG1402	Uncharacterized protein, putat
COG1403	Restriction endonuclease
COG1404	Subtilisin-like serine proteas

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1405	Transcription initiation facto
COG1406	Predicted inhibitor of MCP met
COG1407	Predicted ICC-like phosphoeste
COG1408	Predicted phosphohydrolases
COG1410	Methionine synthase I, cobalam
COG1411	Uncharacterized protein relate
COG1412	Uncharacterized proteins of Pi
COG1413	FOG: HEAT repeat
COG1415	Uncharacterized conserved prot
COG1419	Flagellar GTP-binding protein
COG1420	Transcriptional regulator of h
COG1422	Predicted membrane protein
COG1423	ATP-dependent DNA ligase, homo
COG1426	Uncharacterized protein conser
COG1427	Predicted periplasmic solute-b
COG1428	Deoxynucleoside kinases
COG1429	Cobalamin biosynthesis protein
COG1430	Uncharacterized conserved prot
COG1435	Thymidine kinase
COG1436	Archaeal/vacuolar-type H ⁺ -ATPa
COG1437	Adenylate cyclase, class 2 (th
COG1438	Arginine repressor
COG1439	Predicted nucleic acid-binding

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1440	Phosphotransferase system cell
COG1443	Isopentenylidiphosphate isomera
COG1444	Predicted P-loop ATPase fused
COG1445	Phosphotransferase system fruc
COG1446	Asparaginase
COG1447	Phosphotransferase system cell
COG1448	Aspartate/tyrosine/aromatic am
COG1449	Alpha-amylase/alpha-mannosidas
COG1450	Type II secretory pathway, com
COG1451	Predicted metal-dependent hydr
COG1452	Organic solvent tolerance prot
COG1453	Predicted oxidoreductases of t
COG1455	Phosphotransferase system cell
COG1456	CO dehydrogenase/acetyl-CoA sy
COG1457	Purine-cytosine permease and r
COG1458	Predicted DNA-binding protein
COG1459	Type II secretory pathway, com
COG1460	Uncharacterized protein conser
COG1461	Predicted kinase related to di
COG1464	ABC-type metal ion transport s
COG1465	Predicted alternative 3-dehydr
COG1467	Eukaryotic-type DNA primase, c
COG1468	RecB family exonuclease

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1469	Uncharacterized conserved prot
COG1470	Predicted membrane protein
COG1471	Ribosomal protein S4E
COG1474	Cdc6-related protein, AAA supe
COG1477	Membrane-associated lipoprotei
COG1478	Uncharacterized conserved prot
COG1479	Uncharacterized conserved prot
COG1480	Predicted membrane-associated
COG1481	Uncharacterized protein conser
COG1483	Predicted ATPase (AAA+ superfa
COG1485	Predicted ATPase
COG1487	Predicted nucleic acid-binding
COG1489	DNA-binding protein, stimulate
COG1490	D-Tyr-tRNA ^{Tyr} deacylase
COG1491	Predicted RNA-binding protein
COG1492	Cobyric acid synthase
COG1493	Serine kinase of the HPr prote
COG1494	Fructose-1,6-bisphosphatase/se
COG1495	Disulfide bond formation prote
COG1496	Uncharacterized conserved prot
COG1497	Predicted transcriptional regu
COG1498	Protein implicated in ribosoma
COG1499	NMD protein affecting ribosome

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1500	Predicted exosome subunit
COG1501	Alpha-glucosidases, family 31
COG1503	Peptide chain release factor 1
COG1504	Uncharacterized conserved prot
COG1505	Serine proteases of the peptid
COG1507	Uncharacterized conserved prot
COG1508	DNA-directed RNA polymerase sp
COG1509	Lysine 2,3-aminomutase
COG1510	Predicted transcriptional regu
COG1511	Predicted membrane protein
COG1512	Beta-propeller domains of meth
COG1513	Cyanate lyase
COG1514	2'-5' RNA ligase
COG1515	Deoxyinosine 3'endonuclease (e
COG1516	Flagellin-specific chaperone F
COG1517	Uncharacterized protein predic
COG1518	Uncharacterized protein predic
COG1519	3-deoxy-D-manno-octulosonic-ac
COG1520	FOG: WD40-like repeat
COG1521	Putative transcriptional regul
COG1523	Type II secretory pathway, pul
COG1524	Uncharacterized proteins of th
COG1525	Micrococcal nuclease (thermonu

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1526	Uncharacterized protein requir
COG1527	Archaeal/vacuolar-type H ⁺ -ATPa
COG1528	Ferritin-like protein
COG1533	DNA repair photolyase
COG1535	Isochorismate hydrolase
COG1536	Flagellar motor switch protein
COG1537	Predicted RNA-binding proteins
COG1539	Dihydroneopterin aldolase
COG1540	Uncharacterized proteins, homo
COG1541	Coenzyme F390 synthetase
COG1542	Uncharacterized conserved prot
COG1543	Uncharacterized conserved prot
COG1544	Ribosome-associated protein Y
COG1546	Uncharacterized protein (compe
COG1548	Predicted transcriptional regu
COG1549	Queuine tRNA-ribosyltransferas
COG1550	Uncharacterized protein conser
COG1551	Carbon storage regulator (coul
COG1552	Ribosomal protein L40E
COG1553	Uncharacterized conserved prot
COG1554	Trehalose and maltose hydrolas
COG1555	DNA uptake protein and related
COG1556	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1558	Flagellar basal body rod prote
COG1559	Predicted periplasmic solute-b
COG1561	Uncharacterized stress-induced
COG1564	Thiamine pyrophosphokinase
COG1565	Uncharacterized conserved prot
COG1566	Multidrug resistance efflux pu
COG1568	Predicted methyltransferases
COG1569	Predicted nucleic acid-binding
COG1570	Exonuclease VII, large subunit
COG1571	Predicted DNA-binding protein
COG1572	Uncharacterized conserved prot
COG1574	Predicted metal-dependent hydr
COG1575	1,4-dihydroxy-2-naphthoate oct
COG1576	Uncharacterized conserved prot
COG1577	Mevalonate kinase
COG1578	Uncharacterized conserved prot
COG1579	Zn-ribbon protein, possibly nu
COG1580	Flagellar basal body-associate
COG1581	Archaeal DNA-binding protein
COG1582	Uncharacterized protein, possi
COG1583	Uncharacterized protein predic
COG1584	Predicted membrane protein
COG1585	Membrane protein implicated in

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1586	S-adenosylmethionine decarboxy
COG1587	Uroporphyrinogen-III synthase
COG1588	RNase P/RNase MRP subunit p29
COG1590	Uncharacterized conserved prot
COG1591	Holliday junction resolvase -
COG1592	Rubryerythrin
COG1594	DNA-directed RNA polymerase, s
COG1596	Periplasmic protein involved i
COG1598	Uncharacterized conserved prot
COG1599	Single-stranded DNA-binding re
COG1600	Uncharacterized Fe-S protein
COG1601	Translation initiation factor
COG1602	Uncharacterized conserved prot
COG1603	RNase P/RNase MRP subunit p30
COG1604	Uncharacterized protein predic
COG1606	ATP-utilizing enzymes of the P
COG1608	Predicted archaeal kinase
COG1610	Uncharacterized conserved prot
COG1613	ABC-type sulfate transport sys
COG1614	CO dehydrogenase/acetyl-CoA sy
COG1615	Uncharacterized conserved prot
COG1618	Predicted nucleotide kinase
COG1619	Uncharacterized proteins, homo

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1620	L-lactate permease
COG1621	Beta-fructosidasas (levanase/i
COG1623	Predicted nucleic-acid-binding
COG1625	Fe-S oxidoreductase, related t
COG1626	Neutral trehalase
COG1628	Uncharacterized conserved prot
COG1631	Ribosomal protein L44E
COG1632	Ribosomal protein L15E
COG1634	Uncharacterized Rossmann fold
COG1635	Flavoprotein involved in thiaz
COG1636	Uncharacterized protein conser
COG1637	Predicted nuclease of the RecB
COG1639	Predicted signal transduction
COG1640	4-alpha-glucanotransferase
COG1641	Uncharacterized conserved prot
COG1643	HrpA-like helicases
COG1644	DNA-directed RNA polymerase, s
COG1645	Uncharacterized Zn-finger cont
COG1646	Predicted phosphate-binding en
COG1647	Esterase/lipase
COG1648	Siroheme synthase (precorrin-2
COG1649	Uncharacterized protein conser
COG1650	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1652	Uncharacterized protein contai
COG1654	Biotin operon repressor
COG1655	Uncharacterized protein conser
COG1656	Uncharacterized conserved prot
COG1657	Squalene cyclase
COG1658	Small primase-like proteins (T
COG1660	Predicted P-loop-containing ki
COG1661	Predicted DNA-binding protein
COG1663	Tetraacyldisaccharide-1-P 4'-k
COG1664	Integral membrane protein CcmA
COG1665	Uncharacterized protein conser
COG1666	Uncharacterized protein conser
COG1669	Predicted nucleotidyltransfera
COG1671	Uncharacterized protein conser
COG1672	Predicted ATPase (AAA+ superfa
COG1673	Uncharacterized protein conser
COG1675	Transcription initiation facto
COG1676	tRNA splicing endonuclease
COG1677	Flagellar hook-basal body prot
COG1678	Putative transcriptional regul
COG1679	Uncharacterized conserved prot
COG1681	Archaeal flagellins
COG1682	ABC-type polysaccharide/polyol

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1683	Uncharacterized conserved prot
COG1684	Flagellar biosynthesis pathway
COG1685	Archaeal shikimate kinase
COG1686	D-alanyl-D-alanine carboxypept
COG1687	Predicted branched-chain amino
COG1690	Uncharacterized conserved prot
COG1691	NCAIR mutase (PurE)-related pr
COG1692	Uncharacterized protein conser
COG1693	Uncharacterized protein conser
COG1696	Predicted membrane protein inv
COG1697	DNA topoisomerase VI, subunit
COG1698	Uncharacterized protein conser
COG1699	Uncharacterized protein conser
COG1701	Uncharacterized protein conser
COG1702	Phosphate starvation-inducible
COG1703	Putative periplasmic protein k
COG1704	Uncharacterized conserved prot
COG1705	Muramidase (flagellum-specific
COG1706	Flagellar basal-body P-ring pr
COG1707	ACT domain-containing protein
COG1708	Predicted nucleotidyltransfera
COG1709	Predicted transcriptional regu
COG1712	Predicted dinucleotide-utilizi

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1713	Predicted HD superfamily hydro
COG1714	Predicted membrane protein/dom
COG1715	Restriction endonuclease
COG1716	FOG: FHA domain
COG1717	Ribosomal protein L32E
COG1718	Serine/threonine protein kinas
COG1720	Uncharacterized conserved prot
COG1722	Exonuclease VII small subunit
COG1723	Uncharacterized conserved prot
COG1724	Predicted periplasmic or secre
COG1726	Na ⁺ -transporting NADH:ubiquino
COG1727	Ribosomal protein L18E
COG1728	Uncharacterized protein conser
COG1729	Uncharacterized protein conser
COG1730	Predicted prefoldin, molecular
COG1731	Archaeal riboflavin synthase
COG1732	Periplasmic glycine betaine/ch
COG1734	DnaK suppressor protein
COG1735	Predicted metal-dependent hydr
COG1736	Diphthamide synthase subunit D
COG1738	Uncharacterized conserved prot
COG1739	Uncharacterized conserved prot
COG1740	Ni,Fe-hydrogenase I small subu

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1742	Uncharacterized conserved prot
COG1743	Adenine-specific DNA methylase
COG1744	Uncharacterized ABC-type trans
COG1745	Predicted metal-binding protei
COG1746	tRNA nucleotidyltransferase (C
COG1749	Flagellar hook protein FlgE
COG1750	Archaeal serine proteases
COG1751	Uncharacterized conserved prot
COG1752	Predicted esterase of the alph
COG1754	Uncharacterized C-terminal dom
COG1755	Uncharacterized protein conser
COG1756	Uncharacterized conserved prot
COG1761	DNA-directed RNA polymerase, s
COG1763	Molybdopterin-guanine dinucleo
COG1766	Flagellar biosynthesis/type II
COG1768	Predicted phosphohydrolase
COG1770	Protease II
COG1772	Uncharacterized protein conser
COG1773	Rubredoxin
COG1774	Uncharacterized homolog of PSP
COG1775	Benzoyl-CoA reductase/2-hydrox
COG1777	Predicted transcriptional regu
COG1778	Low specificity phosphatase (H

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1779	C4-type Zn-finger protein
COG1780	Protein involved in ribonucleo
COG1781	Aspartate carbamoyltransferase
COG1782	Predicted metal-dependent RNAs
COG1784	Predicted membrane protein
COG1785	Alkaline phosphatase
COG1786	Uncharacterized conserved prot
COG1787	Predicted endonuclease distant
COG1791	Uncharacterized conserved prot
COG1792	Cell shape-determining protein
COG1794	Aspartate racemase
COG1795	Uncharacterized conserved prot
COG1796	DNA polymerase IV (family X)
COG1797	Cobyrinic acid a,c-diamide syn
COG1798	Diphthamide biosynthesis methy
COG1799	Uncharacterized protein conser
COG1800	Predicted transglutaminase-lik
COG1802	Transcriptional regulators
COG1803	Methylglyoxal synthase
COG1805	Na ⁺ -transporting NADH:ubiquino
COG1806	Uncharacterized protein conser
COG1807	4-amino-4-deoxy-L-arabinose tr
COG1808	Predicted membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1809	Uncharacterized conserved prot
COG1810	Uncharacterized protein conser
COG1811	Uncharacterized membrane prote
COG1812	Archaeal S-adenosylmethionine
COG1813	Predicted transcription factor
COG1815	Flagellar basal body protein
COG1816	Adenosine deaminase
COG1817	Uncharacterized protein conser
COG1818	Predicted RNA-binding protein,
COG1819	Glycosyl transferases, related
COG1820	N-acetylglucosamine-6-phosphat
COG1821	Predicted ATP-utilizing enzyme
COG1823	Predicted Na ⁺ /dicarboxylate sy
COG1824	Permease, similar to cation tr
COG1825	Ribosomal protein L25 (general
COG1827	Predicted small molecule bindi
COG1828	Phosphoribosylformylglycinamid
COG1829	Predicted archaeal kinase (sug
COG1830	DhnA-type fructose-1,6-bisphos
COG1831	Predicted metal-dependent hydr
COG1832	Predicted CoA-binding protein
COG1833	Uncharacterized conserved prot
COG1834	N-Dimethylarginine dimethylami

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1835	Predicted acyltransferases
COG1836	Predicted membrane protein
COG1837	Predicted RNA-binding protein
COG1838	Tartrate dehydratase beta subu
COG1839	Uncharacterized conserved prot
COG1842	Phage shock protein A (IM30),
COG1843	Flagellar hook capping protein
COG1844	Uncharacterized protein conser
COG1847	Predicted RNA-binding protein
COG1849	Uncharacterized protein conser
COG1850	Ribulose 1,5-bisphosphate carb
COG1852	Uncharacterized conserved prot
COG1854	LuxS protein involved in autoi
COG1855	ATPase (PilT family)
COG1856	Uncharacterized homolog of bio
COG1857	Uncharacterized protein predic
COG1858	Cytochrome c peroxidase
COG1859	RNA:NAD 2'-phosphotransferase
COG1860	Uncharacterized protein conser
COG1861	Spore coat polysaccharide bios
COG1862	Preprotein translocase subunit
COG1863	Multisubunit Na ⁺ /H ⁺ antiporter
COG1864	DNA/RNA endonuclease G, NUC1

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1865	Uncharacterized conserved prot
COG1866	Phosphoenolpyruvate carboxykin
COG1867	N2,N2-dimethylguanosine tRNA m
COG1868	Flagellar motor switch protein
COG1869	ABC-type ribose transport syst
COG1871	Chemotaxis protein; stimulates
COG1872	Uncharacterized conserved prot
COG1873	Uncharacterized conserved prot
COG1874	Beta-galactosidase
COG1875	Predicted ATPase related to ph
COG1876	D-alanyl-D-alanine carboxypept
COG1877	Trehalose-6-phosphatase
COG1878	Predicted metal-dependent hydr
COG1881	Phospholipid-binding protein
COG1882	Pyruvate-formate lyase
COG1883	Na ⁺ -transporting methylmalonyl
COG1884	Methylmalonyl-CoA mutase, N-te
COG1885	Uncharacterized protein conser
COG1886	Flagellar motor switch/type II
COG1887	Putative glycosyl/glycerosp
COG1888	Uncharacterized protein conser
COG1889	Fibrillar-like rRNA methylas
COG1890	Ribosomal protein S3AE

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1891	Uncharacterized protein conser
COG1892	Uncharacterized protein conser
COG1893	Ketopantoate reductase
COG1894	NADH:ubiquinone oxidoreductase
COG1896	Predicted hydrolases of HD sup
COG1897	Homoserine trans-succinylase
COG1899	Deoxyhypusine synthase
COG1900	Uncharacterized conserved prot
COG1901	Uncharacterized conserved prot
COG1903	Cobalamin biosynthesis protein
COG1904	Glucuronate isomerase
COG1905	NADH:ubiquinone oxidoreductase
COG1906	Uncharacterized conserved prot
COG1907	Predicted archaeal sugar kinas
COG1908	Coenzyme F420-reducing hydroge
COG1909	Uncharacterized protein conser
COG1910	Periplasmic molybdate-binding
COG1911	Ribosomal protein L30E
COG1912	Uncharacterized conserved prot
COG1913	Predicted Zn-dependent proteas
COG1914	Mn ²⁺ and Fe ²⁺ transporters of
COG1915	Uncharacterized conserved prot
COG1916	Uncharacterized homolog of Prg

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1918	Fe ²⁺ transport system protein
COG1920	Uncharacterized conserved prot
COG1921	Selenocysteine synthase [seryl
COG1922	Teichoic acid biosynthesis pro
COG1923	Uncharacterized host factor I
COG1924	Activator of 2-hydroxyglutaryl
COG1925	Phosphotransferase system, HPr
COG1926	Predicted phosphoribosyltransf
COG1927	Coenzyme F420-dependent N(5),N
COG1928	Dolichyl-phosphate-mannose-pr
COG1929	Glycerate kinase
COG1930	ABC-type cobalt transport syst
COG1931	Uncharacterized protein conser
COG1933	Archaeal DNA polymerase II, la
COG1934	Uncharacterized protein conser
COG1935	Uncharacterized conserved prot
COG1936	Predicted nucleotide kinase (r
COG1937	Uncharacterized protein conser
COG1938	Archaeal enzymes of ATP-grasp
COG1939	Uncharacterized protein conser
COG1941	Coenzyme F420-reducing hydroge
COG1942	Uncharacterized protein, 4-oxa
COG1943	Transposase and inactivated de

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1944	Uncharacterized conserved prot
COG1945	Uncharacterized conserved prot
COG1946	Acyl-CoA thioesterase
COG1947	4-diphosphocytidyl-2C-methyl-D
COG1948	ERCC4-type nuclease
COG1949	Oligoribonuclease (3'-5' exor
COG1950	Predicted membrane protein
COG1951	Tartrate dehydratase alpha sub
COG1952	Preprotein translocase subunit
COG1953	Cytosine/uracil/thiamine/allan
COG1954	Glycerol-3-phosphate responsiv
COG1955	Archaeal flagella assembly pro
COG1956	GAF domain-containing protein
COG1957	Inosine-uridine nucleoside N-r
COG1958	Small nuclear ribonucleoprotei
COG1962	Tetrahydromethanopterin S-meth
COG1963	Uncharacterized protein conser
COG1964	Predicted Fe-S oxidoreductases
COG1965	Protein implicated in iron tra
COG1966	Carbon starvation protein, pre
COG1967	Predicted membrane protein
COG1968	Uncharacterized bacitracin res
COG1970	Large-conductance mechanosensi

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1971	Predicted membrane protein
COG1972	Nucleoside permease
COG1973	Hydrogenase maturation factor
COG1974	SOS-response transcriptional r
COG1976	Translation initiation factor
COG1977	Molybdopterin converting facto
COG1979	Uncharacterized oxidoreductase
COG1980	Archaeal fructose 1,6-bisphosp
COG1981	Predicted membrane protein
COG1982	Arginine/lysine/ornithine deca
COG1983	Putative stress-responsive tra
COG1984	Allophanate hydrolase subunit
COG1985	Pyrimidine reductase, riboflav
COG1986	Uncharacterized conserved prot
COG1987	Flagellar biosynthesis pathway
COG1988	Predicted membrane-bound metal
COG1989	Type II secretory pathway, pre
COG1990	Uncharacterized conserved prot
COG1992	Uncharacterized conserved prot
COG1993	Uncharacterized conserved prot
COG1995	Pyridoxal phosphate biosynthes
COG1997	Ribosomal protein L37AE/L43A
COG1998	Ribosomal protein S27AE

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG1999	Uncharacterized protein SCO1/S
COG2001	Uncharacterized protein conser
COG2002	Regulators of stationary/sporu
COG2003	DNA repair proteins
COG2004	Ribosomal protein S24E
COG2005	N-terminal domain of molybdenu
COG2006	Uncharacterized conserved prot
COG2007	Ribosomal protein S8E
COG2008	Threonine aldolase
COG2009	Succinate dehydrogenase/fumara
COG2010	Cytochrome c, mono- and diheme
COG2011	ABC-type metal ion transport s
COG2012	DNA-directed RNA polymerase, s
COG2013	Uncharacterized conserved prot
COG2014	Uncharacterized conserved prot
COG2015	Alkyl sulfatase and related hy
COG2016	Predicted RNA-binding protein
COG2019	Archaeal adenylate kinase
COG2020	Putative protein-S-isoprenylcy
COG2021	Homoserine acetyltransferase
COG2022	Uncharacterized enzyme of thia
COG2023	RNase P subunit RPR2
COG2024	Phenylalanyl-tRNA synthetase a

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2026	Cytotoxic translational repres
COG2027	D-alanyl-D-alanine carboxypept
COG2028	Uncharacterized conserved prot
COG2029	Uncharacterized conserved prot
COG2031	Short chain fatty acids transp
COG2032	Cu/Zn superoxide dismutase
COG2033	Desulfoferrodoxin
COG2034	Predicted membrane protein
COG2035	Predicted membrane protein
COG2036	Histones H3 and H4
COG2037	Formylmethanofuran:tetrahydrom
COG2039	Pyrrolidone-carboxylate peptid
COG2040	Homocysteine/selenocysteine me
COG2042	Uncharacterized conserved prot
COG2043	Uncharacterized protein conser
COG2044	Predicted peroxiredoxins
COG2045	Phosphosulfolactate phosphohyd
COG2046	ATP sulfurylase (sulfate adeny
COG2047	Uncharacterized protein (ATP-g
COG2048	Heterodisulfide reductase, sub
COG2049	Allophanate hydrolase subunit
COG2051	Ribosomal protein S27E
COG2052	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2053	Ribosomal protein S28E/S33
COG2054	Uncharacterized archaeal kinas
COG2055	Malate/L-lactate dehydrogenase
COG2056	Predicted permease
COG2058	Ribosomal protein L12E/L44/L45
COG2059	Chromate transport protein Chr
COG2060	K ⁺ -transporting ATPase, A chai
COG2061	ACT-domain-containing protein,
COG2063	Flagellar basal body L-ring pr
COG2064	Flp pilus assembly protein Tad
COG2065	Pyrimidine operon attenuation
COG2066	Glutaminase
COG2067	Long-chain fatty acid transpor
COG2068	Uncharacterized MobA-related p
COG2069	CO dehydrogenase/acetyl-CoA sy
COG2071	Predicted glutamine amidotrans
COG2073	Cobalamin biosynthesis protein
COG2074	2-phosphoglycerate kinase
COG2075	Ribosomal protein L24E
COG2077	Peroxiredoxin
COG2078	Uncharacterized conserved prot
COG2079	Uncharacterized protein involv
COG2080	Aerobic-type carbon monoxide d

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2081	Predicted flavoproteins
COG2082	Precorrin isomerase
COG2083	Uncharacterized protein conser
COG2086	Electron transfer flavoprotein
COG2087	Adenosyl cobinamide kinase/ade
COG2088	Uncharacterized protein, invol
COG2089	Sialic acid synthase
COG2090	Uncharacterized protein conser
COG2091	Phosphopantetheinyl transferas
COG2092	Translation elongation factor
COG2093	DNA-directed RNA polymerase, s
COG2094	3-methyladenine DNA glycosylas
COG2095	Multiple antibiotic transporte
COG2096	Uncharacterized conserved prot
COG2097	Ribosomal protein L31E
COG2098	Uncharacterized protein conser
COG2099	Precorrin-6x reductase
COG2100	Predicted Fe-S oxidoreductase
COG2101	TATA-box binding protein (TBP)
COG2102	Predicted ATPases of PP-loop s
COG2103	Predicted sugar phosphate isom
COG2105	Uncharacterized conserved prot
COG2106	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2107	Predicted periplasmic solute-b
COG2108	Uncharacterized conserved prot
COG2109	ATP:corrinoïd adenosyltransfer
COG2110	Predicted phosphatase homologo
COG2112	Predicted Ser/Thr protein kina
COG2113	ABC-type proline/glycine betai
COG2114	Adenylate cyclase, family 3 (s
COG2115	Xylose isomerase
COG2116	Formate/nitrite family of tran
COG2117	Predicted subunit of tRNA(5-me
COG2118	DNA-binding protein
COG2119	Predicted membrane protein
COG2120	Uncharacterized proteins, LmbE
COG2121	Uncharacterized protein conser
COG2122	Uncharacterized conserved prot
COG2123	RNase PH-related exoribonuclea
COG2125	Ribosomal protein S6E (S10)
COG2126	Ribosomal protein L37E
COG2127	Uncharacterized conserved prot
COG2129	Predicted phosphoesterases, re
COG2130	Putative NADP-dependent oxidor
COG2131	Deoxycytidylate deaminase
COG2132	Putative multicopper oxidases

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2134	CDP-diacylglycerol pyrophospha
COG2135	Uncharacterized conserved prot
COG2136	Predicted exosome subunit/U3 s
COG2137	Uncharacterized protein conser
COG2138	Uncharacterized conserved prot
COG2139	Ribosomal protein L21E
COG2142	Succinate dehydrogenase, hydro
COG2144	Selenophosphate synthetase-rel
COG2145	Hydroxyethylthiazole kinase, s
COG2147	Ribosomal protein L19E
COG2148	Sugar transferases involved in
COG2149	Predicted membrane protein
COG2150	Predicted regulator of amino a
COG2151	Predicted metal-sulfur cluster
COG2152	Predicted glycosylase
COG2153	Predicted acyltransferase
COG2154	Pterin-4a-carbinolamine dehydr
COG2155	Uncharacterized conserved prot
COG2156	K ⁺ -transporting ATPase, c chai
COG2157	Ribosomal protein L20A (L18A)
COG2158	Uncharacterized protein contai
COG2159	Predicted metal-dependent hydr
COG2160	L-arabinose isomerase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2162	Arylamine N-acetyltransferase
COG2163	Ribosomal protein L14E/L6E/L27
COG2165	Type II secretory pathway, pse
COG2166	SufE protein probably involved
COG2167	Ribosomal protein L39E
COG2169	Adenosine deaminase
COG2170	Uncharacterized conserved prot
COG2172	Anti-sigma regulatory factor (
COG2173	D-alanyl-D-alanine dipeptidase
COG2174	Ribosomal protein L34E
COG2175	Probable taurine catabolism di
COG2176	DNA polymerase III, alpha subu
COG2177	Cell division protein
COG2178	Predicted RNA-binding protein
COG2179	Predicted hydrolase of the HAD
COG2180	Nitrate reductase delta subuni
COG2181	Nitrate reductase gamma subuni
COG2182	Maltose-binding periplasmic pr
COG2183	Transcriptional accessory prot
COG2184	Protein involved in cell divis
COG2185	Methylmalonyl-CoA mutase, C-te
COG2187	Uncharacterized protein conser
COG2189	Adenine specific DNA methylase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2191	Formylmethanofuran dehidrogena
COG2192	Predicted carbamoyl transferas
COG2193	Bacterioferritin (cytochrome b
COG2194	Predicted membrane-associated,
COG2198	FOG: HPt domain
COG2201	Chemotaxis response regulator
COG2203	FOG: GAF domain
COG2205	Osmosensitive K ⁺ channel histi
COG2206	HD-GYP domain
COG2208	Serine phosphatase RsbU, regul
COG2209	Na ⁺ -transporting NADH:ubiquino
COG2210	Uncharacterized conserved prot
COG2211	Na ⁺ /melibiose symporter and re
COG2212	Multisubunit Na ⁺ /H ⁺ antiporter
COG2213	Phosphotransferase system, man
COG2214	DnaJ-class molecular chaperone
COG2215	ABC-type uncharacterized trans
COG2216	High-affinity K ⁺ transport sys
COG2218	Formylmethanofuran dehidrogena
COG2219	Eukaryotic-type DNA primase, l
COG2221	Dissimilatory sulfite reductas
COG2224	Isocitrate lyase
COG2226	Methylase involved in ubiquino

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2227	2-polyprenyl-3-methyl-5-hydrox
COG2230	Cyclopropane fatty acid syntha
COG2231	Uncharacterized protein relate
COG2232	Predicted ATP-dependent carbol
COG2234	Predicted aminopeptidases
COG2235	Arginine deiminase
COG2237	Predicted membrane protein
COG2238	Ribosomal protein S19E (S16A)
COG2239	Mg/Co/Ni transporter MgtE (con
COG2240	Pyridoxal/pyridoxine/pyridoxam
COG2241	Precorrin-6B methylase 1
COG2242	Precorrin-6B methylase 2
COG2243	Precorrin-2 methylase
COG2245	Predicted membrane protein
COG2246	Predicted membrane protein
COG2248	Predicted hydrolase (metallo-b
COG2250	Uncharacterized conserved prot
COG2251	Predicted nuclease (RecB famil
COG2252	Permeases
COG2255	Holliday junction resolasome,
COG2256	ATPase related to the helicase
COG2257	Uncharacterized homolog of the
COG2258	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2260	Predicted Zn-ribbon RNA-bindin
COG2261	Predicted membrane protein
COG2262	GTPases
COG2263	Predicted RNA methylase
COG2264	Ribosomal protein L11 methylas
COG2266	GTP:adenosylcobinamide-phospa
COG2268	Uncharacterized protein conser
COG2269	Truncated, possibly inactive,
COG2270	Permeases of the major facilit
COG2271	Sugar phosphate permease
COG2272	Carboxylesterase type B
COG2273	Beta-glucanase/Beta-glucan syn
COG2274	ABC-type bacteriocin/lantibiot
COG2301	Citrate lyase beta subunit
COG2302	Uncharacterized conserved prot
COG2304	Uncharacterized protein contai
COG2307	Uncharacterized protein conser
COG2308	Uncharacterized conserved prot
COG2310	Uncharacterized proteins invol
COG2311	Predicted membrane protein
COG2312	Erythromycin esterase homolog
COG2313	Uncharacterized enzyme involve
COG2315	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2317	Zn-dependent carboxypeptidase
COG2318	Uncharacterized protein conser
COG2319	FOG: WD40 repeat
COG2320	Uncharacterized conserved prot
COG2321	Predicted metalloprotease
COG2322	Predicted membrane protein
COG2323	Predicted membrane protein
COG2324	Predicted membrane protein
COG2326	Uncharacterized conserved prot
COG2327	Uncharacterized conserved prot
COG2329	Uncharacterized enzyme involve
COG2332	Cytochrome c-type biogenesis p
COG2333	Predicted hydrolase (metallo-b
COG2334	Putative homoserine kinase typ
COG2335	Secreted and surface protein c
COG2336	Growth regulator
COG2337	Growth inhibitor
COG2339	Predicted membrane protein
COG2340	Uncharacterized protein with S
COG2342	Predicted extracellular endo a
COG2343	Uncharacterized protein conser
COG2344	AT-rich DNA-binding protein
COG2345	Predicted transcriptional regu

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2346	Truncated hemoglobins
COG2350	Uncharacterized protein conser
COG2351	Transthyretin-like protein
COG2352	Phosphoenolpyruvate carboxylas
COG2353	Uncharacterized conserved prot
COG2354	Uncharacterized protein conser
COG2355	Zn-dependent dipeptidase, micr
COG2356	Endonuclease I
COG2358	TRAP-type uncharacterized tran
COG2359	Uncharacterized protein conser
COG2360	Leu/Phe-tRNA-protein transfera
COG2361	Uncharacterized conserved prot
COG2362	D-aminopeptidase
COG2363	Uncharacterized small membrane
COG2364	Predicted membrane protein
COG2365	Protein tyrosine/serine phosph
COG2366	Protein related to penicillin
COG2367	Beta-lactamase class A
COG2368	Aromatic ring hydroxylase
COG2369	Uncharacterized protein, homol
COG2370	Hydrogenase/urease accessory p
COG2371	Urease accessory protein UreE
COG2372	Uncharacterized protein, homol

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2373	Large extracellular alpha-heli
COG2374	Predicted extracellular nuclea
COG2375	Siderophore-interacting protei
COG2377	Predicted molecular chaperone
COG2378	Predicted transcriptional regu
COG2379	Putative glycerate kinase
COG2382	Enterochelin esterase and rela
COG2383	Uncharacterized conserved prot
COG2384	Predicted SAM-dependent methyl
COG2385	Sporulation protein and relate
COG2386	ABC-type transport system invo
COG2388	Predicted acetyltransferase
COG2401	ABC-type ATPase fused to a pre
COG2403	Predicted GTPase
COG2405	Predicted nucleic acid-binding
COG2406	Protein distantly related to b
COG2407	L-fucose isomerase and related
COG2409	Predicted drug exporters of th
COG2411	Uncharacterized conserved prot
COG2412	Uncharacterized conserved prot
COG2413	Predicted nucleotidyltransfera
COG2414	Aldehyde:ferredoxin oxidoreduc
COG2419	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2421	Predicted acetamidase/formamid
COG2423	Predicted ornithine cyclodeami
COG2425	Uncharacterized protein contai
COG2426	Predicted membrane protein
COG2428	Uncharacterized conserved prot
COG2429	Uncharacterized conserved prot
COG2430	Uncharacterized conserved prot
COG2431	Predicted membrane protein
COG2433	Uncharacterized conserved prot
COG2440	Ferredoxin-like protein
COG2441	Predicted butyrate kinase
COG2450	Uncharacterized conserved prot
COG2451	Ribosomal protein L35AE/L33A
COG2453	Predicted protein-tyrosine pho
COG2454	Uncharacterized conserved prot
COG2461	Uncharacterized conserved prot
COG2469	Uncharacterized conserved prot
COG2501	Uncharacterized conserved prot
COG2502	Asparagine synthetase A
COG2508	Regulator of polyketide syntha
COG2509	Uncharacterized FAD-dependent
COG2510	Predicted membrane protein
COG2511	Archaeal Glu-tRNAGln amidotran

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2514	Predicted ring-cleavage extrad
COG2515	1-aminocyclopropane-1-carboxyl
COG2516	Biotin synthase-related enzyme
COG2517	Predicted RNA-binding protein
COG2519	tRNA(1-methyladenosine) methyl
COG2521	Predicted archaeal methyltrans
COG2524	Predicted transcriptional regu
COG2602	Beta-lactamase class D
COG2603	Predicted ATPase
COG2605	Predicted kinase related to ga
COG2607	Predicted ATPase (AAA+ superfa
COG2608	Copper chaperone
COG2609	Pyruvate dehydrogenase complex
COG2610	H ⁺ /gluconate symporter and rel
COG2703	Hemerythrin
COG2704	Anaerobic C4-dicarboxylate tra
COG2706	3-carboxymuconate cyclase
COG2707	Predicted membrane protein
COG2715	Uncharacterized membrane prote
COG2716	Glycine cleavage system regula
COG2717	Predicted membrane protein
COG2718	Uncharacterized conserved prot
COG2719	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2720	Uncharacterized vancomycin res
COG2721	Altronate dehydratase
COG2730	Endoglucanase
COG2731	Beta-galactosidase, beta subun
COG2733	Predicted membrane protein
COG2738	Predicted Zn-dependent proteas
COG2739	Uncharacterized protein conser
COG2740	Predicted nucleic-acid-binding
COG2746	Aminoglycoside N3'-acetyltrans
COG2755	Lysophospholipase L1 and relat
COG2759	Formyltetrahydrofolate synthet
COG2761	Predicted dithiol-disulfide is
COG2766	Putative Ser protein kinase
COG2768	Uncharacterized Fe-S center pr
COG2771	DNA-binding HTH domain-contain
COG2802	Uncharacterized protein, simil
COG2804	Type II secretory pathway, ATP
COG2805	Tfp pilus assembly protein, pi
COG2807	Cyanate permease
COG2808	Transcriptional regulator
COG2811	Archaeal/vacuolar-type H ⁺ -ATPa
COG2812	DNA polymerase III, gamma/tau
COG2813	16S RNA G1207 methylase RsmC

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2815	Uncharacterized protein conser
COG2816	NTP pyrophosphohydrolases cont
COG2818	3-methyladenine DNA glycosylas
COG2819	Predicted hydrolase of the alp
COG2820	Uridine phosphorylase
COG2821	Membrane-bound lytic murein tr
COG2822	Predicted periplasmic lipoprot
COG2823	Predicted periplasmic or secre
COG2824	Uncharacterized Zn-ribbon-cont
COG2825	Outer membrane protein
COG2826	Transposase and inactivated de
COG2827	Predicted endonuclease contain
COG2828	Uncharacterized protein conser
COG2831	Hemolysin activation/secretion
COG2832	Uncharacterized protein conser
COG2834	Outer membrane lipoprotein-sor
COG2835	Uncharacterized conserved prot
COG2836	Uncharacterized conserved prot
COG2837	Predicted iron-dependent perox
COG2838	Monomeric isocitrate dehydroge
COG2839	Uncharacterized protein conser
COG2840	Uncharacterized protein conser
COG2843	Putative enzyme of poly-gamma-

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2844	UTP:GlnB (protein PII) uridyly
COG2846	Regulator of cell morphogenesi
COG2847	Uncharacterized protein conser
COG2848	Uncharacterized conserved prot
COG2850	Uncharacterized conserved prot
COG2851	H ⁺ /citrate symporter
COG2852	Uncharacterized protein conser
COG2853	Surface lipoprotein
COG2854	ABC-type transport system invo
COG2855	Predicted membrane protein
COG2856	Predicted Zn peptidase
COG2857	Cytochrome c1
COG2859	Uncharacterized protein conser
COG2860	Predicted membrane protein
COG2861	Uncharacterized protein conser
COG2862	Predicted membrane protein
COG2863	Cytochrome c553
COG2864	Cytochrome b subunit of format
COG2865	Predicted transcriptional regu
COG2866	Predicted carboxypeptidase
COG2867	Oligoketide cyclase/lipid tran
COG2868	Predicted ribosomal protein
COG2869	Na ⁺ -transporting NADH:ubiquino

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2870	ADP-heptose synthase, bifuncti
COG2871	Na ⁺ -transporting NADH:ubiquino
COG2873	O-acetylhomoserine sulfhydryla
COG2874	Predicted ATPases involved in
COG2875	Precorrin-4 methylase
COG2876	3-deoxy-D-arabino-heptulosonat
COG2877	3-deoxy-D-manno-octulosonic ac
COG2878	Predicted NADH:ubiquinone oxid
COG2879	Uncharacterized small protein
COG2882	Flagellar biosynthesis chapero
COG2884	Predicted ATPase involved in c
COG2887	RecB family exonuclease
COG2888	Predicted Zn-ribbon RNA-bindin
COG2891	Cell shape-determining protein
COG2893	Phosphotransferase system, man
COG2894	Septum formation inhibitor-act
COG2895	GTPases - Sulfate adenylate tr
COG2896	Molybdenum cofactor biosynthes
COG2897	Rhodanese-related sulfurtransf
COG2898	Uncharacterized conserved prot
COG2899	Uncharacterized protein conser
COG2901	Factor for inversion stimulati
COG2902	NAD-specific glutamate dehydro

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2904	Uncharacterized protein conser
COG2905	Predicted signal-transduction
COG2907	Predicted NAD/FAD-binding prot
COG2909	ATP-dependent transcriptional
COG2910	Putative NADH-flavin reductase
COG2911	Uncharacterized protein conser
COG2912	Uncharacterized conserved prot
COG2913	Small protein A (tmRNA-binding
COG2914	Uncharacterized protein conser
COG2915	Uncharacterized protein involv
COG2917	Intracellular septation protei
COG2918	Gamma-glutamylcysteine synthet
COG2919	Septum formation initiator
COG2920	Dissimilatory sulfite reductas
COG2921	Uncharacterized conserved prot
COG2922	Uncharacterized protein conser
COG2923	Uncharacterized protein involv
COG2924	Fe-S cluster protector protein
COG2925	Exonuclease I
COG2927	DNA polymerase III, chi subuni
COG2928	Uncharacterized conserved prot
COG2929	Uncharacterized protein conser
COG2930	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2932	Predicted transcriptional regu
COG2933	Predicted SAM-dependent methyl
COG2935	Putative arginyl-tRNA:protein
COG2936	Predicted acyl esterases
COG2937	Glycerol-3-phosphate O-acyltra
COG2938	Uncharacterized conserved prot
COG2939	Carboxypeptidase C (cathepsin
COG2940	Proteins containing SET domain
COG2941	Ubiquinone biosynthesis protei
COG2942	N-acyl-D-glucosamine 2-epimera
COG2943	Membrane glycosyltransferase
COG2944	Predicted transcriptional regu
COG2945	Predicted hydrolase of the alp
COG2947	Uncharacterized conserved prot
COG2948	Type IV secretory pathway, Vir
COG2949	Uncharacterized membrane prote
COG2951	Membrane-bound lytic murein tr
COG2954	Uncharacterized protein conser
COG2956	Predicted N-acetylglucosaminyl
COG2957	Peptidylarginine deiminase and
COG2958	Uncharacterized protein conser
COG2959	Uncharacterized enzyme of heme
COG2960	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2961	Protein involved in catabolism
COG2962	Predicted permeases
COG2963	Transposase and inactivated de
COG2964	Uncharacterized protein conser
COG2966	Uncharacterized conserved prot
COG2967	Uncharacterized protein affect
COG2968	Uncharacterized conserved prot
COG2969	Stringent starvation protein B
COG2971	Predicted N-acetylglucosamine
COG2972	Predicted signal transduction
COG2974	DNA recombination-dependent gr
COG2975	Uncharacterized protein conser
COG2976	Uncharacterized protein conser
COG2977	Phosphopantetheinyl transferas
COG2978	Putative p-aminobenzoyl-glutam
COG2979	Uncharacterized protein conser
COG2981	Uncharacterized protein involv
COG2982	Uncharacterized protein involv
COG2983	Uncharacterized conserved prot
COG2984	ABC-type uncharacterized trans
COG2985	Predicted permease
COG2986	Histidine ammonia-lyase
COG2987	Urocanate hydratase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG2988	Succinylglutamate desuccinylas
COG2989	Uncharacterized protein conser
COG2990	Uncharacterized protein conser
COG2993	Cbb3-type cytochrome oxidase,
COG2995	Uncharacterized paraquat-induc
COG2996	Uncharacterized protein conser
COG2998	ABC-type tungstate transport s
COG3001	Fructosamine-3-kinase
COG3004	Na ⁺ /H ⁺ antiporter
COG3005	Nitrate/TMAO reductases, membr
COG3007	Uncharacterized paraquat-induc
COG3008	Paraquat-inducible protein B
COG3010	Putative N-acetylmannosamine-6
COG3011	Uncharacterized protein conser
COG3012	Uncharacterized protein conser
COG3013	Uncharacterized conserved prot
COG3017	Outer membrane lipoprotein inv
COG3019	Predicted metal-binding protei
COG3021	Uncharacterized protein conser
COG3022	Uncharacterized protein conser
COG3023	Negative regulator of beta-lac
COG3024	Uncharacterized protein conser
COG3025	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3026	Negative regulator of sigma E
COG3027	Uncharacterized protein conser
COG3028	Uncharacterized protein conser
COG3029	Fumarate reductase subunit C
COG3030	Protein affecting phage T7 exc
COG3033	Tryptophanase
COG3034	Uncharacterized protein conser
COG3037	Uncharacterized protein conser
COG3038	Cytochrome B561
COG3039	Transposase and inactivated de
COG3040	Bacterial lipocalin
COG3041	Uncharacterized protein conser
COG3043	Nitrate reductase cytochrome c
COG3044	Predicted ATPase of the ABC cl
COG3045	Uncharacterized protein conser
COG3046	Uncharacterized protein relate
COG3047	Outer membrane protein W
COG3048	D-serine dehydratase
COG3049	Penicillin V acylase and relat
COG3051	Citrate lyase, alpha subunit
COG3052	Citrate lyase, gamma subunit
COG3053	Citrate lyase synthetase
COG3058	Uncharacterized protein involv

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3063	Tfp pilus assembly protein Pil
COG3065	Starvation-inducible outer mem
COG3067	Na ⁺ /H ⁺ antiporter
COG3070	Regulator of competence-specif
COG3071	Uncharacterized enzyme of heme
COG3072	Adenylate cyclase
COG3075	Anaerobic glycerol-3-phosphate
COG3077	DNA-damage-inducible protein J
COG3080	Fumarate reductase subunit D
COG3081	Nucleoid-associated protein
COG3088	Uncharacterized protein involv
COG3090	TRAP-type C4-dicarboxylate tra
COG3091	Uncharacterized protein conser
COG3093	Plasmid maintenance system ant
COG3098	Uncharacterized protein conser
COG3103	SH3 domain protein
COG3104	Dipeptide/tripeptide permease
COG3106	Predicted ATPase
COG3107	Putative lipoprotein
COG3108	Uncharacterized protein conser
COG3115	Cell division protein
COG3118	Thioredoxin domain-containing
COG3121	P pilus assembly protein, chap

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3123	Uncharacterized protein conser
COG3124	Uncharacterized protein conser
COG3125	Heme/copper-type cytochrome/qu
COG3126	Uncharacterized protein conser
COG3127	Predicted ABC-type transport s
COG3128	Uncharacterized iron-regulated
COG3129	Predicted SAM-dependent methyl
COG3131	Periplasmic glucans biosynthes
COG3132	Uncharacterized protein conser
COG3133	Outer membrane lipoprotein
COG3134	Predicted outer membrane lipop
COG3135	Uncharacterized protein involv
COG3136	Uncharacterized membrane prote
COG3138	Arginine/ornithine N-succinylt
COG3142	Uncharacterized protein involv
COG3143	Chemotaxis protein
COG3144	Flagellar hook-length control
COG3146	Uncharacterized protein conser
COG3147	Uncharacterized protein conser
COG3150	Predicted esterase
COG3151	Uncharacterized protein conser
COG3152	Predicted membrane protein
COG3153	Predicted acetyltransferase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3154	Putative lipid carrier protein
COG3156	Type II secretory pathway, com
COG3157	Hemolysin-coregulated protein
COG3158	K ⁺ transporter
COG3159	Uncharacterized protein conser
COG3160	Regulator of sigma D
COG3161	4-hydroxybenzoate synthetase (
COG3164	Predicted membrane protein
COG3166	Tfp pilus assembly protein Pil
COG3167	Tfp pilus assembly protein Pil
COG3168	Tfp pilus assembly protein Pil
COG3169	Uncharacterized protein conser
COG3170	Tfp pilus assembly protein Fim
COG3174	Predicted membrane protein
COG3175	Cytochrome oxidase assembly fa
COG3176	Putative hemolysin
COG3178	Predicted phosphotransferase r
COG3179	Predicted chitinase
COG3180	Putative ammonia monooxygenase
COG3181	Uncharacterized protein conser
COG3182	Uncharacterized iron-regulated
COG3183	Predicted restriction endonucl
COG3185	4-hydroxyphenylpyruvate dioxyg

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3186	Phenylalanine-4-hydroxylase
COG3188	P pilus assembly protein, pori
COG3189	Uncharacterized conserved prot
COG3191	L-aminopeptidase/D-esterase
COG3192	Ethanolamine utilization prote
COG3193	Uncharacterized protein, possi
COG3194	Ureidoglycolate hydrolase
COG3195	Uncharacterized protein conser
COG3199	Uncharacterized conserved prot
COG3200	3-deoxy-D-arabino-heptulosonate
COG3201	Nicotinamide mononucleotide tr
COG3204	Uncharacterized protein conser
COG3205	Predicted membrane protein
COG3206	Uncharacterized protein involv
COG3207	Pyoverdine/dityrosine biosynth
COG3208	Predicted thioesterase involve
COG3209	Rhs family protein
COG3210	Large exoproteins involved in
COG3211	Predicted phosphatase
COG3212	Predicted membrane protein
COG3213	Uncharacterized protein involv
COG3214	Uncharacterized protein conser
COG3215	Tfp pilus assembly protein Pil

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3216	Uncharacterized protein conser
COG3217	Uncharacterized Fe-S protein
COG3218	ABC-type uncharacterized trans
COG3220	Uncharacterized protein conser
COG3221	ABC-type phosphate/phosphonate
COG3223	Predicted membrane protein
COG3224	Uncharacterized protein conser
COG3226	Uncharacterized protein conser
COG3227	Zinc metalloprotease (elastase
COG3228	Uncharacterized protein conser
COG3231	Aminoglycoside phosphotransfer
COG3232	5-carboxymethyl-2-hydroxyucon
COG3234	Uncharacterized protein conser
COG3237	Uncharacterized protein conser
COG3238	Uncharacterized protein conser
COG3239	Fatty acid desaturase
COG3240	Phospholipase/lecithinase/hemo
COG3243	Poly(3-hydroxyalkanoate) synth
COG3246	Uncharacterized conserved prot
COG3247	Uncharacterized conserved prot
COG3249	Uncharacterized protein conser
COG3250	Beta-galactosidase/beta-glucur
COG3252	Methenyltetrahydromethanopteri

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3253	Uncharacterized conserved prot
COG3254	Uncharacterized conserved prot
COG3255	Putative sterol carrier protei
COG3256	Nitric oxide reductase large s
COG3257	Uncharacterized protein, possi
COG3258	Cytochrome c
COG3259	Coenzyme F420-reducing hydroge
COG3260	Ni,Fe-hydrogenase III small su
COG3261	Ni,Fe-hydrogenase III large su
COG3263	NhaP-type Na ⁺ /H ⁺ and K ⁺ /H ⁺ ant
COG3264	Small-conductance mechanosensi
COG3265	Gluconate kinase
COG3268	Uncharacterized conserved prot
COG3269	Predicted RNA-binding protein,
COG3272	Uncharacterized conserved prot
COG3273	Uncharacterized conserved prot
COG3275	Putative regulator of cell aut
COG3276	Selenocysteine-specific transl
COG3277	RNA-binding protein involved i
COG3278	Cbb3-type cytochrome oxidase,
COG3279	Response regulator of the LytR
COG3280	Maltooligosyl trehalose syntha
COG3281	Uncharacterized protein, proba

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3283	Transcriptional regulator of a
COG3284	Transcriptional activator of a
COG3285	Predicted eukaryotic-type DNA
COG3286	Uncharacterized protein conser
COG3287	Uncharacterized conserved prot
COG3290	Signal transduction histidine
COG3291	FOG: PKD repeat
COG3292	Predicted periplasmic ligand-b
COG3293	Transposase and inactivated de
COG3294	Uncharacterized conserved prot
COG3300	MHYT domain (predicted integra
COG3301	Formate-dependent nitrite redu
COG3304	Predicted membrane protein
COG3307	Lipid A core - O-antigen ligas
COG3310	Uncharacterized protein conser
COG3312	F0F1-type ATP synthase, subuni
COG3313	Predicted Fe-S protein
COG3314	Uncharacterized protein conser
COG3315	O-Methyltransferase involved i
COG3316	Transposase and inactivated de
COG3319	Thioesterase domains of type I
COG3320	Putative dehydrogenase domain
COG3321	Polyketide synthase modules an

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3322	Predicted periplasmic ligand-b
COG3323	Uncharacterized protein conser
COG3324	Predicted enzyme related to la
COG3325	Chitinase
COG3326	Predicted membrane protein
COG3327	Phenylacetic acid-responsive t
COG3328	Transposase and inactivated de
COG3329	Predicted permease
COG3330	Uncharacterized protein conser
COG3331	Penicillin-binding protein-rel
COG3332	Uncharacterized conserved prot
COG3333	Uncharacterized protein conser
COG3335	Transposase and inactivated de
COG3336	Predicted membrane protein
COG3338	Carbonic anhydrase
COG3339	Uncharacterized conserved prot
COG3340	Peptidase E
COG3341	Predicted double-stranded RNA/
COG3342	Uncharacterized conserved prot
COG3343	DNA-directed RNA polymerase, d
COG3344	Retron-type reverse transcript
COG3345	Alpha-galactosidase
COG3346	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3347	Uncharacterized conserved prot
COG3351	Putative archaeal flagellar pr
COG3354	Putative archaeal flagellar pr
COG3355	Predicted transcriptional regu
COG3356	Predicted membrane protein
COG3357	Predicted transcriptional regu
COG3358	Uncharacterized conserved prot
COG3359	Predicted exonuclease
COG3360	Uncharacterized conserved prot
COG3363	Archaeal IMP cyclohydrolase
COG3365	Uncharacterized protein conser
COG3367	Uncharacterized conserved prot
COG3371	Predicted membrane protein
COG3372	Uncharacterized conserved prot
COG3373	Uncharacterized protein conser
COG3374	Predicted membrane protein
COG3376	High-affinity nickel permease
COG3377	Uncharacterized conserved prot
COG3378	Predicted ATPase
COG3380	Predicted NAD/FAD-dependent ox
COG3382	Uncharacterized conserved prot
COG3383	Uncharacterized anaerobic dehy
COG3384	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3385	FOG: Transposase and inactivat
COG3388	Uncharacterized protein conser
COG3389	Uncharacterized protein conser
COG3390	Uncharacterized protein conser
COG3391	Uncharacterized conserved prot
COG3394	Uncharacterized protein conser
COG3395	Uncharacterized protein conser
COG3396	Uncharacterized conserved prot
COG3397	Uncharacterized protein conser
COG3402	Uncharacterized conserved prot
COG3404	Methenyl tetrahydrofolate cycl
COG3407	Mevalonate pyrophosphate decar
COG3412	Uncharacterized protein conser
COG3413	Predicted DNA binding protein
COG3414	Phosphotransferase system, gal
COG3416	Uncharacterized protein conser
COG3420	Nitrous oxidase accessory prot
COG3424	Predicted naringenin-chalcone
COG3425	3-hydroxy-3-methylglutaryl CoA
COG3426	Butyrate kinase
COG3427	Uncharacterized conserved prot
COG3428	Predicted membrane protein
COG3429	Glucose-6-P dehydrogenase subu

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3435	Gentisate 1,2-dioxygenase
COG3436	Transposase and inactivated de
COG3437	Response regulator containing
COG3442	Predicted glutamine amidotrans
COG3443	Predicted periplasmic or secre
COG3444	Phosphotransferase system, man
COG3447	Predicted integral membrane se
COG3448	CBS-domain-containing membrane
COG3449	DNA gyrase inhibitor
COG3451	Type IV secretory pathway, Vir
COG3452	Predicted periplasmic ligand-b
COG3453	Uncharacterized protein conser
COG3454	Metal-dependent hydrolase invo
COG3455	Uncharacterized protein conser
COG3458	Acetyl esterase (deacetylase)
COG3459	Cellobiose phosphorylase
COG3464	Transposase and inactivated de
COG3467	Predicted flavin-nucleotide-bi
COG3469	Chitinase
COG3472	Uncharacterized conserved prot
COG3473	Maleate cis-trans isomerase
COG3474	Cytochrome c2
COG3475	LPS biosynthesis protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3476	Tryptophan-rich sensory protei
COG3478	Predicted nucleic-acid-binding
COG3479	Phenolic acid decarboxylase
COG3480	Predicted secreted protein con
COG3481	Predicted HD-superfamily hydro
COG3482	Uncharacterized conserved prot
COG3483	Tryptophan 2,3-dioxygenase (ve
COG3484	Predicted proteasome-type prot
COG3486	Lysine/ornithine N-monooxygena
COG3487	Uncharacterized iron-regulated
COG3488	Predicted thiol oxidoreductase
COG3489	Predicted periplasmic lipoprot
COG3490	Uncharacterized protein conser
COG3491	Isopenicillin N synthase and r
COG3492	Uncharacterized protein conser
COG3493	Na ⁺ /citrate symporter
COG3494	Uncharacterized protein conser
COG3495	Uncharacterized protein conser
COG3496	Uncharacterized conserved prot
COG3497	Phage tail sheath protein FI
COG3498	Phage tail tube protein FII
COG3500	Phage protein D
COG3501	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3502	Uncharacterized protein conser
COG3503	Predicted membrane protein
COG3504	Type IV secretory pathway, Vir
COG3505	Type IV secretory pathway, Vir
COG3506	Uncharacterized conserved prot
COG3507	Beta-xylosidase
COG3508	Homogentisate 1,2-dioxygenase
COG3509	Poly(3-hydroxybutyrate) depoly
COG3510	Cephalosporin hydroxylase
COG3511	Phospholipase C
COG3515	Uncharacterized protein conser
COG3516	Uncharacterized protein conser
COG3517	Uncharacterized protein conser
COG3519	Uncharacterized protein conser
COG3520	Uncharacterized protein conser
COG3522	Uncharacterized protein conser
COG3523	Uncharacterized protein conser
COG3525	N-acetyl-beta-hexosaminidase
COG3526	Uncharacterized protein conser
COG3527	Alpha-acetolactate decarboxyla
COG3530	Uncharacterized protein conser
COG3531	Predicted protein-disulfide is
COG3533	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3534	Alpha-L-arabinofuranosidase
COG3535	Uncharacterized conserved prot
COG3536	Uncharacterized protein conser
COG3537	Putative alpha-1,2-mannosidase
COG3538	Uncharacterized conserved prot
COG3539	P pilus assembly protein, pili
COG3540	Phosphodiesterase/alkaline pho
COG3541	Predicted nucleotidyltransfera
COG3542	Uncharacterized conserved prot
COG3543	Uncharacterized conserved prot
COG3544	Uncharacterized protein conser
COG3545	Predicted esterase of the alph
COG3546	Mn-containing catalase
COG3549	Plasmid maintenance system kil
COG3550	Uncharacterized protein relate
COG3552	Protein containing von Willebr
COG3554	Uncharacterized protein conser
COG3555	Aspartyl/asparaginyl beta-hydr
COG3556	Predicted membrane protein
COG3557	Uncharacterized domain/protein
COG3558	Uncharacterized protein conser
COG3559	Putative exporter of polyketid
COG3560	Predicted oxidoreductase relat

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3564	Uncharacterized protein conser
COG3566	Uncharacterized protein conser
COG3567	Uncharacterized protein conser
COG3568	Metal-dependent hydrolase
COG3569	Topoisomerase IB
COG3570	Streptomycin 6-kinase
COG3571	Predicted hydrolase of the alp
COG3572	Gamma-glutamylcysteine synthet
COG3573	Predicted oxidoreductase
COG3575	Uncharacterized protein conser
COG3576	Predicted flavin-nucleotide-bi
COG3577	Predicted aspartyl protease
COG3579	Aminopeptidase C
COG3580	Uncharacterized protein conser
COG3581	Uncharacterized protein conser
COG3583	Uncharacterized protein conser
COG3586	Uncharacterized conserved prot
COG3588	Fructose-1,6-bisphosphate aldo
COG3589	Uncharacterized conserved prot
COG3590	Predicted metalloendopeptidase
COG3591	V8-like Glu-specific endopepti
COG3592	Uncharacterized conserved prot
COG3593	Predicted ATP-dependent endonu

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3594	Fucose 4-O-acetylase and relat
COG3595	Uncharacterized conserved prot
COG3596	Predicted GTPase
COG3600	Uncharacterized phage-associat
COG3601	Predicted membrane protein
COG3602	Uncharacterized protein conser
COG3603	Uncharacterized conserved prot
COG3604	Transcriptional regulator cont
COG3605	Signal transduction protein co
COG3607	Predicted lactoylglutathione l
COG3608	Predicted deacylase
COG3609	Predicted transcriptional regu
COG3610	Uncharacterized conserved prot
COG3611	Replication initiation/membran
COG3613	Nucleoside 2-deoxyribosyltrans
COG3614	Predicted periplasmic ligand-b
COG3616	Predicted amino acid aldolase
COG3619	Predicted membrane protein
COG3620	Predicted transcriptional regu
COG3622	Hydroxypyruvate isomerase
COG3623	Putative L-xylulose-5-phosphat
COG3624	Uncharacterized enzyme of phos
COG3625	Uncharacterized enzyme of phos

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3626	Uncharacterized enzyme of phos
COG3627	Uncharacterized enzyme of phos
COG3628	Phage baseplate assembly prote
COG3633	Na ⁺ /serine symporter
COG3634	Alkyl hydroperoxide reductase,
COG3635	Predicted phosphoglycerate mut
COG3636	Predicted transcriptional regu
COG3638	ABC-type phosphate/phosphonate
COG3639	ABC-type phosphate/phosphonate
COG3640	CO dehydrogenase maturation fa
COG3641	Predicted membrane protein, pu
COG3642	Mn ²⁺ -dependent serine/threonin
COG3643	Glutamate formiminotransferase
COG3644	Uncharacterized protein conser
COG3648	Uricase (urate oxidase)
COG3651	Uncharacterized protein conser
COG3652	Predicted outer membrane prote
COG3653	N-acyl-D-aspartate/D-glutamate
COG3655	Predicted transcriptional regu
COG3657	Uncharacterized protein conser
COG3658	Cytochrome b
COG3661	Alpha-glucuronidase
COG3662	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3663	G:T/U mismatch-specific DNA gl
COG3665	Uncharacterized conserved prot
COG3666	Transposase and inactivated de
COG3667	Uncharacterized protein involv
COG3668	Plasmid stabilization system p
COG3669	Alpha-L-fucosidase
COG3670	Lignostilbene-alpha,beta-dioxy
COG3672	Predicted periplasmic protein
COG3673	Uncharacterized conserved prot
COG3679	Uncharacterized conserved prot
COG3681	Uncharacterized conserved prot
COG3682	Predicted transcriptional regu
COG3684	Tagatose-1,6-bisphosphate aldo
COG3685	Uncharacterized protein conser
COG3686	Predicted membrane protein
COG3687	Predicted metal-dependent hydr
COG3688	Predicted RNA-binding protein
COG3689	Predicted membrane protein
COG3693	Beta-1,4-xylanase
COG3694	ABC-type uncharacterized trans
COG3695	Predicted methylated DNA-prote
COG3696	Putative silver efflux pump
COG3697	Phosphoribosyl-dephospho-CoA t

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3698	Predicted periplasmic protein
COG3701	Type IV secretory pathway, Trb
COG3703	Uncharacterized protein involv
COG3704	Type IV secretory pathway, Vir
COG3705	ATP phosphoribosyltransferase
COG3706	Response regulator containing
COG3707	Response regulator with putati
COG3708	Uncharacterized protein conser
COG3709	Uncharacterized component of p
COG3710	DNA-binding winged-HTH domains
COG3714	Predicted membrane protein
COG3715	Phosphotransferase system, man
COG3716	Phosphotransferase system, man
COG3717	5-keto 4-deoxyuronate isomeras
COG3718	Uncharacterized enzyme involve
COG3720	Putative heme degradation prot
COG3723	Recombinational DNA repair pro
COG3724	Succinylarginine dihydrolase
COG3727	DNA G:T-mismatch repair endonu
COG3730	Phosphotransferase system sorb
COG3732	Phosphotransferase system sorb
COG3733	Cu ²⁺ -containing amine oxidase
COG3734	2-keto-3-deoxy-galactonokinase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3735	Uncharacterized protein conser
COG3736	Type IV secretory pathway, com
COG3737	Uncharacterized conserved prot
COG3738	Uncharacterized protein conser
COG3739	Uncharacterized integral membr
COG3741	N-formylglutamate amidohydrola
COG3745	Flp pilus assembly protein Cpa
COG3748	Predicted membrane protein
COG3749	Uncharacterized protein conser
COG3750	Uncharacterized protein conser
COG3752	Predicted membrane protein
COG3753	Uncharacterized protein conser
COG3755	Uncharacterized protein conser
COG3757	Lysozyme M1 (1,4-beta-N-acetyl
COG3758	Uncharacterized protein conser
COG3759	Predicted membrane protein
COG3760	Uncharacterized conserved prot
COG3761	NADH:ubiquinone oxidoreductase
COG3762	Predicted membrane protein
COG3763	Uncharacterized protein conser
COG3764	Sortase (surface protein trans
COG3766	Predicted membrane protein
COG3768	Predicted membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3770	Murein endopeptidase
COG3772	Phage-related lysozyme (murami
COG3773	Cell wall hydrolyses involved
COG3775	Phosphotransferase system, gal
COG3781	Predicted membrane protein
COG3785	Uncharacterized conserved prot
COG3786	Uncharacterized protein conser
COG3789	Uncharacterized protein conser
COG3791	Uncharacterized conserved prot
COG3793	Tellurite resistance protein
COG3795	Uncharacterized protein conser
COG3797	Uncharacterized protein conser
COG3799	Methylaspartate ammonia-lyase
COG3800	Predicted transcriptional regu
COG3802	Uncharacterized protein conser
COG3803	Uncharacterized protein conser
COG3806	Anti-sigma factor
COG3807	Uncharacterized protein conser
COG3808	Inorganic pyrophosphatase
COG3811	Uncharacterized protein conser
COG3812	Uncharacterized protein conser
COG3813	Uncharacterized protein conser
COG3814	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3816	Uncharacterized protein conser
COG3817	Predicted membrane protein
COG3818	Predicted acetyltransferase, G
COG3819	Predicted membrane protein
COG3820	Uncharacterized protein conser
COG3821	Predicted membrane protein
COG3823	Glutamine cyclotransferase
COG3824	Uncharacterized protein conser
COG3825	Uncharacterized protein conser
COG3826	Uncharacterized protein conser
COG3827	Uncharacterized protein conser
COG3828	Uncharacterized protein conser
COG3829	Transcriptional regulator cont
COG3830	ACT domain-containing protein
COG3832	Uncharacterized conserved prot
COG3833	ABC-type maltose transport sys
COG3835	Sugar diacid utilization regul
COG3836	2,4-dihydroxyhept-2-ene-1,7-di
COG3837	Uncharacterized conserved prot
COG3840	ABC-type thiamine transport sy
COG3844	Kynureninase
COG3845	ABC-type uncharacterized trans
COG3846	Type IV secretory pathway, Trb

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3848	Phosphohistidine swiveling dom
COG3850	Signal transduction histidine
COG3852	Signal transduction histidine
COG3853	Uncharacterized protein involv
COG3854	Uncharacterized protein conser
COG3855	Uncharacterized protein conser
COG3856	Uncharacterized conserved prot
COG3857	ATP-dependent nuclease, subuni
COG3859	Predicted membrane protein
COG3860	Uncharacterized protein conser
COG3861	Uncharacterized protein conser
COG3862	Uncharacterized protein with c
COG3865	Uncharacterized protein conser
COG3866	Pectate lyase
COG3867	Arabinogalactan endo-1,4-beta-
COG3868	Uncharacterized conserved prot
COG3869	Arginine kinase
COG3870	Uncharacterized protein conser
COG3872	Predicted metal-dependent enzy
COG3876	Uncharacterized protein conser
COG3878	Uncharacterized protein conser
COG3880	Uncharacterized protein with c
COG3882	Predicted enzyme involved in m

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3883	Uncharacterized protein conser
COG3885	Uncharacterized conserved prot
COG3886	Predicted HKD family nuclease
COG3887	Predicted signaling protein co
COG3888	Predicted transcriptional regu
COG3889	Predicted solute binding prote
COG3892	Uncharacterized protein conser
COG3893	Inactivated superfamily I heli
COG3894	Uncharacterized metal-binding
COG3895	Predicted periplasmic protein
COG3897	Predicted methyltransferase
COG3898	Uncharacterized membrane-bound
COG3900	Predicted periplasmic protein
COG3901	Regulator of nitric oxide redu
COG3906	Uncharacterized protein conser
COG3910	Predicted ATPase
COG3917	2-hydroxychromene-2-carboxylat
COG3920	Signal transduction histidine
COG3921	Uncharacterized protein conser
COG3930	Uncharacterized protein conser
COG3932	Uncharacterized ABC-type trans
COG3935	Putative primosome component a
COG3938	Proline racemase

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3944	Capsular polysaccharide biosyn
COG3945	Uncharacterized conserved prot
COG3946	Type IV secretory pathway, Vir
COG3948	Phage-related baseplate assemb
COG3949	Uncharacterized membrane prote
COG3951	Rod binding protein
COG3954	Phosphoribulokinase
COG3956	Protein containing tetrapyrrol
COG3957	Phosphoketolase
COG3958	Transketolase, C-terminal subu
COG3959	Transketolase, N-terminal subu
COG3960	Glyoxylate carboligase
COG3961	Pyruvate decarboxylase and rel
COG3962	Acetolactate synthase
COG3963	Phospholipid N-methyltransfera
COG3965	Predicted Co/Zn/Cd cation tran
COG3966	Protein involved in D-alanine
COG3967	Short-chain dehydrogenase invo
COG3968	Uncharacterized protein relate
COG3969	Predicted phosphoadenosine pho
COG3970	Fumarylacetoacetate (FAA) hydr
COG3971	2-keto-4-pentenoate hydratase
COG3973	Superfamily I DNA and RNA heli

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG3975	Predicted protease with the C-
COG3977	Alanine-alpha-ketoisovalerate
COG3980	Spore coat polysaccharide bios
COG3981	Predicted acetyltransferase
COG4002	Predicted phosphotransacetylas
COG4008	Predicted metal-binding transc
COG4012	Uncharacterized protein conser
COG4020	Uncharacterized protein conser
COG4021	Uncharacterized conserved prot
COG4028	Predicted P-loop ATPase/GTPase
COG4029	Uncharacterized protein conser
COG4031	Predicted metal-binding protei
COG4032	Predicted thiamine-pyrophospha
COG4034	Uncharacterized protein conser
COG4044	Uncharacterized protein conser
COG4046	Uncharacterized protein conser
COG4047	Uncharacterized protein conser
COG4048	Uncharacterized protein conser
COG4049	Uncharacterized protein contai
COG4050	Uncharacterized protein conser
COG4052	Uncharacterized protein relate
COG4054	Methyl coenzyme M reductase, b
COG4056	Methyl coenzyme M reductase, s

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4059	Tetrahydromethanopterin S-meth
COG4060	Tetrahydromethanopterin S-meth
COG4061	Tetrahydromethanopterin S-meth
COG4063	Tetrahydromethanopterin S-meth
COG4064	Tetrahydromethanopterin S-meth
COG4065	Uncharacterized protein conser
COG4070	Predicted peptidyl-prolyl cis-
COG4072	Uncharacterized protein conser
COG4079	Uncharacterized protein conser
COG4080	RecB-family nuclease
COG4083	Predicted membrane protein
COG4086	Predicted secreted protein
COG4088	Predicted nucleotide kinase
COG4089	Predicted membrane protein
COG4090	Uncharacterized protein conser
COG4091	Predicted homoserine dehydroge
COG4095	Uncharacterized conserved prot
COG4096	Type I site-specific restricti
COG4097	Predicted ferric reductase
COG4098	Superfamily II DNA/RNA helicas
COG4100	Cystathionine beta-lyase famil
COG4101	Predicted mannose-6-phosphate
COG4102	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4105	DNA uptake lipoprotein
COG4106	Trans-aconitate methyltransfer
COG4107	ABC-type phosphonate transport
COG4108	Peptide chain release factor R
COG4109	Predicted transcriptional regu
COG4110	Uncharacterized protein involv
COG4115	Uncharacterized protein conser
COG4116	Uncharacterized protein conser
COG4117	Thiosulfate reductase cytochro
COG4119	Predicted NTP pyrophosphohydro
COG4120	ABC-type uncharacterized trans
COG4121	Uncharacterized conserved prot
COG4122	Predicted O-methyltransferase
COG4123	Predicted O-methyltransferase
COG4125	Predicted membrane protein
COG4126	Hydantoin racemase
COG4129	Predicted membrane protein
COG4130	Predicted sugar epimerase
COG4132	ABC-type uncharacterized trans
COG4133	ABC-type transport system invo
COG4134	ABC-type uncharacterized trans
COG4136	ABC-type uncharacterized trans
COG4137	ABC-type uncharacterized trans

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4143	ABC-type thiamine transport sy
COG4145	Na+/panthothenate symporter
COG4147	Predicted symporter
COG4148	ABC-type molybdate transport s
COG4149	ABC-type molybdate transport s
COG4150	ABC-type sulfate transport sys
COG4152	ABC-type uncharacterized trans
COG4154	Fucose dissimilation pathway p
COG4160	ABC-type arginine/histidine tr
COG4172	ABC-type uncharacterized trans
COG4174	ABC-type uncharacterized trans
COG4175	ABC-type proline/glycine betai
COG4176	ABC-type proline/glycine betai
COG4178	ABC-type uncharacterized trans
COG4181	Predicted ABC-type transport s
COG4187	Arginine degradation protein (
COG4188	Predicted diene lactone hydrola
COG4189	Predicted transcriptional regu
COG4191	Signal transduction histidine
COG4194	Predicted membrane protein
COG4198	Uncharacterized conserved prot
COG4199	Uncharacterized protein conser
COG4206	Outer membrane cobalamin recep

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4208	ABC-type sulfate transport sys
COG4209	ABC-type polysaccharide transp
COG4211	ABC-type glucose/galactose tra
COG4213	ABC-type xylose transport syst
COG4214	ABC-type xylose transport syst
COG4215	ABC-type arginine transport sy
COG4221	Short-chain alcohol dehydrogen
COG4222	Uncharacterized protein conser
COG4223	Uncharacterized protein conser
COG4224	Uncharacterized protein conser
COG4225	Predicted unsaturated glucuron
COG4226	Uncharacterized protein encode
COG4229	Predicted enolase-phosphatase
COG4230	Delta 1-pyrroline-5-carboxylat
COG4231	Indolepyruvate ferredoxin oxid
COG4232	Thiol:disulfide interchange pr
COG4233	Uncharacterized protein predic
COG4235	Cytochrome c biogenesis factor
COG4239	ABC-type uncharacterized trans
COG4240	Predicted kinase
COG4246	Uncharacterized protein conser
COG4247	3-phytase (myo-inositol-hexaph
COG4251	Bacteriophytochrome (light-reg

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4252	Predicted transmembrane sensor
COG4254	Uncharacterized protein conser
COG4257	Streptogramin lyase
COG4262	Predicted spermidine synthase
COG4263	Nitrous oxide reductase
COG4266	Allantoicase
COG4267	Predicted membrane protein
COG4269	Predicted membrane protein
COG4271	Predicted nucleotide-binding p
COG4272	Predicted membrane protein
COG4274	Uncharacterized conserved prot
COG4275	Uncharacterized conserved prot
COG4277	Predicted DNA-binding protein
COG4279	Uncharacterized conserved prot
COG4281	Acyl-CoA-binding protein
COG4282	Protein involved in beta-1,3-g
COG4283	Uncharacterized conserved prot
COG4284	UDP-glucose pyrophosphorylase
COG4285	Uncharacterized conserved prot
COG4286	Uncharacterized conserved prot
COG4287	PhoPQ-activated pathogenicity-
COG4288	Uncharacterized protein conser
COG4290	Guanyl-specific ribonuclease S

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4291	Predicted membrane protein
COG4292	Predicted membrane protein
COG4294	UV damage repair endonuclease
COG4295	Uncharacterized protein conser
COG4297	Uncharacterized protein contai
COG4299	Uncharacterized conserved prot
COG4300	Predicted permease, cadmium re
COG4301	Uncharacterized conserved prot
COG4302	Ethanolamine ammonia-lyase, sm
COG4303	Ethanolamine ammonia-lyase, la
COG4305	Endoglucanase C-terminal domai
COG4306	Uncharacterized protein conser
COG4307	Uncharacterized protein conser
COG4309	Uncharacterized conserved prot
COG4311	Sarcosine oxidase delta subuni
COG4312	Uncharacterized protein conser
COG4313	Protein involved in meta-pathw
COG4314	Predicted lipoprotein involved
COG4317	Uncharacterized protein conser
COG4318	Uncharacterized protein conser
COG4319	Ketosteroid isomerase homolog
COG4324	Predicted aminopeptidase
COG4325	Predicted membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4326	Sporulation control protein
COG4328	Uncharacterized protein conser
COG4329	Predicted membrane protein
COG4330	Predicted membrane protein
COG4331	Predicted membrane protein
COG4336	Uncharacterized conserved prot
COG4337	Uncharacterized protein conser
COG4339	Uncharacterized protein conser
COG4340	Uncharacterized protein conser
COG4341	Predicted HD phosphohydrolase
COG4344	Uncharacterized protein conser
COG4345	Uncharacterized protein conser
COG4346	Predicted membrane-bound dolc
COG4347	Predicted membrane protein
COG4353	Uncharacterized conserved prot
COG4354	Predicted bile acid beta-gluco
COG4357	Uncharacterized conserved prot
COG4360	ATP adenylyltransferase (5',5'
COG4362	Nitric oxide synthase, oxygena
COG4365	Uncharacterized protein conser
COG4371	Predicted membrane protein
COG4372	Uncharacterized protein conser
COG4377	Predicted membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4385	Bacteriophage P2-related tail
COG4394	Uncharacterized protein conser
COG4395	Uncharacterized protein conser
COG4398	Uncharacterized protein conser
COG4399	Uncharacterized protein conser
COG4401	Chorismate mutase
COG4402	Uncharacterized protein conser
COG4403	Lantibiotic modifying enzyme
COG4405	Uncharacterized protein conser
COG4412	Uncharacterized protein conser
COG4420	Predicted membrane protein
COG4424	Uncharacterized protein conser
COG4425	Predicted membrane protein
COG4427	Uncharacterized protein conser
COG4430	Uncharacterized protein conser
COG4443	Uncharacterized protein conser
COG4445	Hydroxylase for synthesis of 2
COG4447	Uncharacterized protein relate
COG4448	L-asparaginase II
COG4451	Ribulose bisphosphate carboxyl
COG4452	Inner membrane protein involve
COG4453	Uncharacterized protein conser
COG4456	Virulence-associated protein a

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4463	Transcriptional repressor of c
COG4464	Capsular polysaccharide biosyn
COG4465	Pleiotropic transcriptional re
COG4466	Uncharacterized protein conser
COG4467	Uncharacterized protein conser
COG4468	Galactose-1-phosphate uridyldr
COG4469	Competence protein
COG4470	Uncharacterized protein conser
COG4472	Uncharacterized protein conser
COG4473	Predicted ABC-type exoprotein
COG4474	Uncharacterized protein conser
COG4475	Uncharacterized protein conser
COG4476	Uncharacterized protein conser
COG4477	Negative regulator of septatio
COG4478	Predicted membrane protein
COG4479	Uncharacterized protein conser
COG4481	Uncharacterized protein conser
COG4483	Uncharacterized protein conser
COG4485	Predicted membrane protein
COG4492	ACT domain-containing protein
COG4493	Uncharacterized protein conser
COG4496	Uncharacterized protein conser
COG4499	Predicted membrane protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4508	Uncharacterized protein conser
COG4521	ABC-type taurine transport sys
COG4525	ABC-type taurine transport sys
COG4531	ABC-type Zn ²⁺ transport system
COG4533	ABC-type uncharacterized trans
COG4535	Putative Mg ²⁺ and Co ²⁺ transpo
COG4536	Putative Mg ²⁺ and Co ²⁺ transpo
COG4537	Competence protein ComGC
COG4539	Predicted membrane protein
COG4542	Protein involved in propanedio
COG4547	Cobalamin biosynthesis protein
COG4550	Predicted membrane protein
COG4551	Predicted protein tyrosine pho
COG4552	Predicted acetyltransferase in
COG4553	Poly-beta-hydroxyalkanoate dep
COG4555	ABC-type Na ⁺ transport system,
COG4558	ABC-type hemin transport syste
COG4559	ABC-type hemin transport syste
COG4565	Response regulator of citrate/
COG4566	Response regulator
COG4567	Response regulator consisting
COG4569	Acetaldehyde dehydrogenase (ac
COG4573	Predicted tagatose 6-phosphate

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4576	Carbon dioxide concentrating m
COG4577	Carbon dioxide concentrating m
COG4579	Isocitrate dehydrogenase kinas
COG4580	Maltoporin (phage lambda and m
COG4581	Superfamily II RNA helicase
COG4582	Uncharacterized protein conser
COG4583	Sarcosine oxidase gamma subuni
COG4584	Transposase and inactivated de
COG4586	ABC-type uncharacterized trans
COG4587	ABC-type uncharacterized trans
COG4588	Accessory colonization factor
COG4589	Predicted CDP-diglyceride synt
COG4591	ABC-type transport system, inv
COG4592	ABC-type Fe ²⁺ -enterobactin tra
COG4597	ABC-type amino acid transport
COG4598	ABC-type histidine transport s
COG4603	ABC-type uncharacterized trans
COG4604	ABC-type enterochelin transpor
COG4605	ABC-type enterochelin transpor
COG4606	ABC-type enterochelin transpor
COG4607	ABC-type enterochelin transpor
COG4615	ABC-type siderophore export sy
COG4618	ABC-type protease/lipase trans

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4623	Predicted soluble lytic transg
COG4624	Iron only hydrogenase large su
COG4625	Uncharacterized protein with a
COG4626	Phage terminase-like protein,
COG4630	Xanthine dehydrogenase, iron-s
COG4631	Xanthine dehydrogenase, molybd
COG4635	Flavodoxin
COG4637	Predicted ATPase
COG4642	Uncharacterized protein conser
COG4643	Uncharacterized protein conser
COG4644	Transposase and inactivated de
COG4645	Uncharacterized protein conser
COG4647	Acetone carboxylase, gamma sub
COG4649	Uncharacterized protein conser
COG4650	Sigma54-dependent transcriptio
COG4651	Kef-type K ⁺ transport system,
COG4653	Predicted phage phi-C31 gp36 m
COG4655	Predicted membrane protein
COG4656	Predicted NADH:ubiquinone oxid
COG4657	Predicted NADH:ubiquinone oxid
COG4658	Predicted NADH:ubiquinone oxid
COG4659	Predicted NADH:ubiquinone oxid
COG4660	Predicted NADH:ubiquinone oxid

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4662	ABC-type tungstate transport s
COG4663	TRAP-type mannitol/chloroaroma
COG4664	TRAP-type mannitol/chloroaroma
COG4665	TRAP-type mannitol/chloroaroma
COG4666	TRAP-type uncharacterized tran
COG4667	Predicted esterase of the alph
COG4668	Mannitol/fructose-specific pho
COG4669	Type III secretory pathway, li
COG4670	Acyl CoA:acetate/3-ketoacid Co
COG4671	Predicted glycosyl transferase
COG4672	Phage-related protein
COG4674	Uncharacterized ABC-type trans
COG4675	Microcystin-dependent protein
COG4676	Uncharacterized protein conser
COG4677	Pectin methylesterase
COG4678	Muramidase (phage lambda lysoz
COG4679	Phage-related protein
COG4680	Uncharacterized protein conser
COG4681	Uncharacterized protein conser
COG4685	Uncharacterized protein conser
COG4687	Uncharacterized protein conser
COG4689	Acetoacetate decarboxylase
COG4695	Phage-related protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4698	Uncharacterized protein conser
COG4700	Uncharacterized protein conser
COG4702	Uncharacterized conserved prot
COG4704	Uncharacterized protein conser
COG4708	Predicted membrane protein
COG4717	Uncharacterized conserved prot
COG4721	Predicted membrane protein
COG4722	Phage-related protein
COG4741	Predicted secreted endonucleas
COG4748	Uncharacterized conserved prot
COG4749	Uncharacterized protein conser
COG4753	Response regulator containing
COG4755	Uncharacterized protein conser
COG4756	Predicted cation transporter
COG4758	Predicted membrane protein
COG4760	Predicted membrane protein
COG4764	Uncharacterized protein conser
COG4765	Uncharacterized protein conser
COG4766	Ethanolamine utilization prote
COG4768	Uncharacterized protein contai
COG4769	Predicted membrane protein
COG4774	Outer membrane receptor for mo
COG4775	Outer membrane protein/protect

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4778	ABC-type phosphonate transport
COG4779	ABC-type enterobactin transpor
COG4782	Uncharacterized protein conser
COG4783	Putative Zn-dependent protease
COG4784	Putative Zn-dependent protease
COG4787	Flagellar basal body rod prote
COG4789	Type III secretory pathway, co
COG4790	Type III secretory pathway, co
COG4791	Type III secretory pathway, co
COG4794	Type III secretory pathway, co
COG4795	Type II secretory pathway, com
COG4796	Type II secretory pathway, com
COG4800	Predicted transcriptional regu
COG4801	Predicted acyltransferase
COG4803	Predicted membrane protein
COG4804	Uncharacterized conserved prot
COG4806	L-rhamnose isomerase
COG4807	Uncharacterized protein conser
COG4809	Archaeal ADP-dependent phospho
COG4813	Trehalose utilization protein
COG4816	Ethanolamine utilization prote
COG4818	Predicted membrane protein
COG4819	Ethanolamine utilization prote

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4820	Ethanolamine utilization prote
COG4822	Cobalamin biosynthesis protein
COG4824	Phage-related holin (Lysis pro
COG4825	Uncharacterized membrane-ancho
COG4826	Serine protease inhibitor
COG4830	Ribosomal protein S26
COG4832	Uncharacterized conserved prot
COG4833	Predicted glycosyl hydrolase
COG4834	Uncharacterized protein conser
COG4836	Predicted membrane protein
COG4838	Uncharacterized protein conser
COG4844	Uncharacterized protein conser
COG4846	Membrane protein involved in c
COG4848	Uncharacterized protein conser
COG4849	Uncharacterized protein conser
COG4850	Uncharacterized conserved prot
COG4851	Protein involved in sex pherom
COG4852	Predicted membrane protein
COG4853	Uncharacterized protein conser
COG4856	Uncharacterized protein conser
COG4857	Predicted kinase
COG4860	Uncharacterized protein conser
COG4861	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4862	Negative regulator of genetic
COG4864	Uncharacterized protein conser
COG4865	Glutamate mutase epsilon subun
COG4866	Uncharacterized conserved prot
COG4867	Uncharacterized protein with a
COG4868	Uncharacterized protein conser
COG4870	Cysteine protease
COG4872	Predicted membrane protein
COG4874	Uncharacterized protein conser
COG4878	Uncharacterized protein conser
COG4879	Uncharacterized protein conser
COG4880	Secreted protein containing C-
COG4881	Predicted membrane protein
COG4886	Leucine-rich repeat (LRR) prot
COG4887	Uncharacterized metal-binding
COG4889	Predicted helicase
COG4901	Ribosomal protein S25
COG4905	Predicted membrane protein
COG4908	Uncharacterized protein contai
COG4912	Predicted DNA alkylation repai
COG4914	Predicted nucleotidyltransfera
COG4917	Ethanolamine utilization prote
COG4923	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4926	Phage-related protein
COG4928	Predicted P-loop ATPase
COG4932	Predicted outer membrane prote
COG4941	Predicted RNA polymerase sigma
COG4942	Membrane-bound metallopeptidas
COG4944	Uncharacterized protein conser
COG4946	Uncharacterized protein relate
COG4949	Uncharacterized membrane-ancho
COG4951	Uncharacterized protein conser
COG4952	Predicted sugar isomerase
COG4953	Membrane carboxypeptidase/peni
COG4956	Integral membrane protein (PIN
COG4960	Flp pilus assembly protein, pr
COG4961	Flp pilus assembly protein Tad
COG4962	Flp pilus assembly protein, AT
COG4963	Flp pilus assembly protein, AT
COG4964	Flp pilus assembly protein, se
COG4965	Flp pilus assembly protein Tad
COG4968	Tfp pilus assembly protein Pil
COG4969	Tfp pilus assembly protein, ma
COG4972	Tfp pilus assembly protein, AT
COG4973	Site-specific recombinase XerC
COG4976	Predicted methyltransferase (c

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG4977	Transcriptional regulator cont
COG4981	Enoyl reductase domain of yeas
COG4982	3-oxoacyl-[acyl-carrier protei
COG4983	Uncharacterized conserved prot
COG4986	ABC-type anion transport syste
COG4987	ABC-type transport system invo
COG4988	ABC-type transport system invo
COG4989	Predicted oxidoreductase
COG4993	Glucose dehydrogenase
COG4997	Uncharacterized conserved prot
COG4998	Predicted endonuclease (RecB f
COG5000	Signal transduction histidine
COG5002	Signal transduction histidine
COG5006	Predicted permease, DMT superf
COG5007	Predicted transcriptional regu
COG5008	Tfp pilus assembly protein, AT
COG5009	Membrane carboxypeptidase/peni
COG5010	Flp pilus assembly protein Tad
COG5011	Uncharacterized protein conser
COG5012	Predicted cobalamin binding pr
COG5013	Nitrate reductase alpha subuni
COG5014	Predicted Fe-S oxidoreductase
COG5016	Pyruvate/oxaloacetate carboxyl

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5020	Mannosyltransferase
COG5021	Ubiquitin-protein ligase
COG5022	Myosin heavy chain
COG5023	Tubulin
COG5026	Hexokinase
COG5027	Histone acetyltransferase (MYS
COG5028	Vesicle coat complex COPII, su
COG5029	Prenyltransferase, beta subuni
COG5032	Phosphatidylinositol kinase an
COG5033	Transcription initiation facto
COG5038	Ca ²⁺ -dependent lipid-binding p
COG5039	Exopolysaccharide biosynthesis
COG5041	Casein kinase II, beta subunit
COG5042	Purine nucleoside permease
COG5043	Vacuolar protein sorting-assoc
COG5044	RAB proteins geranylgeranyltra
COG5047	Vesicle coat complex COPII, su
COG5049	5'-3' exonuclease
COG5050	sn-1,2-diacylglycerol ethanolol
COG5052	Protein involved in membrane t
COG5054	Mitochondrial sulfhydryl oxida
COG5057	Phosphotyrosyl phosphatase act
COG5059	Kinesin-like protein

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5068	Regulator of arginine metaboli
COG5070	Nucleotide-sugar transporter
COG5071	26S proteasome regulatory comp
COG5075	Uncharacterized conserved prot
COG5076	Transcription factor involved
COG5078	Ubiquitin-protein ligase
COG5080	Rab GTPase interacting factor,
COG5081	Predicted membrane protein
COG5083	Uncharacterized protein involv
COG5084	Cleavage and polyadenylation s
COG5088	Rad5p-binding protein
COG5092	N-myristoyl transferase
COG5095	Transcription initiation facto
COG5096	Vesicle coat complex, various
COG5098	Chromosome condensation comple
COG5099	RNA-binding protein of the Puf
COG5100	Nuclear pore protein
COG5103	Cell division control protein,
COG5105	Mitotic inducer, protein phosp
COG5107	Pre-mRNA 3'-end processing (cl
COG5110	26S proteasome regulatory comp
COG5111	DNA-directed RNA polymerase II
COG5113	Ubiquitin fusion degradation p

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5114	Histone acetyltransferase comp
COG5116	26S proteasome regulatory comp
COG5117	Protein involved in the nuclea
COG5126	Ca ²⁺ -binding protein (EF-Hand
COG5129	Nuclear protein with HMG-like
COG5132	Cell cycle control protein, G1
COG5133	Uncharacterized conserved prot
COG5137	Histone chaperone involved in
COG5140	Ubiquitin fusion-degradation p
COG5141	PHD zinc finger-containing pro
COG5143	Synaptobrevin/VAMP-like protei
COG5144	RNA polymerase II transcriptio
COG5145	DNA excision repair protein
COG5147	Myb superfamily proteins, incl
COG5150	Class 2 transcription represso
COG5151	RNA polymerase II transcriptio
COG5153	Putative lipase essential for
COG5154	RNA-binding protein required f
COG5155	Separase, a protease involved
COG5158	Proteins involved in synaptic
COG5159	26S proteasome regulatory comp
COG5160	Protease, Ulp1 family
COG5161	Pre-mRNA cleavage and polyaden

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5163	Protein required for biogenesi
COG5165	Nucleosome-binding factor SPN,
COG5166	Uncharacterized conserved prot
COG5167	Protein involved in vacuole im
COG5169	Heat shock transcription facto
COG5170	Serine/threonine protein phosp
COG5171	Ran GTPase-activating protein
COG5175	Transcriptional repressor
COG5176	Splicing factor (branch point
COG5178	U5 snRNP spliceosome subunit
COG5179	Transcription initiation facto
COG5180	Protein interacting with poly(
COG5181	U2 snRNP spliceosome subunit
COG5182	Splicing factor 3b, subunit 2
COG5183	Protein involved in mRNA turno
COG5184	Alpha-tubulin suppressor and r
COG5185	Protein involved in chromosome
COG5186	Poly(A) polymerase
COG5187	26S proteasome regulatory comp
COG5190	TFIIF-interacting CTD phosphat
COG5191	Uncharacterized conserved prot
COG5193	La protein, small RNA-binding
COG5194	Component of SCF ubiquitin lig

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5196	ER lumen protein retaining rec
COG5197	Predicted membrane protein
COG5199	Calponin
COG5202	Predicted membrane protein
COG5206	Glycosylphosphatidylinositol t
COG5208	CCAAT-binding factor, subunit
COG5209	Uncharacterized protein involv
COG5210	GTPase-activating protein
COG5211	RNA polymerase II-interacting
COG5213	Polyadenylation factor I compl
COG5214	DNA polymerase alpha-primase c
COG5215	Karyopherin (importin) beta
COG5216	Uncharacterized conserved prot
COG5217	Microtubule-binding protein in
COG5219	Uncharacterized conserved prot
COG5220	Cdk activating kinase (CAK)/RN
COG5221	Dopey and related predicted le
COG5222	Uncharacterized conserved prot
COG5226	mRNA capping enzyme, guanylylt
COG5228	mRNA deadenylase subunit
COG5237	Predicted membrane protein
COG5238	Ran GTPase-activating protein
COG5239	mRNA deadenylase, exonuclease

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5240	Vesicle coat complex COPI, gam
COG5241	Nucleotide excision repair end
COG5249	Golgi protein involved in Golg
COG5253	Phosphatidylinositol-4-phospha
COG5255	Uncharacterized protein conser
COG5256	Translation elongation factor
COG5257	Translation initiation factor
COG5258	GTPase
COG5261	Protein involved in regulation
COG5264	Vacuolar transporter chaperone
COG5265	ABC-type transport system invo
COG5267	Uncharacterized protein conser
COG5269	Ribosome-associated chaperone
COG5270	PUA domain (predicted RNA-bind
COG5271	AAA ATPase containing von Will
COG5272	Ubiquitin
COG5273	Uncharacterized protein contai
COG5277	Actin and related proteins
COG5278	Predicted periplasmic ligand-b
COG5280	Phage-related minor tail prote
COG5281	Phage-related minor tail prote
COG5283	Phage-related tail protein
COG5295	Autotransporter adhesin

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5310	Homospermidine synthase
COG5316	Uncharacterized conserved prot
COG5317	Uncharacterized protein conser
COG5319	Uncharacterized protein conser
COG5321	Uncharacterized protein conser
COG5322	Predicted dehydrogenase
COG5328	Uncharacterized protein conser
COG5329	Phosphoinositide polyphosphata
COG5336	Uncharacterized protein conser
COG5338	Uncharacterized protein conser
COG5343	Uncharacterized protein conser
COG5347	GTPase-activating protein that
COG5349	Uncharacterized protein conser
COG5352	Uncharacterized protein conser
COG5353	Uncharacterized protein conser
COG5360	Uncharacterized protein conser
COG5361	Uncharacterized conserved prot
COG5368	Uncharacterized protein conser
COG5373	Predicted membrane protein
COG5375	Uncharacterized protein conser
COG5383	Uncharacterized protein conser
COG5385	Uncharacterized protein conser
COG5394	Uncharacterized protein conser

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5397	Uncharacterized conserved prot
COG5400	Uncharacterized protein conser
COG5402	Uncharacterized conserved prot
COG5405	ATP-dependent protease HslVU (
COG5406	Nucleosome binding factor SPN,
COG5407	Preprotein translocase subunit
COG5429	Uncharacterized secreted prote
COG5434	Endopolygalacturonase
COG5438	Predicted multitransmembrane p
COG5443	Flagellar biosynthesis regulat
COG5445	Predicted secreted protein
COG5452	Uncharacterized conserved prot
COG5458	Uncharacterized conserved prot
COG5461	Type IV pili component
COG5465	Uncharacterized conserved prot
COG5467	Uncharacterized conserved prot
COG5470	Uncharacterized conserved prot
COG5477	Predicted small integral membr
COG5478	Predicted small integral membr
COG5482	Uncharacterized conserved prot
COG5485	Predicted ester cyclase
COG5487	Small integral membrane protei
COG5493	Uncharacterized conserved prot

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
COG5496	Predicted thioesterase
COG5499	Predicted transcription regula
COG5500	Predicted integral membrane pr
COG5508	Uncharacterized conserved smal
COG5526	Uncharacterized conserved prot
COG5528	Predicted integral membrane pr
COG5530	Predicted integral membrane pr
COG5531	SWIB-domain-containing protein
COG5533	Ubiquitin C-terminal hydrolase
COG5545	Predicted P-loop ATPase and in
COG5552	Uncharacterized conserved prot
COG5563	Predicted integral membrane pr
COG5565	Bacteriophage terminase large
COG5572	Predicted integral membrane pr
COG5579	Uncharacterized conserved prot
COG5594	Uncharacterized integral membr
COG5602	Histone deacetylase complex, S
COG5605	Predicted small integral membr
COG5607	Uncharacterized conserved prot
COG5644	Uncharacterized conserved prot
COG5649	Uncharacterized conserved prot
COG5656	Importin, protein involved in
COG5662	Predicted transmembrane transc

Continúa ...

Cuadro 21: COGs eliminados (continuación)

COG	Descripción
-----	-------------

Fin tabla

12. Apendice 4: Cuadro COGs finales

Cuadro con los COGs finales con el número de grupos en que fueron divididos por MCL (ver 4.4).

Cuadro 22: 584 COGs finales con el número de grupos en que fueron divididos (ver 4.4).

COG	Descripción	Grupos
COG0008	Glutamyl- and glutaminyl-tRNA	2
COG0010	Arginase/agmatinase/formimiono	3
COG0019	Diaminopimelate decarboxylase	3
COG0025	NhaP-type Na ⁺ /H ⁺ and K ⁺ /H ⁺ ant	3
COG0028	Thiamine pyrophosphate-requiri	3
COG0031	Cysteine synthase	2
COG0037	Predicted ATPase of the PP-loo	2
COG0042	tRNA-dihydrouridine synthase	3
COG0044	Dihydroorotase and related cyc	2
COG0052	Ribosomal protein S2	2
COG0057	Glyceraldehyde-3-phosphate deh	2
COG0067	Glutamate synthase domain 1	2
COG0070	Glutamate synthase domain 3	2
COG0071	Molecular chaperone (small hea	3
COG0073	EMAP domain	3
COG0079	Histidinol-phosphate/aromatic	2
COG0085	DNA-directed RNA polymerase, b	2
COG0086	DNA-directed RNA polymerase, b	2
COG0088	Ribosomal protein L4	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0089	Ribosomal protein L23	2
COG0091	Ribosomal protein L22	2
COG0095	Lipoate-protein ligase A	2
COG0102	Ribosomal protein L13	2
COG0110	Acetyltransferase (isoleucine	6
COG0111	Phosphoglycerate dehydrogenase	2
COG0115	Branched-chain amino acid amin	4
COG0119	Isopropylmalate/homocitrate/ci	3
COG0121	Predicted glutamine amidotrans	2
COG0122	3-methyladenine DNA glycosylas	2
COG0129	Dihydroxyacid dehydratase/phos	2
COG0142	Geranylgeranyl pyrophosphate s	3
COG0144	tRNA and rRNA cytosine-C5-meth	2
COG0145	N-methylhydantoinase A/acetone	2
COG0147	Anthranilate/para-aminobenzoat	2
COG0149	Triosephosphate isomerase	2
COG0154	Asp-tRNAAsn/Glu-tRNAGln amidot	3
COG0160	4-aminobutyrate aminotransfera	2
COG0161	Adenosylmethionine-8-amino-7-o	4
COG0174	Glutamine synthetase	2
COG0175	3'-phosphoadenosine 5'-phospho	3
COG0179	2-keto-4-pentenoate hydratase/	2
COG0183	Acetyl-CoA acetyltransferase	7

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0184	Ribosomal protein S15P/S13E	2
COG0186	Ribosomal protein S17	2
COG0187	Type IIA topoisomerase (DNA gy	3
COG0188	Type IIA topoisomerase (DNA gy	2
COG0189	Glutathione synthase/Ribosomal	3
COG0195	Transcription elongation facto	2
COG0197	Ribosomal protein L16/L10E	2
COG0198	Ribosomal protein L24	2
COG0199	Ribosomal protein S14	3
COG0200	Ribosomal protein L15	2
COG0202	DNA-directed RNA polymerase, a	2
COG0204	1-acyl-sn-glycerol-3-phosphate	3
COG0207	Thymidylate synthase	2
COG0212	5-formyltetrahydrofolate cyclo	2
COG0227	Ribosomal protein L28	2
COG0231	Translation elongation factor	2
COG0235	Ribulose-5-phosphate 4-epimera	3
COG0237	Dephospho-CoA kinase	2
COG0243	Anaerobic dehydrogenases, typi	3
COG0244	Ribosomal protein L10	2
COG0246	Mannitol-1-phosphate/altronate	2
COG0247	Fe-S oxidoreductase	3
COG0250	Transcription antiterminator	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0251	Putative translation initiatio	5
COG0254	Ribosomal protein L31	2
COG0255	Ribosomal protein L29	2
COG0256	Ribosomal protein L18	2
COG0257	Ribosomal protein L36	2
COG0258	5'-3' exonuclease (including N	2
COG0262	Dihydrofolate reductase	2
COG0267	Ribosomal protein L33	2
COG0271	Stress-induced morphogen (acti	2
COG0277	FAD/FMN-containing dehydrogena	7
COG0280	Phosphotransacetylase	2
COG0284	Orotidine-5'-phosphate decarbo	2
COG0287	Prephenate dehydrogenase	2
COG0288	Carbonic anhydrase	2
COG0294	Dihydropteroate synthase and r	2
COG0296	1,4-alpha-glucan branching enz	2
COG0300	Short-chain dehydrogenases of	3
COG0304	3-oxoacyl-(acyl-carrier-protei	3
COG0310	ABC-type Co2+ transport system	2
COG0312	Predicted Zn-dependent proteas	3
COG0316	Uncharacterized conserved prot	2
COG0317	Guanosine polyphosphate pyroph	2
COG0318	Acyl-CoA synthetases (AMP-form	6

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0328	Ribonuclease HI	2
COG0329	Dihydrodipicolinate synthase/N	2
COG0330	Membrane protease subunits, st	5
COG0331	(acyl-carrier-protein) S-malon	2
COG0333	Ribosomal protein L32	3
COG0334	Glutamate dehydrogenase/leucin	2
COG0346	Lactoylglutathione lyase and r	6
COG0348	Polyferredoxin	2
COG0349	Ribonuclease D	2
COG0358	DNA primase (bacterial type)	2
COG0361	Translation initiation factor	2
COG0363	6-phosphogluconolactonase/Gluc	2
COG0365	Acyl-coenzyme A synthetases/AM	6
COG0366	Glycosidases	4
COG0382	4-hydroxybenzoate polyprenyltr	3
COG0383	Alpha-mannosidase	2
COG0385	Predicted Na ⁺ -dependent transp	2
COG0388	Predicted amidohydrolase	3
COG0389	Nucleotidyltransferase/DNA pol	2
COG0391	Uncharacterized conserved prot	2
COG0392	Predicted integral membrane pr	2
COG0394	Protein-tyrosine-phosphatase	3
COG0395	ABC-type sugar transport syste	7

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0399	Predicted pyridoxal phosphate-	2
COG0400	Predicted esterase	2
COG0402	Cytosine deaminase and related	3
COG0404	Glycine cleavage system T prot	2
COG0406	Fructose-2,6-bisphosphatase	4
COG0410	ABC-type branched-chain amino	3
COG0411	ABC-type branched-chain amino	2
COG0412	Dienelactone hydrolase and rel	2
COG0417	DNA polymerase elongation subu	2
COG0420	DNA repair exonuclease	2
COG0425	Predicted redox protein, regul	3
COG0431	Predicted flavoprotein	4
COG0433	Predicted ATPase	2
COG0436	Aspartate/tyrosine/aromatic am	6
COG0438	Glycosyltransferase	22
COG0443	Molecular chaperone	2
COG0444	ABC-type dipeptide/oligopeptid	2
COG0446	Uncharacterized NAD(FAD)-depen	5
COG0448	ADP-glucose pyrophosphorylase	2
COG0451	Nucleoside-diphosphate-sugar e	11
COG0454	Histone acetyltransferase HPA2	21
COG0457	FOG: TPR repeat	7
COG0459	Chaperonin GroEL (HSP60 family	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0463	Glycosyltransferases involved	15
COG0464	ATPases of the AAA+ class	2
COG0470	ATPase involved in DNA replica	3
COG0471	Di- and tricarboxylate transpo	3
COG0472	UDP-N-acetylmuramyl pentapepti	2
COG0477	Permeases of the major facilit	49
COG0483	Archaeal fructose-1,6-bisphosp	2
COG0488	ATPase components of ABC trans	4
COG0489	ATPases involved in chromosome	2
COG0491	Zn-dependent hydrolases, inclu	12
COG0492	Thioredoxin reductase	3
COG0493	NADPH-dependent glutamate synt	3
COG0494	NTP pyrophosphohydrolases incl	6
COG0495	Leucyl-tRNA synthetase	2
COG0498	Threonine synthase	3
COG0500	SAM-dependent methyltransferas	33
COG0501	Zn-dependent protease with cha	3
COG0503	Adenine/guanine phosphoribosyl	3
COG0508	Pyruvate/2-oxoglutarate dehydr	2
COG0510	Predicted choline kinase invol	2
COG0511	Biotin carboxyl carrier protei	2
COG0513	Superfamily II DNA and RNA hel	4
COG0515	Serine/threonine protein kinas	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0516	IMP dehydrogenase/GMP reductas	2
COG0517	FOG: CBS domain	12
COG0518	GMP synthase - Glutamine amido	3
COG0520	Selenocysteine lyase	2
COG0522	Ribosomal protein S4 and relat	2
COG0523	Putative GTPases (G3E family)	2
COG0524	Sugar kinases, ribokinase fami	7
COG0526	Thiol-disulfide isomerase and	11
COG0531	Amino acid transporters	5
COG0534	Na ⁺ -driven multidrug efflux pu	6
COG0535	Predicted Fe-S oxidoreductases	3
COG0537	Diadenosine tetraphosphate (Ap	3
COG0538	Isocitrate dehydrogenases	2
COG0542	ATPases with chaperone activit	5
COG0543	2-polyprenylphenol hydroxylase	3
COG0546	Predicted phosphatases	2
COG0547	Anthranilate phosphoribosyltra	2
COG0553	Superfamily II DNA/RNA helicas	2
COG0557	Exoribonuclease R	2
COG0558	Phosphatidylglycerophosphate s	2
COG0559	Branched-chain amino acid ABC-	3
COG0560	Phosphoserine phosphatase	2
COG0561	Predicted hydrolases of the HA	8

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0563	Adenylate kinase and related k	2
COG0564	Pseudouridylate synthases, 23S	5
COG0566	rRNA methylases	3
COG0568	DNA-directed RNA polymerase, s	2
COG0569	K ⁺ transport systems, NAD-bind	2
COG0574	Phosphoenolpyruvate synthase/p	2
COG0575	CDP-diglyceride synthetase	2
COG0577	ABC-type antimicrobial peptide	4
COG0579	Predicted dehydrogenase	2
COG0582	Integrase	4
COG0583	Transcriptional regulator	37
COG0584	Glycerophosphoryl diester phos	2
COG0586	Uncharacterized membrane-assoc	3
COG0589	Universal stress protein UspA	3
COG0591	Na ⁺ /proline symporter	3
COG0592	DNA polymerase sliding clamp s	2
COG0596	Predicted hydrolases or acyltr	15
COG0598	Mg ²⁺ and Co ²⁺ transporters	2
COG0599	Uncharacterized homolog of gam	2
COG0600	ABC-type nitrate/sulfonate/bic	2
COG0601	ABC-type dipeptide/oligopeptid	5
COG0602	Organic radical activating enz	2
COG0604	NADPH:quinone reductase and re	7

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0607	Rhodanese-related sulfurtransf	6
COG0608	Single-stranded DNA-specific e	2
COG0609	ABC-type Fe ³⁺ -siderophore tran	4
COG0612	Predicted Zn-dependent peptida	3
COG0613	Predicted metal-dependent phos	2
COG0614	ABC-type Fe ³⁺ -hydroxamate tran	5
COG0616	Periplasmic serine proteases (2
COG0618	Exopolyphosphatase-related pro	2
COG0619	ABC-type cobalt transport syst	2
COG0621	2-methylthioadenine synthetase	2
COG0622	Predicted phosphoesterase	3
COG0624	Acetylnithine deacetylase/Su	8
COG0625	Glutathione S-transferase	9
COG0626	Cystathionine beta-lyases/cyst	3
COG0628	Predicted permease	4
COG0630	Type IV secretory pathway, Vir	2
COG0635	Coproporphyrinogen III oxidase	2
COG0636	F0F1-type ATP synthase, subuni	4
COG0637	Predicted phosphatase/phosphoh	2
COG0638	20S proteasome, alpha and beta	3
COG0639	Diadenosine tetraphosphatase a	5
COG0640	Predicted transcriptional regu	6
COG0642	Signal transduction histidine	25

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0644	Dehydrogenases (flavoproteins)	2
COG0647	Predicted sugar phosphatases o	2
COG0648	Endonuclease IV	2
COG0651	Formate hydrogenlyase subunit	3
COG0654	2-polyprenyl-6-methoxyphenol h	6
COG0655	Multimeric flavodoxin WrbA	2
COG0656	Aldo/keto reductases, related	2
COG0657	Esterase/lipase	4
COG0659	Sulfate permease and related t	3
COG0663	Carbonic anhydrases/acetyltran	2
COG0664	cAMP-binding proteins - catabo	4
COG0665	Glycine/D-amino acid oxidases	6
COG0667	Predicted oxidoreductases (rel	6
COG0668	Small-conductance mechanosensi	3
COG0671	Membrane-associated phospholip	3
COG0673	Predicted dehydrogenases and r	9
COG0674	Pyruvate:ferredoxin oxidoreduc	2
COG0675	Transposase and inactivated de	3
COG0678	Peroxiredoxin	2
COG0681	Signal peptidase I	2
COG0683	ABC-type branched-chain amino	2
COG0685	5,10-methylenetetrahydrofolate	2
COG0688	Phosphatidylserine decarboxyla	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0693	Putative intracellular proteas	3
COG0697	Permeases of the drug/metaboli	15
COG0701	Predicted permeases	2
COG0702	Predicted nucleoside-diphospha	3
COG0704	Phosphate uptake regulator	2
COG0705	Uncharacterized membrane prote	2
COG0714	MoxR-like ATPases	5
COG0715	ABC-type nitrate/sulfonate/bic	3
COG0717	Deoxycytidine deaminase	2
COG0719	ABC-type transport system invo	3
COG0720	6-pyruvoyl-tetrahydropterin sy	2
COG0723	Rieske Fe-S protein	2
COG0725	ABC-type molybdate transport s	2
COG0726	Predicted xylanase/chitin deac	3
COG0730	Predicted permeases	4
COG0731	Fe-S oxidoreductases	2
COG0732	Restriction endonuclease S sub	2
COG0735	Fe ²⁺ /Zn ²⁺ uptake regulation pr	4
COG0737	5'-nucleotidase/2',3'-cyclic p	2
COG0739	Membrane proteins related to m	3
COG0740	Protease subunit of ATP-depend	2
COG0741	Soluble lytic murein transglyc	3
COG0744	Membrane carboxypeptidase (pen	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0745	Response regulators consisting	9
COG0746	Molybdopterin-guanine dinucleo	2
COG0747	ABC-type dipeptide transport s	6
COG0750	Predicted membrane-associated	3
COG0753	Catalase	2
COG0760	Parvulin-like peptidyl-prolyl	3
COG0762	Predicted integral membrane pr	2
COG0764	3-hydroxymyristoyl/3-hydroxyde	2
COG0765	ABC-type amino acid transport	4
COG0766	UDP-N-acetylglucosamine enolpy	2
COG0768	Cell division protein FtsI/pen	2
COG0769	UDP-N-acetylmuramyl tripeptide	2
COG0772	Bacterial cell division membra	2
COG0773	UDP-N-acetylmuramate-alanine l	2
COG0775	Nucleoside phosphorylase	2
COG0776	Bacterial nucleoid DNA-binding	3
COG0778	Nitroreductase	7
COG0782	Transcription elongation facto	2
COG0784	FOG: CheY-like receiver	13
COG0789	Predicted transcriptional regu	8
COG0790	FOG: TPR repeat, SEL1 subfamil	2
COG0791	Cell wall-associated hydrolase	3
COG0793	Periplasmic protease	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG0794	Predicted sugar phosphate isom	2
COG0795	Predicted permeases	2
COG0803	ABC-type metal ion transport s	2
COG0811	Biopolymer transport proteins	2
COG0824	Predicted thioesterase	3
COG0826	Collagenase and related protea	2
COG0834	ABC-type amino acid transport/	9
COG0835	Chemotaxis signal transduction	2
COG0840	Methyl-accepting chemotaxis pr	7
COG0841	Cation/multidrug efflux pump	5
COG0842	ABC-type multidrug transport s	2
COG0843	Heme/copper-type cytochrome/qu	3
COG0845	Membrane-fusion protein	10
COG0848	Biopolymer transport protein	3
COG0850	Septum formation inhibitor	2
COG0859	ADP-heptose:LPS heptosyltransf	2
COG0861	Membrane protein TerC, possibl	3
COG0863	DNA modification methylase	3
COG1011	Predicted hydrolase (HAD super	3
COG1012	NAD-dependent aldehyde dehydro	15
COG1013	Pyruvate:ferredoxin oxidoreduc	2
COG1014	Pyruvate:ferredoxin oxidoreduc	5
COG1018	Flavodoxin reductases (ferredo	4

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1020	Non-ribosomal peptide syntheta	3
COG1024	Enoyl-CoA hydratase/carnithine	8
COG1028	Dehydrogenases with different	32
COG1032	Fe-S oxidoreductase	6
COG1051	ADP-ribose pyrophosphatase	3
COG1052	Lactate dehydrogenase and rela	3
COG1053	Succinate dehydrogenase/fumara	2
COG1060	Thiamine biosynthesis enzyme T	4
COG1061	DNA or RNA helicases of superf	2
COG1063	Threonine dehydrogenase and re	5
COG1064	Zn-dependent alcohol dehydroge	2
COG1070	Sugar (pentulose and hexulose)	3
COG1073	Hydrolases of the alpha/beta s	5
COG1076	DnaJ-domain-containing protein	2
COG1077	Actin-like ATPase involved in	2
COG1082	Sugar phosphate isomerases/epi	6
COG1098	Predicted RNA binding protein	2
COG1100	GTPase SAR1 and related small	2
COG1108	ABC-type Mn ²⁺ /Zn ²⁺ transport s	2
COG1109	Phosphomannomutase	3
COG1112	Superfamily I DNA and RNA heli	2
COG1115	Na ⁺ /alanine symporter	2
COG1116	ABC-type nitrate/sulfonate/bic	4

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1120	ABC-type cobalamin/Fe ³⁺ -sidero	5
COG1121	ABC-type Mn/Zn transport syste	2
COG1122	ABC-type cobalt transport syst	3
COG1126	ABC-type polar amino acid tran	2
COG1129	ABC-type sugar transport syste	2
COG1131	ABC-type multidrug transport s	14
COG1132	ABC-type multidrug transport s	8
COG1136	ABC-type antimicrobial peptide	4
COG1145	Ferredoxin	4
COG1149	MinD superfamily P-loop ATPase	2
COG1161	Predicted GTPases	2
COG1167	Transcriptional regulators con	4
COG1173	ABC-type dipeptide/oligopeptid	6
COG1175	ABC-type sugar transport syste	7
COG1178	ABC-type Fe ³⁺ transport system	2
COG1180	Pyruvate-formate lyase-activat	2
COG1181	D-alanine-D-alanine ligase and	2
COG1186	Protein chain release factor B	2
COG1187	16S rRNA uridine-516 pseudouri	3
COG1188	Ribosome-associated heat shock	2
COG1191	DNA-directed RNA polymerase sp	2
COG1192	ATPases involved in chromosome	2
COG1199	Rad3-related DNA helicases	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1205	Distinct helicase family with	2
COG1208	Nucleoside-diphosphate-sugar p	4
COG1215	Glycosyltransferases, probably	3
COG1216	Predicted glycosyltransferases	2
COG1226	Kef-type K ⁺ transport systems,	4
COG1228	Imidazolonepropionase and rela	3
COG1233	Phytoene dehydrogenase and rel	3
COG1234	Metal-dependent hydrolases of	3
COG1235	Metal-dependent hydrolases of	3
COG1236	Predicted exonuclease of the b	3
COG1238	Predicted membrane protein	2
COG1239	Mg-chelatase subunit ChII	2
COG1249	Pyruvate/2-oxoglutarate dehydr	5
COG1250	3-hydroxyacyl-CoA dehydrogenas	6
COG1252	NADH dehydrogenase, FAD-contai	2
COG1253	Hemolysins and related protein	3
COG1263	Phosphotransferase system IIC	4
COG1264	Phosphotransferase system IIB	4
COG1267	Phosphatidylglycerophosphatase	2
COG1275	Tellurite resistance protein a	2
COG1277	ABC-type transport system invo	2
COG1278	Cold shock proteins	2
COG1280	Putative threonine efflux prot	7

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1284	Uncharacterized conserved prot	3
COG1286	Uncharacterized membrane prote	2
COG1302	Uncharacterized protein conser	2
COG1304	L-lactate dehydrogenase (FMN-d	2
COG1305	Transglutaminase-like enzymes,	2
COG1309	Transcriptional regulator	24
COG1335	Amidases related to nicotinami	3
COG1344	Flagellin and related hook-ass	2
COG1349	Transcriptional regulators of	7
COG1353	Predicted hydrolase of the HD	2
COG1354	Uncharacterized conserved prot	2
COG1357	Uncharacterized low-complexity	3
COG1358	Ribosomal protein HS6-type (S1	2
COG1360	Flagellar motor protein	2
COG1363	Cellulase M and related protei	2
COG1373	Predicted ATPase (AAA+ superfa	3
COG1376	Uncharacterized protein conser	2
COG1381	Recombinational DNA repair pro	2
COG1392	Phosphate transport regulator	2
COG1393	Arsenate reductase and related	3
COG1396	Predicted transcriptional regu	5
COG1409	Predicted phosphohydrolases	2
COG1414	Transcriptional regulator	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1418	Predicted HD superfamily hydro	3
COG1432	Uncharacterized conserved prot	2
COG1433	Uncharacterized conserved prot	2
COG1434	Uncharacterized conserved prot	2
COG1454	Alcohol dehydrogenase, class I	2
COG1463	ABC-type transport system invo	2
COG1466	DNA polymerase III, delta subu	2
COG1472	Beta-glucosidase-related glyco	2
COG1473	Metal-dependent amidase/aminoa	2
COG1475	Predicted transcriptional regu	2
COG1476	Predicted transcriptional regu	2
COG1482	Phosphomannose isomerase	2
COG1484	DNA replication protein	3
COG1486	Alpha-galactosidases/6-phospho	2
COG1488	Nicotinic acid phosphoribosylt	2
COG1502	Phosphatidylserine/phosphatidy	4
COG1506	Dipeptidyl aminopeptidases/acy	3
COG1522	Transcriptional regulators	15
COG1529	Aerobic-type carbon monoxide d	4
COG1530	Ribonucleases G and E	2
COG1534	Predicted RNA-binding protein	2
COG1538	Outer membrane protein	2
COG1545	Predicted nucleic-acid-binding	3

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1560	Lauroyl/myristoyl acyltransfer	2
COG1562	Phytoene/squalene synthetase	2
COG1573	Uracil-DNA glycosylase	3
COG1589	Cell division septal protein	2
COG1593	TRAP-type C4-dicarboxylate tra	2
COG1595	DNA-directed RNA polymerase sp	5
COG1597	Sphingosine kinase and enzymes	2
COG1605	Chorismate mutase	2
COG1607	Acyl-CoA hydrolase	2
COG1609	Transcriptional regulators	7
COG1611	Predicted Rossmann fold nucleo	3
COG1612	Uncharacterized protein requir	2
COG1622	Heme/copper-type cytochrome/qu	3
COG1624	Uncharacterized conserved prot	2
COG1629	Outer membrane receptor protei	5
COG1638	TRAP-type C4-dicarboxylate tra	3
COG1651	Protein-disulfide isomerase	3
COG1653	ABC-type sugar transport syste	8
COG1668	ABC-type Na ⁺ efflux pump, perm	2
COG1670	Acetyltransferases, including	9
COG1674	DNA segregation ATPase FtsK/Sp	2
COG1680	Beta-lactamase class C and oth	3
COG1694	Predicted pyrophosphatase	3

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1695	Predicted transcriptional regu	5
COG1721	Uncharacterized conserved prot	4
COG1725	Predicted transcriptional regu	2
COG1733	Predicted transcriptional regu	3
COG1737	Transcriptional regulators	2
COG1741	Pirin-related protein	2
COG1748	Saccharopine dehydrogenase and	2
COG1757	Na ⁺ /H ⁺ antiporter	2
COG1758	DNA-directed RNA polymerase, s	3
COG1759	ATP-utilizing enzymes of ATP-g	2
COG1760	L-serine deaminase	2
COG1762	Phosphotransferase system mann	4
COG1764	Predicted redox protein, regul	2
COG1765	Predicted redox protein, regul	2
COG1767	Triphosphoribosyl-dephospho-Co	2
COG1776	Chemotaxis protein CheC, inhib	2
COG1788	Acyl CoA:acetate/3-ketoacid Co	2
COG1793	ATP-dependent DNA ligase	2
COG1801	Uncharacterized conserved prot	2
COG1804	Predicted acyl-CoA transferase	3
COG1814	Uncharacterized membrane prote	2
COG1826	Sec-independent protein secret	4
COG1840	ABC-type Fe ³⁺ transport system	3

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG1841	Ribosomal protein L30/L7E	2
COG1845	Heme/copper-type cytochrome/qu	3
COG1846	Transcriptional regulators	12
COG1853	Conserved protein/domain typic	3
COG1879	ABC-type sugar transport syste	2
COG1898	dTDP-4-dehydrorhamnose 3,5-epi	2
COG1902	NADH:flavin oxidoreductases, O	3
COG1917	Uncharacterized conserved prot	3
COG1932	Phosphoserine aminotransferase	2
COG1940	Transcriptional regulator/suga	5
COG1959	Predicted transcriptional regu	4
COG1960	Acyl-CoA dehydrogenases	12
COG1961	Site-specific recombinases, DN	2
COG1975	Xanthine and CO dehydrogenases	2
COG1994	Zn-dependent proteases	2
COG2017	Galactose mutarotase and relat	4
COG2025	Electron transfer flavoprotein	2
COG2030	Acyl dehydratase	3
COG2038	NaMN:DMB phosphoribosyltransfe	2
COG2041	Sulfite oxidase and related en	3
COG2050	Uncharacterized protein, possi	5
COG2057	Acyl CoA:acetate/3-ketoacid Co	2
COG2062	Phosphohistidine phosphatase S	2

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG2070	Dioxygenases related to 2-nitr	2
COG2072	Predicted flavoprotein involve	2
COG2076	Membrane transporters of catio	2
COG2084	3-hydroxyisobutyrate dehydroge	2
COG2085	Predicted dinucleotide-binding	2
COG2111	Multisubunit Na ⁺ /H ⁺ antiporter	2
COG2124	Cytochrome P450	2
COG2128	Uncharacterized conserved prot	2
COG2133	Glucose/sorbosone dehydrogenas	2
COG2141	Coenzyme F420-dependent N5,N10	6
COG2146	Ferredoxin subunits of nitrite	2
COG2171	Tetrahydrodipicolinate N-succi	2
COG2186	Transcriptional regulators	3
COG2188	Transcriptional regulators	6
COG2190	Phosphotransferase system IIA	2
COG2195	Di- and tripeptidases	2
COG2197	Response regulator containing	4
COG2199	FOG: GGDEF domain	5
COG2200	FOG: EAL domain	4
COG2202	FOG: PAS/PAC domain	8
COG2204	Response regulator containing	2
COG2207	AraC-type DNA-binding domain-c	12
COG2217	Cation transport ATPase	4

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG2220	Predicted Zn-dependent hydrola	4
COG2222	Predicted phosphosugar isomera	2
COG2223	Nitrate/nitrite transporter	2
COG2225	Malate synthase	2
COG2233	Xanthine/uracil permeases	2
COG2236	Predicted phosphoribosyltransf	2
COG2244	Membrane protein involved in t	2
COG2249	Putative NADPH-quinone reducta	3
COG2259	Predicted membrane protein	2
COG2265	SAM-dependent methyltransferas	2
COG2267	Lysophospholipase	2
COG2303	Choline dehydrogenase and rela	4
COG2309	Leucyl aminopeptidase (aminope	2
COG2357	Uncharacterized protein conser	2
COG2376	Dihydroxyacetone kinase	2
COG2390	Transcriptional regulator, con	2
COG2391	Predicted transporter componen	4
COG2513	PEP phosphonomutase and relate	2
COG2518	Protein-L-isoaspartate carboxy	2
COG2520	Predicted methyltransferase	2
COG2606	Uncharacterized conserved prot	2
COG2710	Nitrogenase molybdenum-iron pr	3
COG2723	Beta-glucosidase/6-phospho-bet	3

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG2764	Uncharacterized protein conser	2
COG2801	Transposase and inactivated de	5
COG2814	Arabinose efflux permease	3
COG2872	Predicted metal-dependent hydr	2
COG2885	Outer membrane protein and rel	3
COG2890	Methylase of polypeptide chain	2
COG2908	Uncharacterized protein conser	2
COG2931	RTX toxins and related Ca ²⁺ -bi	2
COG3000	Sterol desaturase	2
COG3119	Arylsulfatase A and related en	2
COG3145	Alkylated DNA repair protein	2
COG3173	Predicted aminoglycoside phosp	2
COG3177	Uncharacterized conserved prot	2
COG3288	NAD/NADP transhydrogenase alph	2
COG3386	Gluconolactonase	2
COG3387	Glucoamylase and related glyco	2
COG3408	Glycogen debranching enzyme	2
COG3485	Protocatechuate 3,4-dioxygenas	2
COG3547	Transposase and inactivated de	2
COG3599	Cell division initiation prote	2
COG3711	Transcriptional antiterminator	2
COG3839	ABC-type sugar transport syste	2
COG3842	ABC-type spermidine/putrescine	3

Continúa ...

Cuadro 22: COGs finales (continuación)

COG	Descripción	Grupos
COG4166	ABC-type oligopeptide transpor	2
COG4177	ABC-type branched-chain amino	2
COG4264	Siderophore synthetase compone	2
COG4585	Signal transduction histidine	3
COG4608	ABC-type oligopeptide transpor	3
COG4638	Phenylpropionate dioxygenase a	3
COG4770	Acetyl/propionyl-CoA carboxyla	2
COG4786	Flagellar basal body rod prote	2
COG4799	Acetyl-CoA carboxylase, carbox	2
COG4948	L-alanine-DL-glutamate epimera	3
COG4974	Site-specific recombinase XerD	2
COG4992	Ornithine/acetylnithine amin	3

Fin tabla

13. Apéndice 5: Cuadro COGs anotados

Cuadro con los COGs anotados por PROSITE (ver 5.2), PFAM (ver 5.1) y manualmente. Los últimos 4 COGs fueron documentados manualmente (ver 5.3).

Cuadro 23: 75 COGs anotados por PROSITE, PFAM. Los últimos 4 COGs fueron documentados manualmente (ver 5.3)

COG	Secuencias semillas	PFAM	PROSITE
COG0079_G1	41	-	PS00599 (3/41 7.31 %)
COG0079_G2	20	-	PS00105 (11/20 55 %)
		-	PS00599 (1/20 5 %)
COG0122_G1	24	PF00730 (24/24 100 %)	-
COG0122_G2	5	PF07934 (5/5 100 %)	-
		PF00730 (3/5 60 %)	-
COG0144_G1	51	PF01189 (49/51 96.07 %)	-
COG0144_G2	5	PF01472 (4/5 80 %)	-
		PF01189 (5/5 100 %)	-
COG0184_G1	31	PF00312 (31/31 100 %)	-
COG0184_G2	13	PF00312 (13/13 100 %)	-
		PF08069 (13/13 100 %)	-
COG0195_G1	29	PF08529 (29/29 100 %)	-
		PF00575 (16/29 55.17 %)	-
		PF00013 (4/29 13.79 %)	-
COG0195_G2	10	PF07650 (10/10 100 %)	-
		PF00013 (1/10 10 %)	-
COG0197_G1	31	PF00252 (31/31 100 %)	PS00586 (83.87 %)
		PF00826 (4/31 12.90 %)	PS00701 (93.54 %)

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG0197_G2	14	PF00826 (14/14 100 %)	PS01257 (14/14 100 %)
		PF00252 (1/17 7.14 %)	-
COG0202_G1	29	PF01000 (29/29 100 %)	-
		PF01193 (29/29 100 %)	-
		PF03118 (28/29 96.55 %)	-
COG0202_G2	15	PF01000 (14/14 100 %)	-
		PF01193 (14/15 93.33 %)	-
COG0231_G1	32	PF09285 (32/32 100 %)	PS01257 (28/32 87.5 %)
		PF01132 (31/32 96.87 %)	-
		PF08207 (29/32 90.6 %)	-
COG0231_G2	12	PF01287 (11/12 91.6 %)	PS00302 (12/12 100 %)
		PF00467 (11/12 91.6 %)	-
		PF01132 (1/12 8.33 %)	-
COG0237_G1	29	PF01121 (29/29 100 %)	-
COG0237_G2	10	PF01583 (4/10 40 %)	-
		PF01202 (4/10 40 %)	-
COG0250_G1	29	PF02357 (29/29 100 %)	PS01014 (22/29 75.86 %)
		PF00467 (26/29 89.65 %)	PS00430 (1/29 3.44 %)
COG0250_G2	10	PF00467 (10/10 100 %)	PS01108 (4/10 40 %)
		PF02357 (2/10 20 %)	-
COG0258_G1	30	PF02739 (30/30 100 %)	-
		PF01367 (30/30 100 %)	-
		PF09281 (1/30 3.33 %)	-

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG0258_G2	12	PF00867 (12/12 100 %)	-
		PF00752 (11/12 91.66 %)	-
		PF01367 (2/12 16.66 %)	-
		PF02151 (1/12 8.33 %)	-
COG0262_G1	26	PF00186 (26/26 100 %)	PS00075 (17/26 65.38 %)
COG0262_G2	11	PF01872 (11/11 100 %)	-
		PF00186 (1/11 9.09 %)	-
COG0288_G1	26	-	PS00705 (14/26 53.84 %)
		-	PS00704 (11/26 42.30 %)
COG0288_G2	5	-	-
COG0294_G1	28	-	PS00793 (23/28 82.14 %)
		-	PS00792 (20/28 71.42 %)
COG0294_G2	7	-	-
COG0331_G1	26	PF00698 (26/26 100 %)	-
COG0331_G2	6	-	-
COG0334_G1	35	-	PS00074 (85.71 %)
COG0334_G2	8	-	-
COG0358_G1	29	PF01751 (29/29 100 %)	-
		PF08275 (29/29 100 %)	-
		PF01807 (29/29 100 %)	-
COG0358_G2	10	PF01751 (10/10 100 %)	-
COG0361_G1	30	PF01176 (30/30 100 %)	-
		PF04689 (25/30 83.33 %)	-

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG0361_G2	13	PF01176 (10/10 100 %)	PS01262 (6/13 46.15 %)
COG0363_G1	19	-	-
COG0363_G2	14	-	PS01161 (8/14 57.14 %)
COG0392_G1	12	PF03706 (12/12 100 %)	-
COG0392_G2	10	-	-
COG0417_G1	20	-	PS00116 (9/20 45 %)
COG0417_G2	5	-	-
COG0433_G1	16	PF01935 (16/16 100 %)	-
COG0433_G2	7	PF05872 (7/7 100 %)	-
COG0448_G1	17 -	PS00809 (11/17 64.70 %)	
		-	PS00808 (10/17 58.82 %)
		-	PS00810 (9/17 52.94 %)
COG0448_G2	6	-	-
COG0464_G1	25	-	PS00674 (22/25 88 %)
COG0464_G2	5	-	-
COG0471_G1	13	PF00939 (13/13 100 %)	-
		PF03600 (5/13 38.46 %)	-
		PF06808 (3/13 23.07 %)	-
COG0471_G2	10	PF02080 (10/10 100 %)	-
		PF03600 (10/10 100 %)	-
		PF00939 (4/10 40 %)	-
COG0471_G3	6	PF03600 (2/6 33.33 %)	-
COG0510_G1	6	PF01633 (6/6 100 %)	-

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
		PF01636 (6/6 100 %)	-
COG0510_G2	PF01636 (5/5 100 %)	-	
COG0516_G1	37	-	-
COG0516_G2	7	-	PS00487 (5/7 71.42 %)
COG0557_G1	28	-	PS01175 (25/28 89.28 %)
COG0557_G2	6	-	PS001175 (1/6 16.66 %)
COG0563_G1	38	-	PS00113 (38/38 100 %)
COG0563_G2	15	-	-
COG0574_G1	37	-	PS00370 (27/37 72.97 %)
		-	PS00742 (26/37 70.27 %)
COG0574_G2	7	-	PS00228 (1/7 14.28 %)
COG0575_G1	32	PF01148 (32/32 100 %)	PS01315 (26/32 81.25 %)
COG0575_G2	11	PF01864 (11/11 100 %)	-
COG0579_G1	16	PF01266 (16/16 100 %)	-
		PF04324 (5/16 31.25 %)	-
COG0579_G2	8	PF06039 (8/8 100 %)	-
		PF01266 (7/8 87.5 %)	-
COG0592_G1	30	PF00712 (29/30 96.6 %)	PS00041 (1/30 3.33 %)
		PF02767 (28/30 93.33 %)	-
		PF02768 (28/30 93.33 %)	-
COG0592_G2	14	PF00705 (14/14 100 %)	PS01251 (12/14 85.71 %)
		PF02747 (14/14 100 %)	PS00293 (12/14 85.71 %)
COG0612_G1	38	-	PS00143 (23/38 60.52 %)

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG0612_G2	7	-	-
COG0612_G3	5	-	-
COG0616_G1	33	PF01343 (33/33 100 %)	-
COG0616_G2	9	PF01972 (9/9 100 %)	-
		PF00574 (3/9 33.33 %)	-
		PF00378 (1/9 11.11 %)	-
COG0655_G1	16	PF00258 (16/16 100 %)	-
		PF03358 (1/16 6.25 %)	-
COG0655_G2	6	PF03358 (6/6 100 %)	-
COG0704_G1	28	PF01895 (28/28 100 %)	-
COG0704_G2	11	PF04014 (11/11 100 %)	-
		PF01895 (11/11 100 %)	-
COG0731_G1	12	PF04055 (12/12 100 %)	-
		PF08608 (12/12 100 %)	-
COG0731_G2	11	PF04055 (10/11 90.90 %)	-
COG0740_G1	33	-	PS00382 (31/33 93.93 %)
		-	PS00381 (27/33 81.81 %)
COG0740_G2	5	-	-
COG0744_G1	37	PF00912 (37/37 100 %)	-
		PF00905 (37/37 100 %)	-
COG0744_G2	7	PF00912 (7/7 100 %)	-
COG0769_G1	28	PF01225 (25/28 89.28 %)	-
		PF02875 (25/28 89.28 %)	-

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
		PF08245 (25/28 89.28 %)	-
COG0769_G2	8	PF08353 (7/8 87.5 %)	-
		PF08245 (8/8 100 %)	-
COG0782_G1	30	-	PS00829 (27/30 90 %)
		-	PS00830 (24/30 80 %)
COG0782_G2	7	-	PS00830 (1/7 14.28 %)
COG0793_G1	28	PF03572 (28/28 100 %)	-
		PF00595 (27/28 96.42 %)	-
COG0793_G2	6	PF03572 (6/6 100 %)	-
		PF00595 (1/6 16.66 %)	-
COG0826_G1	25	-	PS01276 (22/25 88 %)
COG0826_G2	5 -	-	
COG1053_G1	38	-	PS00504 (28/38 73.68 %)
COG1053_G2	5 -	-	
COG1076_G1	7	PF07319 (7/7 100 %)	-
		PF07743 (6/7 85.71 %)	-
COG1076_G2	6	PF07319 (6/6 100 %)	-
		PF07049 (1/6 16.66 %)	-
COG1100_G1	12	PF03029 (12/12 100 %)	-
COG1100_G2	5	PF00071 (5/5 100 %)	-
		PF08477 (5/5 100 %)	-
		PF00025 (4/5 80 %)	-
COG1180_G1	13	-	-

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG1180_G2	13	-	PS01087 (8/13 61.53 %)
COG1181_G1	30	PF07478 (27/30 90 %)	PS00843 (29/30 96.66 %)
		PF01820 (26/30 86.66 %)	PS00844 (29/30 96.66 %)
		PF02222 (20/30 66.66 %)	-
		PF02655 (8/30 26.66 %)	-
		PF08443 (3/30 10 %)	
COG1181_G2	7	PF08443 (5/7 71.42 %)	PS00690 (1/7 14.28 %)
		PF02955 (3/7 42.85 %)	PS00866 (1/7 14.28 %)
		PF02222 (3/7 42.85 %)	-
		PF01071 (2/7 28.57 %)	-
COG1186_G1	24	PF00472 (24/24 100 %)	-
		PF03462 (24/24 100 %)	-
COG1186_G2	6	PF00472 (6/6 100 %)	-
COG1199_G1	17	PF06733 (17/17 100 %)	-
		PF00270 (2/17 11.76 %)	-
COG1199_G2	17	PF00270 (16/17 94.11 %)	-
		PF06733 (1/17 5.88 %)	-
		PF00271 (1/17 5.88 %)	-
COG1205_G1	14	PF00271 (14/14 100 %)	-
		PF00270 (14/14 100 %)	-
		PF04851 (1/14 7.14 %)	-
COG1205_G2	5	PF00270 (5/5 100 %)	-
		Helicase C (1/5 20 %)	-

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG1304_G1	21	-	PS00557 (17/21 80.95 %)
COG1304_G2	20	-	PS00583 (1/20 5 %)
COG1305_G1	11	PF01841 (11/11 100 %)	-
		PF03647 (1/11 9.09 %)	-
COG1305_G2	8	PF08379 (8/8 100 %)	-
		PF01841 (8/8 100 %)	-
COG1488_G1	26	PF04095 (16/26 61.53 %)	-
		PF01729 (15/26 57.69 %)	-
		PF02749 (8/26 30.76 %)	-
COG1488_G2	11	PF04095 (11/11 100 %)	-
COG1530_G1	24	PF00575 (24/24 100 %)	-
COG1530_G2	6	PF04167 (6/6 100 %)	-
COG1622_G1	24	PF00116 (23/24 95.83 %)	-
		PF02790 (10/24 41.66 %)	-
COG1622_G2	10	PF06481 (10/10 100 %)	-
		PF00116 (7/10 70 %)	-
		PF02790 (1/10 10 %)	-
COG1622_G3	5	PF00116 (5/5 100 %)	-
COG1760_G1	23	PF03313 (23/23 100 %)	-
		PF03315 (16/23 69.56 %)	-
COG1760_G2	7	PF03315 (7/7 100 %)	-
		PF01842 (6/7 85.71 %)	-
COG1788_G1	17	-	PS1273 (10/14 71.42 %)

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG1788_G2	7	-	-
COG1793_G1	19	-	PS00697 (17/19 89.47 %)
COG1793_G2	19 -	PS00333 (8/19 42.10 %)	
COG2017_G1	16	-	PS00545 (8/16 50 %)
COG2017_G2	8	-	-
COG2057_G1	13	-	PS01274 (8/13 61.53 %)
COG2057_G2	6	-	PS00696 (2/6 33.33 %)
COG2225_G1	9	-	PS00510 (7/9 77.77 %)
COG2225_G2	7	-	-
COG2236_G1	11	-	PS00103 (5/11 45.45 %)
COG2236_G2	8	-	-
COG2376_G1	13	PF02734 (13/13 100 %)	-
		PF02733 (2/13 15.38 %)	-
COG2376_G2	8	PF02733 (8/8 100 %)	-
		PF02734 (2/8 25 %)	-
COG2513_G1	15	-	PS00161 (12/15 80 %)
COG2513_G2	10 -	-	
COG2518_G1	24	-	PS01279 (20/24 83.33 %)
COG2518_G2	5	-	-
COG2764_G1	10	PF00903 (7/10 70 %)	-
		PF06983 (6/10 60 %)	-
COG2764_G2	8	PF00903 (8/8 100 %)	-
COG3173_G1	7	-	PS00109 (4/7 57.14 %)

Continúa ...

Cuadro 23: COGs anotados (continuación)

COG	Sec. semillas	PFAM	PROSITE
COG3173_G2	5	-	-
COG3288_G1	10	-	PS00837 (6/10 60 %)
COG3288_G2	7	-	-
COG3711_G1	7	PF00874 (7/7 100 %)	-
		PF08279 (6/7 85.71 %)	-
		PF05043 (3/7 42.85 %)	-
		PF08220 (1/7 14.28 %) -	-
COG3711_G2	6	PF00874 (6/6 100 %)	-
		PF03123 (6/6 100 %)	-
COG0095			
COG0558			
COG1393			
COG0147			

Fin tabla

14. Apéndice 6: División de perfiles Pfam PF04715 y PF01066

Los perfiles Pfam PF04715 (antianilato sintasa) y PF01066 (CDP-alcohol fosfatidiltransferasas) fueron divididos utilizando el enfoque aplicado a los COGs en este trabajo. Las secuencias de los elementos que constituyen cada perfil se obtuvieron del GenBank, la eliminación de redundancia en las secuencias, se hizo utilizando el programa CD-HIT (<http://bioinformatics.org/cd-hit/>). Los hits bidireccionales se construyeron con los mejores hit encontrados por BLAST de cada proteína, además de verificar que las proteínas que presentaban BBTs fueran de familias diferentes.

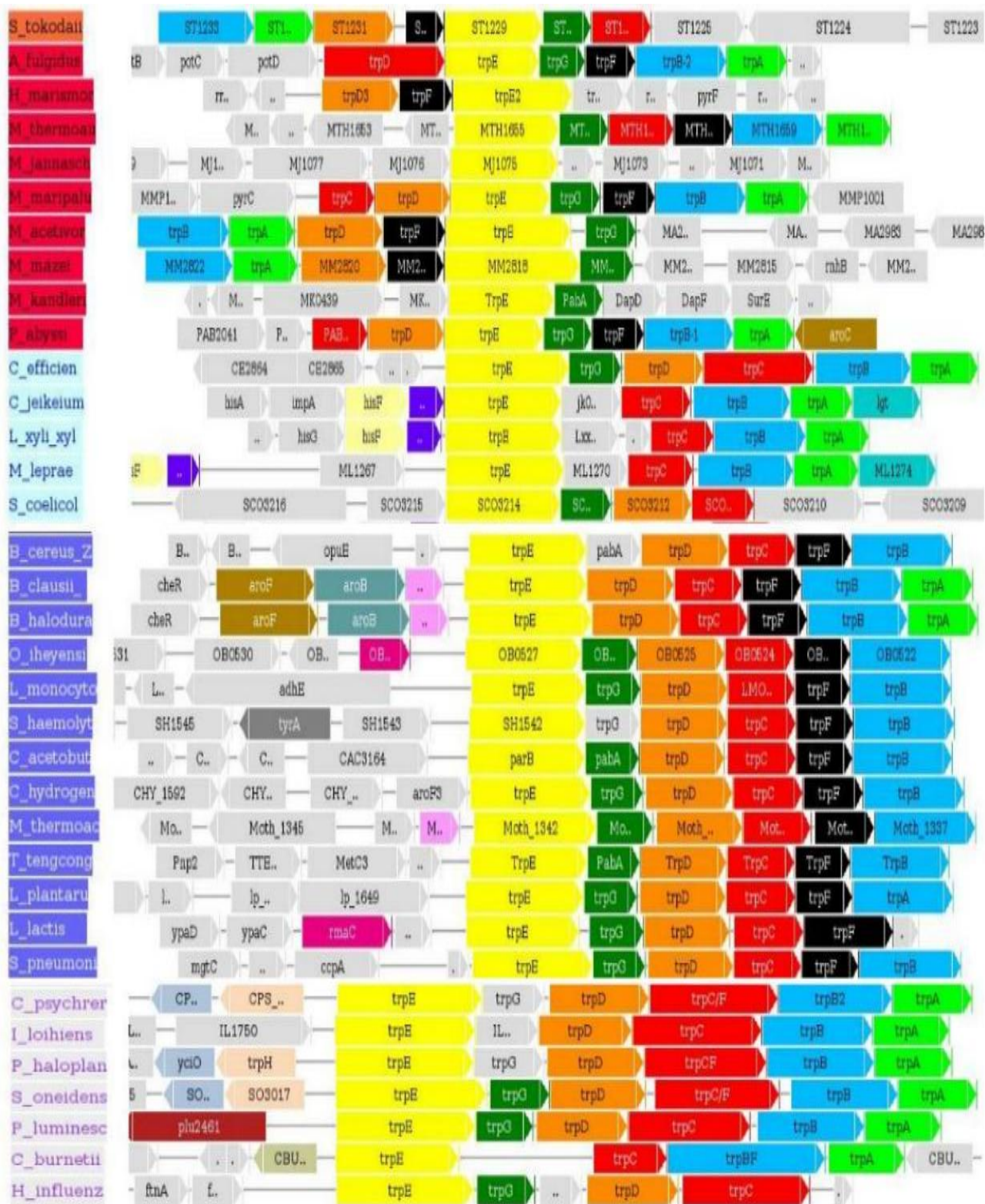


Figura 44: Contexto genómico de algunas secuencias contenidas en el grupo 1, obtenido de la división de Pfam PF04715 (antaranilato sintasa) utilizando la metodología descrita en este trabajo.



Figura 45: Contexto genómico de algunas secuencias contenidas en el grupo 2, obtenido de la división de Pfam PF04715 (antaranilato sintasa) utilizando la metodología descrita en este trabajo.

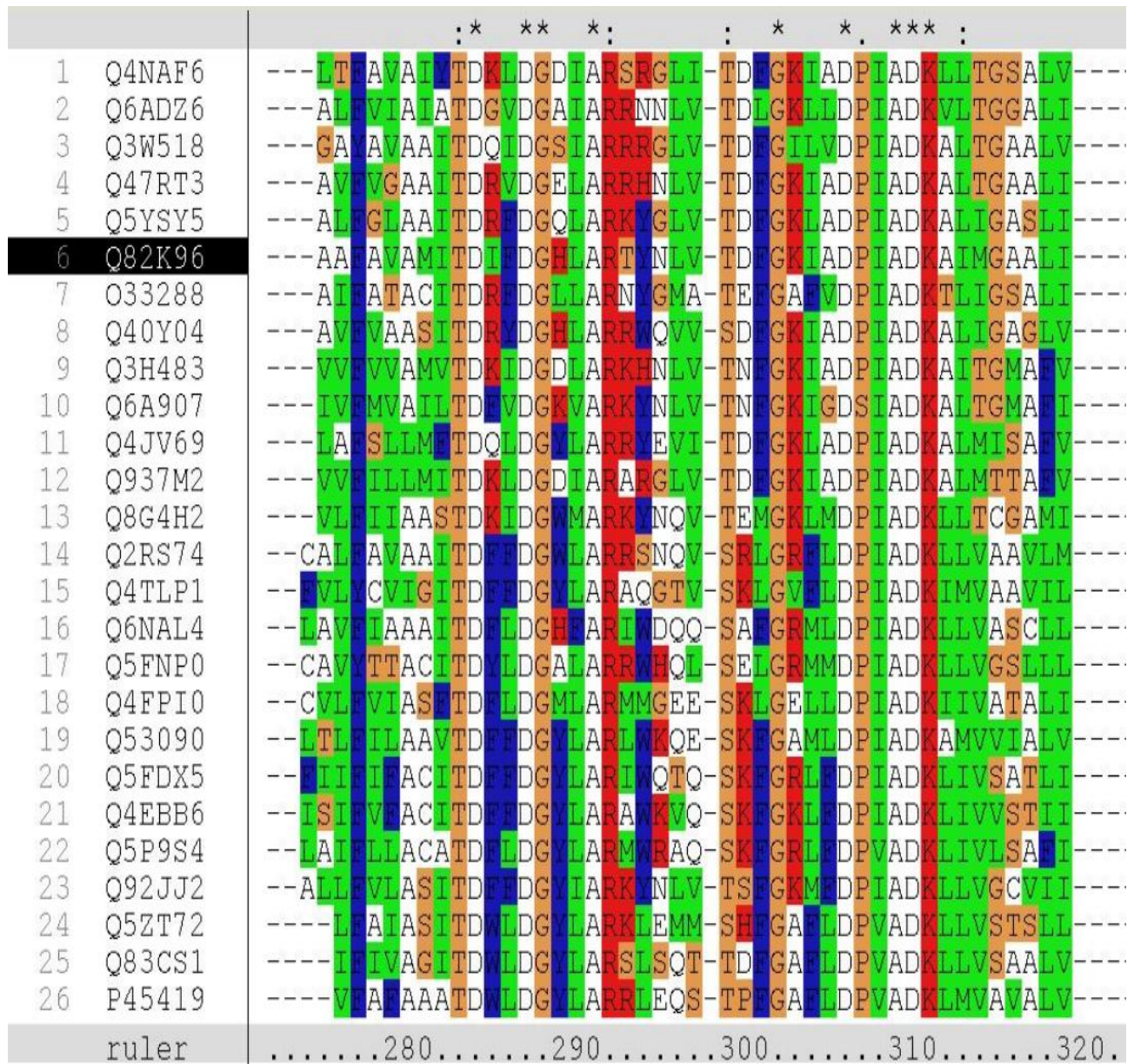


Figura 46: Alineación de algunas secuencias del grupo 1 en que fue dividido el perfil Pfam PF01066 (CDP-alcohol fosfatidiltransferasas), donde se observa el motivo específico de la familia PgsA (ver Sección 5.3.2). La alineación fue hecha con ClustalX.

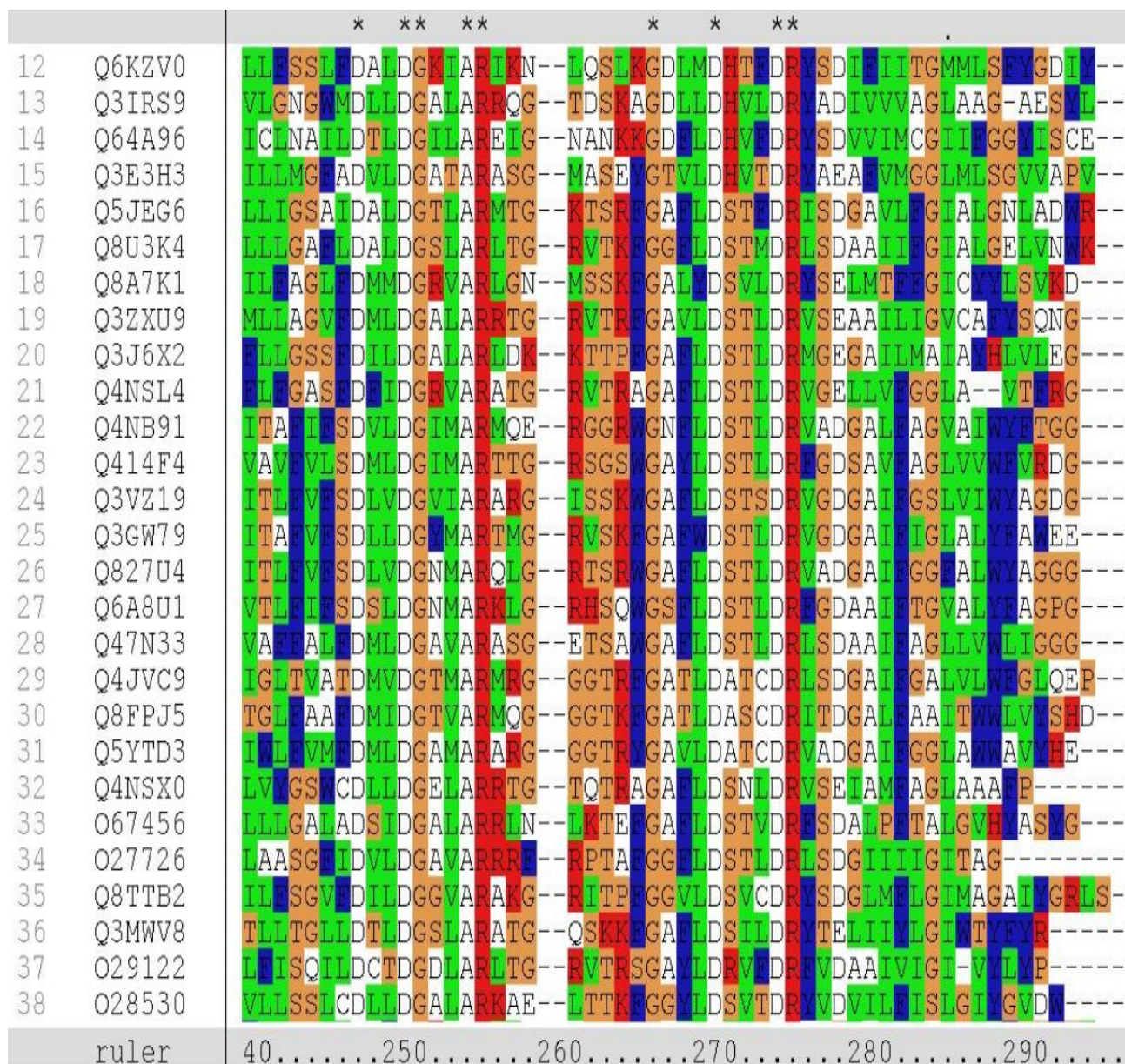


Figura 48: Alineación de algunas secuencias del grupo 3 en que fue dividido el perfil Pfam PF01066 (CDP-alcohol fosfatidiltransferasas), donde se observa el motivo específico de la familia Pis (ver Sección 5.3.2). La alineación fue hecha con ClustalX.

Referencias

- [1] Abhiman, S. and Sonnhammer, E.L.L. 2005. FunShift: a database of function shift analysis on protein subfamilies. *Nucleic Acids Res.* v.33(Database Issue):D197-200.
- [2] Abreu-Goodger, C.; Ontiveros-Palacios, N.; Ciria, R. and Merino, E. 2004. Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends in Genetics.* 20: 475–479.
- [3] Abhiman, S. and Sonnhammer, E.L.L. 2005. Large-scale prediction of function shift in protein families with a focus on enzymatic function. *Proteins: Structure, Function, and Bioinformatics.* 60:758-768.
- [4] Altschul, S.F.; Gish, W.; Miller, W.; Meyers, E. W. and Lipman, D. J. 1990. Basic Local Alignment Search Tool. *J. Mol. Biol.* 215: 403–410.
- [5] Anand, R.; Hoskins, A.A.; Stubbe, J. and Ealick, E.S. 2004. Domain organization of *Salmonella typhimurium* formylglycinamide ribonucleotide amidotransferase revealed by X-ray crystallography. *Biochemistry.* 43: 10328-10342.
- [6] Aravind, L.; Mazumder, R.; Vasudevan, S. and Koonin, V. 2002. Trends in protein evolution inferred from sequence and structure analysis. *Curr. Opin. Struct. Biol.* 12(3): 392-399.
- [7] Bab-Dinitz, E.; Shmueli, H.; Maupin-Furlow, J.; Eichler, J. and Shaanan, B. 2006. *Haloferax volcanii* PitA: an example of functional interaction between the Pfam chlorite dismutase and antibiotic biosynthesis monooxygenase families?. *Bioinformatics.* 22(6): 671-675.
- [8] Bateman, A.; Birney, E.; Eddy, S.R.; Howe, K.L. and Sonnhammer, E.L.L. 2000. The Pfam Protein Families Database. *Nucleic Acid Research.* 28(1): 263-266.
- [9] Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S.R.; Griffiths-Jones, S.; Howe, K.L.; Marshall, M. and Sonnhammer, E.L.L. 2002. The Pfam protein families database. *Nucleic Acid Research.* 30(1):276-280.

- [10] Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E.L.L.; Studholme, D.J.; Yeats, C. and Eddy, S.R. 2004. The Pfam protein families database. *Nucleic Acid Research*. 32. Database Issue:D138-D141.
- [11] Finn, R.D; Mistry, J.; Schuster-Böckler, B.; Griffiths-Jones, S.; Hollich, V.; Lassmann, T; Moxon, S.; Marshall, M.; Khanna, A.; Durbin, R.; Eddy, S.R.; Sonnhammer, E.L.L. and Bateman, A. 2006. Pfam: clans, web tools and services. *Nucleic Acid Research*. 345. Database Issue:D247-D251.
- [12] Breitkreutz, B.J.; Stark, C. and Tyers, M. 2003. Osprey: a network visualization system. *Genome Biol*. 4: R22.
- [13] Chang, S.H.; Heacock, P.N.; Clancey, C.J. and Dowhan, W. 1998. The PEL1 gene (renamed PGS1) encodes 3A the phosphatidylglycerophosphate synthase of *Saccharomyces cerevisiae*. *J. Biol. Chem*. 273(16): 9829-9836.
- [14] Chandonia, J.M. and Brenner, S. 2005. Update on the pfam5000 strategy for selection of structural genomics targets. *Conf Proc IEEE Eng Med Biol Soc*. 1:751-755.
- [15] Choi, I. and Kim, S. 2006. Evolution of protein structural classes and protein sequences families. *PNAS*. 103(38): 14056-14061.
- [16] Choi, I. and Kim, S.H. 2007. Global extent of horizontal gene transfer. *PNAS*. 104(11): 4489-4494.
- [17] Ciria, R.; Abreu-Goodger, C.; Morett, E. and Merino, E. 2004. GeConT: gene context analysis. *Bioinformatics*. 20: 2307-2308.
- [18] Dandekar, T; Snel, B.; Huynen, M. and Bork, P. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*. 23(9):324-328.
- [19] Dopazo, J.; Zanders, E.; Dragoni, I.; Amphlett, G. and Falciani, F. 2001. Methods and approaches in the analysis of gene expression data. *Journal of Immunological Methods*. 250: 93-112.
- [20] Eisen, M.B; Spellman, P.T; Brown, P.O. and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci*. 95(25):14863-14868.

- [21] Enright, A.J.; Van Dongen, S. and Ouzounis, C.A. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* 30(7):1575–1584.
- [22] Fitch, W.M. 2000. Homology a personal view on some of the problems. *TIG*.16(5):227–231.
- [23] Fujiwara, K.; Toma, S.; Okamura-Ikeda, K.; Motokawa, Y.; Nakagawa, A. and Taniguchi, H. 2005. Crystal structure of lipoate-protein ligase A from *Escherichia coli*. *Journal of Biological Chemistry*. 280(39): 33645-33651.
- [24] Galperin, Y.M. and Koonin, V. E. 2000. Who's your neighbor?. New computational approaches for functional genomics. *Nature Biotechnology*. 18: 609-613.
- [25] Ge, H.; Walhout, J.M. and Vidal, M. 2003. Integrating 'omic' information: a bridge between genomics and systems biology. *TIG*. 19(10): 551–560.
- [26] Goodacre, R.; Timmins, M.E.; Burton, R.; Kaderbhai, N.; Woodward, M.A.; Kell, B.D. and Rooney, J.P. 1998. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiol*. 144: 1157-1170.
- [27] Handl, J.; Knowles, J. and Kell, B.D. 2005. Computational cluster validation in post-genomic data analysis. *Bioinformatics*. 21(15): 3201-3212.
- [28] Hannenhalli, S.S. and Russell, R.B. 2000. Analysis and prediction of functional sub-types from protein sequence alignments. *J. Mol. Biol*. 303: 61-76.
- [29] He, Z.; Stigers-Lavoie, K. D.; Bartlett, P. A. and Toney, M. D. 2004. Conservation of mechanism in three chorismate-utilizing enzymes. *J. Am. Chem. Soc*. 126: 2378–2385.
- [30] Hiller, M.; Huse, K.; Platzer, M. and Backofen, R. 2005. Non-EST based prediction of exon skipping and intron retention events using Pfam information. *Nucleic Acid Res*. 33(17): 5611-5621.
- [31] Hoffmann, K.; Bucher, P.; Falquet, L. and Bairoch, A. 1999. The PROSITE database, its status in 1999. 27(1):215-219.

- [32] Huang, S.H.; Chen, Y.H.; Kong, G.; Chen, S.H.; Besemer, I.; Borodovsky, M. and Jong, A. 2001. A novel genetic island of meningitic *Escherichia coli* K1 containing the *ibeA* invasion gene (*GimA*): functional annotation and carbon-source-regulated invasion of human brain microvascular endothelial cells. *Funct Integr Genomics*. 1(5):312-322.
- [33] Hulo, N.; Bairoch, A.; Bulliard, V.; Cerutti, L.; De Castro, E.; Langendijk-Genevaux, P.S.; Pagni, M. and Sigrist C.J.A. 2006. The PROSITE database. *Nucleic Acid Res*. 34. Database Issue: D227-D230.
- [34] Huynen, M.A.; Snel, B.; Late III, W. and Bork, P. 2000. Exploitation of gene context. *Current Opinion in Cell Biology*. 10: 366-370.
- [35] Ikeda, M. 2005. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. *Appl Microbiol Biotechnol*. 69(6):615-626.
- [36] Jain, K.A.; Murty, N. M. and Flynn, J.P. 1999. Data clustering: A review. *ACM Computing Surveys*. 31(3):264-323.
- [37] James, T.Y.; Boulianne, R.P.; Bottoli, A.P.; Granado, J.D.; Aebi, M. and Kües, U. 2002. The *pab1* gene of *Coprinus cinereus* encodes a bifunctional protein for *para*-aminobenzoic acid (PABA) synthesis: implications for the evolution of fused PABA synthases. *J. Basic. Microbiol*. 42: 91-103.
- [38] Kapetanovic, M.I.; Rosenfeld, S. and Izmirlian, G. 2004. Overview of commonly used bioinformatics methods and their applications. *Ann. N.Y. Acad. Sci*. 1020: 10-21.
- [39] Kaplan, N.; Friedlich, M.; Fromer, M. and Linial, M. 2004. A functional hierarchical organization of the protein sequence space. *BMC Bioinformatics*. 5: 196.
- [40] Kim, J.D.; Kim, H.K.; Lee, H.H.; Lee, J-S.; Ha, Y.J.; Yoon, J.H. and Suh, W.S. 2005. Crystal structure of lipocate-protein ligase A bound with the activated intermediate. *Journal of Biological Chemistry*. 280(45): 38081-38089.
- [41] Koonin, E.V.; Wolf, Y.I. and Karev, G.P. 2002. The structure of the protein universe and genome evolution. *Nature*. 420: 218-223

- [42] Koonin, E.V.; Wolf, Y.I. and Aravind, L. 2001. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative genomic approach. *Genome Research*. 11:240–252.
- [43] Koonin, E. V. ; Aravind, L. and Kondrashov, S. 2000. The impact of comparative genomics on our understanding of evolution. *Cell*. 101: 573–576.
- [44] Kwon, K. and Beckett, D. 2000. Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Science*. 9: 1530-1539.
- [45] Lattimore, S.B.; Van Dogen, S. and Crabbe, C.J. 2005. GeneMCL in microarray analysis. *Computational Biology and Chemistry*. 29(5): 354-359.
- [46] Li,L.; Stoeckert, J.C. and Roos, S.D. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13(9): 2178-2189.
- [47] Lockhart, D. J. and Winzler, E. A. 2000. Genomics, gene expression and DNA arrays. *Nature*. 405:827–836.
- [48] Li, H.; Pellegrini, M. and Eisenberg, D. 2005. Detection of parallel functional modules by comparative analysis of genome sequences. *Nat Biotechnol*. 23(2):253-260.
- [49] Lynch, M. and Force, A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics*. 154(1):459-73.
- [50] Maquitman, A.G.; Rechtsteiner, A.; Verspoor, K.; Strauss, C.E. and Rocha, L.M. 2006. Large-scale testing of bibliome informatics using Pfam protein families. *Pac Symp Biocomput*. 76-87.
- [51] Marcotte, E.M. 2000. Computational genetics: finding protein function by nonhomology methods. *Curr. Opin. Struct. Biol*. 10: 359-365.
- [52] Martin,P.; DeMel, S.; Shi, J.; Gladysheva, T.; Gatti, D.L.; Rosen, B.P. and Edwards, B.F. 2001. Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. *Structure*. 9(11): 1071-1081.

- [53] McManus, E.; Luisi, F.B. and Perham, N.R. 2006. Structure of a putative lipoate protein ligase from *Thermoplasma acidophilum* and the mechanism of target selection for post-translational modification. *J. Mol. Biol.* 356: 625-637.
- [54] Messens, J.; Martins, J.C.; Brosens, E.; Van Belle, K.; Jacobs, D.M.; Willen, R. and Wyns, L. 2002. Kinetics and active site dynamics of *Staphylococcus aureus* arsenate reductase. *J. Biol. Inorg. Chem.* 7: 146-156.
- [55] Mindell, D.P. and Meyer, A. 2001. Homology evolving. *Trends in Ecology and Evolution.* 16(8):434-440.
- [56] Mukhopadhyay, R.; Rpsen, B.P.; Phung, L.T. and Silver, S. 2002. Microbial arsenic: from geocycles to genes and enzymes. *FEMS Microbiol. Reviews.* 26: 311-325.
- [57] Nagase, T.; Ishikawa, K.; Nakajima, D.; Ohira, M.; Seki, N.N.; Miyajima, N.N.; Tanaka, A.A.; Kotani, H.H.; Nomura, N.N. and Ohara, O.O. 1997. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA research.* 4(2):141-150.
- [58] Nakano, S.; Küster-Schöck, E.; Grossman, A.D. and Zuber, P. 2003. Spx-dependent global transcriptional control is induced by thiol-specific oxidative stress in *Bacillus subtilis*. *PNAS.* 100(23):13603-13608.
- [59] Nakano, S.; Erwin, K.N.; Ralle, M. and Zuber, P. 2005. Redox-sensitive transcriptional control by a thiol/disulphide switch in the global regulator, Spx. *Mol. Microbiol.* 55(2): 498-510.
- [60] Natale, D.A.; Shankavaram, V.T.; Galperin, M.Y.; Wolf, Y.I.; Aravind, L. and Koonin, E.V. 2000. Towards understanding the first genome sequences of a crenarchaeon by genome annotation using clusters of orthologous groups of protein (COGs). *Genome Biology.* 1: 1-19.
- [61] Narain, A.S. and Madan, M.B. Contextual inference of protein function. In L. Aravind (eds), *Protein Function and Annotation*. Wiley. USA. Archivo PDF disponible en <http://www.mrc-lmb.cam.ac.uk/genomes/madanm/wiley.pdf>

- [62] Newberry, K.J.; Nakano, S.; Zuber, P. and Brennan, G. 2005. Crystal structure of the *Bacillus subtilis* anti-alpha, global transcriptional regulator, Spx, in complex with the α C-terminal of RNA polymerase. 2005. PNAS. 102(44): 15839-15844.
- [63] Overbeek, R.; Fonstein, M.; D'Souza, M.; Pusch, G.D. and Maltsev, N. 1999. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci U S A. 96(6):2896-2901.
- [64] Pamp, S.J., Frees, D., Engelmann, S. Hecker M. and Ingmer, H. 2006. Spx is a global effector impacting stress tolerance and biofilm formation in *Staphylococcus aureus*. J. Bacteriol. 188(13): 4861-4870
- [65] Pandey, A. and Mann, M. 2000. Proteomics to study genes and genomes. Nature. 405: 837-846.
- [66] Pellegrini, M.; Marcotte, E. M.; Thompson, M. J.; Eisenberg, D. and Yeates, T.O. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc. Natl. Acad. Sci. 96: 44285-4288.
- [67] Quackenbush, J. 2001. Computational analysis of microarray data. Nature Review Genetics. 2: 418-427.
- [68] Reche, P.A. 2000. Lipoylating and biotinylating enzymes contain a homologous catalytic module. Protein Science. 9: 1922-1929.
- [69] Remm, M. ; Storm, C. E. V. and Sonnhammer. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol. 314:1041-1052.
- [70] Rosen, B. P. 2002. Biochemistry of arsenic detoxification. FEBS Letters. 529:86-92.
- [71] Sentandreu, M.; ELorza, M.V. and Sentandreu, R. 1997. Isolation of a putative prolyl-tRNA synthetase (CaPRS) gene from *Candida albicans*. Yeast. 13(14):1375-1381.
- [72] Sigrist, C.J.; Cerutti, L.; Hulo, N.; Gattiker, A.; Falquet, L.; Pagni, M.; Bairoch, A. and Bucher, P. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 3(3):265-274.

- [73] Slonim, K.D. 2002. From patterns to pathways: gene expression data analysis comes of age. *Nat. Genet.* 32: 502-508.
- [74] Sohlenkamp, C.; López-Lara, I.M. and Geiger, O. 2003. Biosynthesis of phosphatidylcholine in bacteria. *Progress in Lipid Research.* 42: 115-162.
- [75] Sonnhammer, L.L.E. and Koonin, V.E. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends in Genetics.* 18(12): 619-620.
- [76] Stevens, S.Y.; Hu, W.; Gladysheva, T.; Rosen, B.P.; Zuiderweg, E.R.P. and Lee, L. 1999. Secondary structure and fold homology of the ArsC protein from the *Escherichia coli* arsenic resistance plasmid R773. *Biochemistry.* 38(31):10178-10186.
- [77] Stolz, J.F., Basu, P. and Oremland, R.S. 2002. Microbial transformation of elements: the case of arsenic and selenium. *Int. Microbiol.* 5:201-207.
- [78] Sybesma, W.; Starrenburg, M.; Kleerebezem, M.; Mierau, I.; De Vos, W.M. and Hugenholtz, J. 2003. Increased production of folate by metabolic engineering of *Lactococcus lactis*. *Applied and Environmental Microbiology.* 69(6): 3069-3076.
- [79] Tatusov, R.L.; Galperin, M.Y.; Natale, D.A. and Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acid Research.* 28(1): 33-36.
- [80] Tatusov, R.L.; Koonin, E.V. and Lipman, D.J. 1997. A genomic perspective on protein families. *Science.* 278: 631-637.
- [81] Tringe, G.S.; Von Mering, C.; Kobayashi, A.; Salamov, A.A.; Chen, K.; Chang, W.H.; Podar, M.; Short, M.J.; Mathur, J.E.; Detter, C.J.; Bork, P.; Hugenholtz, P. and Rubin, M. E. 2005. Comparative metagenomics of microbial communities. *Science.* 308: 554-557.
- [82] Uchiyama, I. 2006. Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes. *Nucleic Acids Research.* 34(2): 647-658.
- [83] Van Dogen, S. 2000. Graph Clustering by Flow Simulation. PhD thesis, University of Utrecht. URL:<http://www.library.uu.nl/digiarchief/dip/diss/1895620/inhoud.htm>

- [84] Ward, R. and Durrett, R. 2004. Subfunctionalization: How often does it occur? How long does it take?. *Theoretical Population Biology*. 66(2): 93-100.
- [85] <http://www.ercim.org/publication/Ercim_News/enw42/nieland.jpg>. Consultado en el 2007.
- [86] <<http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?COG1697>>. Consultado en el 2007.
- [87] <<http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?COG2927>>. Consultado en el 2007.
- [88] Xie, D; Lin, X.; Wang, M.; Jiang, Z. and Feng, H. 2005. Using motifs in the prediction of eukaryotic protein subcellular localization. *Conf Proc IEEE Eng Med Biol Soc*. 3:2802-2804.
- [89] Zuber, P. 2004. Spx-RNA polymerase interaction and global transcriptional control during oxidative stress. *J. of Bacteriol*. 186(7): 1911–1918.