

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS  
AVANZADOS DEL INSTITUTO POLITÉCNICO  
NACIONAL

LABORATORIO NACIONAL DE GENÓMICA PARA LA BIODIVERSIDAD

EVOLUCIÓN DE LA FUNCIÓN ENZIMÁTICA MEDIADA POR  
PÉRDIDA DE GENES

Tesis que presenta

**M. EN C. ANA LILIA JUÁREZ-VÁZQUEZ**

Para obtener el grado de

**DOCTOR EN BIOTECNOLOGÍA DE PLANTAS**

Director de Tesis

**FRANCISCO BARONA GÓMEZ**

Irapuato, Guanajuato, México, 2017



El presente trabajo de investigación se desarrolló en el Laboratorio de Evolución de la Diversidad Metabólica del Laboratorio Nacional de Genómica para la Biodiversidad, LANGEBIO ó, Unidad de Genómica Avanzada, UGA. Bajo la dirección del Dr. Francisco Barona Gómez, y con la asesoría interna de los doctores: Rafael Montiel Duarte, Cei Abreu-Goodger, Mauricio Carrillo-Tripp todos profesores titulares del Laboratorio Nacional de Genómica para la Biodiversidad y del Dr. Luis Delaye Arredondo, profesor titular del Departamento de Ingeniería Genética del Cinvestav Unidad Irapuato y la asesoría externa de la Dra. Rosario A. Muñoz Clares, profesora titular del Departamento de Química de la Facultad de Química de la Universidad Nacional Autónoma de México, UNAM.

El trabajo se realizó de Septiembre de 2011 a Febrero de 2016.

Se agradece al **Consejo Nacional de Ciencia y Tecnología, CONACYT** por la beca de estudios de doctorado con el número de registro **234761** y número de apoyo **324991** otorgada durante el periodo de Septiembre de 2011 a Agosto de 2015.

*You cannot hope to build a better world without improving the individuals. To that end, each of us must work for his own improvement and, at the same time, share a general responsibility for all humanity, our particular duty being to aid those to whom we think we can be most useful. M.C.*

## **Agradecimientos**

En primer lugar, agradezco al Dr. Francisco Barona-Gómez por haberme dado la oportunidad de integrarme a su grupo de trabajo.

En segundo lugar, agradezco a mi comité tutorial por todas las críticas hechas a este trabajo.

Dr. Cei Abreu Goodger

Dr. Rafael Montiel Duarte

Dr. Mauricio Carrillo-Tripp

Dr. Luis Delaye Arredondo y

Dra. Rosario Muñoz Clares

Especialmente, agradezco a la Dra. Rosario Muñoz Clares porque a pesar de sus múltiples actividades, siempre estuvo dispuesta a escuchar los avances de éste proyecto.

En tercer lugar, agradezco a mis compañeros y colaboradores los Doctores: Lianet Noda, Ernesto Verduzco, Kari Verdel y Nelly Sélem, quienes me involucraron en el entendimiento de la evolución de proteínas compartiéndome su pasión y conocimientos.

En cuarto lugar, agradezco a mis compañeros del galerón 7: Pablo, Kari G, Paulina, Dany, Miguel Vallebuena, Brenda, Erasmo, José Luis, Luis Yañez, Sandra, Lorena, Noé, Cuauhtémoc, Mitzi, Ximena, etc. Especialmente a Dany por las exquisitas discusiones.

A los auxiliares Hilda Ramos, Mariana Vallejo, Romina Berni y Corina Díaz, por el apoyo para el material utilizado en ésta tesis y a Christian Martínez, Víctor Villa, Julián Santoyo y Gabriel Jiménez por su soporte en bioinformática. Especialmente a Víctor Villa por su colaboración en el diseño de clarisse.

A Araceli Fernández por su ayuda en el uso del clúster MAZORKA.

A Erika Cruz y Sergio Campos del laboratorio del Dr. Alexander de Luna, por prestarme a Evo para los experimentos de adecuación.

A Roberto Álvarez y Cesaré Ovando del laboratorio del Dr. Cei Abreu por su ayuda con el uso de R.

A Mayra Flores del cuarto de medios por su buena disposición y eficiencia al tener siempre el material y el laboratorio de enseñanza a tiempo.

Al Dr. Sean Rovito por sus valiosos consejos y comentarios a mi trabajo.

A los chicos de metamorfosis del CIMAT, en especial a Mariana y Lucy por darme la oportunidad de compartir mi pasión con los chicos del estado de Guanajuato.

A mis amigos: Brenda, Adris (Adriana), Abue (Eric), Alina, Luis, Ale y Janet por apoyarme siempre, por compartir sus alegrías y tristezas y porque a pesar de mis múltiples ausencias siempre estuvieron cerca de mí.

Por último y no por ser menos importante, agradezco al universo por haberme colocado como un miembro más de la familia Juárez-Vázquez, una familia maravillosa, de la cual me siento muy orgullosa de formar parte. A mis padres, que me han apoyado. A mis hermanos que son las personas más increíbles del mundo, no sólo por soportarme si no, por ser como son. En especial, agradezco a mi hermana Janet, quien siempre me ha procurado y por su apoyo económico. Sin su ayuda ésta tesis tal vez nunca se hubiera terminado. A mi cachorro, quien siempre me anima a seguir adelante y aunque parezca increíble fue el primero en leer ésta tesis ¡¡¡Los amo familia!!!

## Resumen

Los genomas bacterianos están en constante cambio debido a la pérdida y ganancia de genes. Se ha propuesto que la evolución de la especificidad por sustrato de las enzimas, mediada por la promiscuidad enzimática, precede a la adquisición de un nuevo gen mediante Transferencia Horizontal de Genes o duplicación génica. Recientemente, se ha visto que la pérdida de genes es un mecanismo que ha permitido la evolución y adaptación de los genomas bacterianos. Sin embargo, una de las preguntas pendientes tiene que ver con ¿cómo éstas pérdidas de genes afectan la evolución de la especificidad por sustrato?, ¿qué pasa con aquellas enzimas bifuncionales que se retienen?, ¿será posible que la pérdida de genes actúe como una fuerza que permita la especialización o sub-funcionalización enzimática? ¿Qué mecanismos moleculares están involucrados en éste proceso? ¿Qué tipo de presión de selección actúa sobre estas enzimas durante dichos procesos?

Esta tesis postula que el uso de la filogenómica bacteriana, como marco conceptual, nos permitirá detectar episodios evolutivos en donde se pierdan genes pero, se retengan enzimas. Para ello, nos enfocaremos en un modelo evolutivo en donde convergen las rutas de biosíntesis de L-Histidina y L-Triptófano, la isomerasa tipo  $(\beta\alpha)_8$  llamada PriA presente en la familia *Actinomycetaceae*, y especies de *Chlamydia*. Esta enzima parece ser el modelo idóneo, para tratar de responder las preguntas antes planteadas debido a que puede caer en una “trampa evolutiva” impuesta por la dinámica genómica, es decir, la pérdida de genes. El análisis de estos episodios evolutivos nos dió la oportunidad de entender el efecto a nivel de la especificidad por sustrato de la pérdida de genes, lo que fue caracterizable a nivel experimental.

## Abstract

Bacterial genomes are constantly changing due to gene gain and loss. It has been proposed that the evolution of substrate specificity of enzymes, mediated by enzymatic promiscuity, precedes to the acquisition of new genes by Horizontal gene transfer or gene duplication. Recently, it has been seen that gene loss is a mechanism that has allowed the adaptation of bacterial genomes. However, how these gene losses affect the evolution of substrate specificity? What about those bi-functional enzymes that are retained, which are the molecular mechanisms by which the enzymes become specialized? What kind of selection pressure act during gene loss? These questions have remained unclear due to the dearth of enzyme models with specific features.

This thesis postulates the use of bacterial phylogenomics, as a conceptual framework, to detect evolutionary episodes where genes are lost, but enzymes are retained. To do this, we will focus on an evolutionary model where the L-Histidine and L-Tryptophan biosynthesis pathways converge, the ( $\beta\alpha$ ) 8 isomerase called, PriA present in the family Actinomycetaceae and Chlamydia species. This enzyme seems to be the ideal model, to try to answer the aforementioned questions because it can fall into an "evolutionary trap" imposed by the genomic dynamics, *i.e.*, gene loss. The analysis of these evolutionary episodes gave us the opportunity to understand the effect of gene loss under the evolution of substrate specificity, which was characterizable at the experimental level.



<b>INTRODUCCIÓN .....</b>	<b>6</b>
<b>1.1 Dinámica genómica en bacterias mediada por pérdida y ganancia de genes.....</b>	<b>7</b>
<b>1.1.1 Mecanismos que promueven la ganancia de genes .....</b>	<b>8</b>
1.1.1.1 Duplicación génica.....	8
1.1.1.2 Transferencia Horizontal de genes.....	9
1.1.1.3 Modelo evolutivo para el surgimiento de funciones mediante duplicación génica.....	10
<b>1.1.2 Mecanismos que promueven la pérdida de genes.....</b>	<b>12</b>
1.1.2.1 Re-arreglos del genoma, pérdida e inactivación de genes.....	13
1.1.2.2 Mutación, como mecanismo de inactivación de genes y su efecto en organismos con genomas pequeños.....	13
<b>1.2 Evolución del metabolismo bacteriano .....</b>	<b>15</b>
1.2.1 Evolución del metabolismo bacteriano mediado por pérdida y ganancia de genes .....	16
1.2.1.1 Evolución de las rutas de biosíntesis de aminoácidos a través de ganancia de genes.....	18
1.2.1.2 Evolución de las rutas de biosíntesis de aminoácidos a través de pérdida de genes .....	19
1.2.1.3 Evolución de la biosíntesis de L-Histidina y L-Triptófano .....	21
<b>1.3. Estudio de la evolución de proteínas desde una perspectiva postgenómica.....</b>	<b>23</b>
1.3.1 Estudios sobre evolución de una función y conectividad metabólica.....	24
1.3.2 Evolución de la función mediada por la promiscuidad enzimática .....	25
1.3.3 Restricciones biofísicas sobre la evolución de la función.....	26
1.3.3.1 Modelos de evolución de la función que integran adecuación y estabilidad.....	27
1.3.4 El papel de la Epistásis en la evolución de la función.....	28
<b>2. Antecedentes .....</b>	<b>31</b>
<b>2.1 Historia evolutiva del filo Actinobacteria.....</b>	<b>31</b>
2.1.1 Actinobacterias bajo dinámica genómica de pérdida de genes: la familia Actinomycetaceae .....	33
<b>2.2 Convergencia de las rutas de biosíntesis de L-Histidina y L-Triptófano en Actinobacteria .....</b>	<b>35</b>
2.2.1 Intermediarios evolutivos entre HisA y PriA.....	37
2.2.2 Relación estructura función de PriA .....	41
2.2.3 Dinámica genómica y la evolución de subfamilias de PRIA.....	46
2.2.3.1 Co-evolución de PriA y el gen trpF, subfamilia HisAp.....	46
2.2.3.2. Evolución de la especificidad por sustrato de PriA promovida por THG: subfamilia subHisA .....	49
<b>3. Justificación .....</b>	<b>51</b>
<b>4. Hipótesis.....</b>	<b>52</b>
<b>5. Objetivos.....</b>	<b>53</b>
<b>5.1. Objetivo general.....</b>	<b>53</b>
5.1.1. Objetivos específicos .....	53
<b>6. MATERIALES Y MÉTODOS .....</b>	<b>55</b>
<b>6. 1. Métodos bioinformáticos.....</b>	<b>55</b>
6.1.1 Secuenciación, ensamble y anotación de genomas .....	55
6.1.1.1. Secuenciación y análisis de lecturas (reads).....	55

6.1.1.2. Ensamble de genomas.....	55
6.1.1.3. Anotación de genomas.....	56
6.1.2 Obtención del genoma núcleo .....	56
6.1.3 Estimación de tasas de sustitución en secuencias de proteínas .....	65
6.1.3.1 Alineamiento de nucleótidos.....	65
6.1.3.2. Cálculo de las tasas de sustitución utilizando codeml.....	66
6.1.3.3. Análisis de máxima verosimilitud (LRT).....	66
Hipótesis nula .....	67
Hipótesis alternativa .....	67
6.1.4 Análisis filogenómicos .....	68
6.1.4.1 Análisis filogenómico de secuencias concatenadas .....	69
6.1.5. Análisis del contexto genómico.....	69
6.1.6 Pruebas estadísticas para detectar diferencias entre Linajes .....	69
6.1.7. Análisis de pérdidas y ganancias de genes.....	70
<b>6.2 Métodos de biología molecular y bioquímica .....</b>	<b>70</b>
6.2.1 Clonación de genes homólogos de priA .....	70
6.2.2 Expresión heteróloga de homólogos de PriA.....	72
6.2.3 Purificación de homólogos de PriA.....	72
6.2.3.1 Cromatografía de afinidad a níquel .....	72
6.2.3.2 Purificación por exclusión molecular.....	73
6.2.4 Ensayos enzimáticos acoplados para detectar actividad de PRA y ProFAR isomerasa en los homólogos de PriA.....	75
6.2.4.1 Purificación de las enzimas HisF, TrpC y TrpD .....	75
6.2.4.2 Determinación de los parámetros cinéticos para la actividad de ProFAR isomerasa.....	76
6.2.4.3 determinación de los parámetros cinéticos para la actividad de PRA isomerasa .....	77
<b>6.3 Métodos de biología estructural .....</b>	<b>79</b>
6.3.1 Cristalografía de rayos X, para la elucidación estructural de la PriA de <i>A. urogenitalis</i> .....	79
6.3.1.1 Obtención de cristales .....	79
6.3.1.2 Colección de datos.....	80
6.3.1.3 Resolución de la estructura y afinamiento.....	80
6.3.2 Construcción de modelos por homología de PriAs, subHisA2s y subTrpFs.....	81
<b>6.4 Métodos de Microbiología.....</b>	<b>81</b>
6.4.1 Ensayos de complementación de las actividades HisA y TrpF en mutantes de <i>E. coli</i> hisA y trpF menos.....	81
6.4.2 Construcción de la mutante de priA en <i>A. oris</i> MG-1 .....	82
6.4.3 Ensayos de complementación de las actividades hisA y trpF utilizando utilizando la mutante en <i>A. oris</i> MG-1 $\Delta$ priA .....	83
6.4.4 Caracterización fenotípica de la cepa <i>A. oris</i> MG1 .....	83
6.4.5 Extracción de DNA genómico de <i>A. oris</i> MG1, <i>A. oris</i> $\Delta$ priA y la cepa UTB-8.....	85
6.4.6 Aislamiento de cepas de <i>Actinomyces</i> provenientes de muestras orales.....	85
<b>7. Resultados y Discusión .....</b>	<b>87</b>
<b>7.1 Resolución filogenómica de Actinobacterias basales .....</b>	<b>87</b>
<b>7.2 Muestreo bucal y obtención de nuevas Actinobacterias.....</b>	<b>93</b>
7.2.1 Secuenciación genómica y análisis de las cepas <i>A. oris</i> MG-1 y UTB18 provenientes del género <i>Actinomyces</i> .....	99
<b>7.3 Resolución filogenómica de la familia <i>Actinomycetaceae</i> .....</b>	<b>100</b>
<b>7.4 Cálculo de las tasas de sustitución nucleotídica entre los Linajes.....</b>	<b>105</b>

<b>7.5 Tamaño del genoma y contenido de (G+C) de los linajes evolutivos.....</b>	<b>107</b>
<b>7.6 Diferencias en el número de secuencias codificantes entre los linajes.....</b>	<b>112</b>
<b>7.7 Evolución del metabolismo de la familia <i>Actinomyceataceae</i> .....</b>	<b>115</b>
7.7.1 Análisis de pérdidas y ganancias de familias de proteínas en Actinomycetaceae .....	115
7.7.2 Familias de proteínas específicas al Linaje II .....	117
7.7.3 Diferencias en las rutas de biosíntesis de aminoácidos entre los distintos Linajes .....	123
7.7.3.1 Relación entre la pérdida de biosíntesis de aminoácidos y la proporción de aminoácidos presentes en las secuencias de proteínas.....	127
7.7.4 Esencialidad de las rutas de biosíntesis de L-histidina y L-triptófano en <i>Actinomyces</i>	128
7.7.5 Análisis de co-ocurrencia de los genes his y trp .....	133
7.7.6 Asociaciones funcionales de los ortólogos de PriA basados en la co-ocurrencia.....	137
<b>7.8 Presiones de selección detectadas en los ortólogos de PriA .....</b>	<b>138</b>
<b>7.9 Caracterización funcional de enzimas retenidas: nuevas subfamilias en la biosíntesis de L-histidina y L-triptófano en <i>Actinomycetaceae</i> .....</b>	<b>140</b>
7.9.1 Caracterización in vivo de ortólogos de PriA provenientes de los tres escenarios en cepas de <i>E. coli</i> hisA- y trpF- .....	140
7.9.2 Caracterización in vivo de los homólogos de PriA en <i>A. oris</i> MG-1 .....	144
<b>7.9.3 Caracterización <i>in vitro</i> de los homólogos de PriA.....</b>	<b>146</b>
7.9.4 Evolución de la función enzimática de la sub-familia PriA (Hipótesis 1) .....	148
7.9.5 Evolución de la función enzimática de la sub-familia SubHisA2 (Hipótesis H2).....	149
7.9.6 Evolución de la función enzimática de la subfamilia SubTrpF (hipótesis H3).....	150
<b>7.10 Análisis estructural DE la evolución de la especificidad por sustrato en los ortólogos de PriA. ....</b>	<b>151</b>
7.10.1 Determinación de estructuras cristalográficas de PriAs de <i>Actinomyces</i> .....	151
7.10.2 Diferencias a nivel de secuencia y estructura entre PriA y SubHisA2.....	155
7.10.3 Diferencias a nivel de secuencia y estructura entre PriA y SubTrpF.....	158
<b>9. Conclusiones .....</b>	<b>160</b>
<b>10. Perspectivas .....</b>	<b>163</b>
<b>11. RESUMEN DE PUBLICACIONES RELACIONADAS Y CONTRIBUCIONES.....</b>	<b>164</b>

# INTRODUCCIÓN

La era postgenómica trajo consigo un aumento en la eficiencia de los métodos de secuenciación, así como la necesidad de crear algoritmos para analizar los resultados de dichos métodos. La fuerte demanda en dichos métodos y análisis dió como resultado la reducción de los costos de secuenciación. Esto permitió obtener información completa de los genomas de muchos organismos, facilitando el llevar a cabo estudios filogenéticos, de genómica comparada y, hasta de reconstrucciones metabólicas *in silico* que nos han ayudado a entender los mecanismos por los cuales los organismos se adaptan a diferentes condiciones ambientales. Así mismo, a medida que se han realizado estudios de genómica comparada, nos hemos dado cuenta que existe una enorme diversidad de genes y funciones entre organismos. Estas diferencias a nivel de funciones están asociadas a las expansiones, mediadas por la Duplicación Génica y la Transferencia Horizontal de Genes (THG) (Wagner 2008), y por pérdidas de función asociadas a la fijación de mutaciones sin sentido o deletéreas que en algunos casos son resultado de la deriva genética (*genetic drift*)<sup>1</sup> (McCutcheon & Moran 2011). Sólo por citar un ejemplo, dentro del filo Actinobacteria, el cual comprende organismos con alto contenido de (G+C), existen géneros que presentan una enorme diversidad debido a ésta dinámica genómica, formando nuevos linajes bacterianos (Zhao, *et al.* 2014). Por otra parte, a pesar de que se ha reconocido que la expansión y reducción en las familias de proteínas depende de la pérdida y ganancia de genes, los mecanismos moleculares por los cuales se generan nuevas familias o subfamilias de proteínas están poco entendidos. En parte, por la falta de metodologías que nos permitan detectar ganancias y pérdidas de genes y modelos de enzimas. Recientemente, la filogenómica se ha comenzado a usar para detectar dichos escenarios (Albalat y Cañestro 2016) y con ello hemos comenzado a responder preguntas como: ¿cuál es la relevancia de

---

<sup>1</sup> La deriva génica propone que los cambios en la frecuencia alélica ocurren al azar, sin dirección, por lo que la deriva conduce a la fijación de unos alelos y a la pérdida de otros, en poblaciones pequeñas.

perder un gen para ciertos filos?, ¿cómo es el camino que se sigue para pasar de ser un gen/función esencial a dispensable, y finalmente perderse?, ¿la pérdida de genes es un camino efectivo para la adaptación? ¿Qué pasa con las enzimas que se retienen en alguna ruta metabólica durante la pérdida de genes?, ¿puede la pérdida de genes dirigir la evolución de la función enzimática?

Este trabajo de tesis se centra en tratar de responder algunas de las preguntas antes mencionadas. Para poner en contexto dichas preguntas, a continuación revisaremos los procesos por los cuales evolucionan los genomas, centrados en la pérdida de genes y seguidos por la evolución de las rutas metabólicas y enzimas. Además, analizamos algunos de los modelos que se han utilizado para tratar de entender la evolución de la función enzimática y, finalmente describiremos el modelo de una enzima bi-funcional que nos ha permitido estudiar el efecto de la pérdida de ciertos genes sobre la evolución de la especificidad por sustrato. Cabe mencionar que para tratar de entender la evolución de la especificidad por sustrato mediada por pérdida de genes, se integran disciplinas como la filogenómica, bioquímica, biología estructural y evolución molecular, mismas que son introducidas en este sentido.

## 1.1 DINÁMICA GENÓMICA EN BACTERIAS MEDIADA POR PÉRDIDA Y GANANCIA DE GENES

Como se mencionó anteriormente, la era pangenómica ha permitido analizar la diversificación de los genomas bacterianos. A medida que se analizan más genomas bacterianos, se ha observado que sólo una pequeña parte de la información que contienen esta conservada en todas las bacterias. A éste conjunto de genes se le conoce como genoma núcleo, la otra parte del genoma, comprende lo que se conoce como el genoma flexible, el cual esta compuesto por aquellos genes que son únicos a ciertos fila, familias, géneros ó especies. Estas diferencias en el genoma flexible, son el resultado de las pérdidas y ganancias de genes que ocurren en los genomas. Dicha dinámica, esta dirigida por mecanismos que han permitido la expansión y reducción del metabolismo debido a que diferentes presiones de selección han actuado ya sea fijando o eliminando a los genes.

## 1.1.1 MECANISMOS QUE PROMUEVEN LA GANANCIA DE GENES

### 1.1.1.1 DUPLICACIÓN GÉNICA

Como se mencionó anteriormente uno de los eventos que permite la adquisición de una función enzimática nueva es la duplicación génica. Así, en 1970, Ohno propuso que cuando ocurre un proceso de duplicación génica, el gen original sigue estando sujeto a la selección purificante, mientras que la copia se libera de la presión de selección purificante, lo que le permite acumular mutaciones al azar, las cuales en muchos casos resultan en la adquisición de una nueva función (neo-funcionalización). Estos eventos han sido detectados frecuentemente en los genomas bacterianos, y por muchos años se ha propuesto que la duplicación génica es el mecanismo principal para la adquisición de una nueva función, los genes duplicados se conocen como parálogos.

Recientemente, se ha observado que cuando ocurre la duplicación génica bajo condiciones no selectivas, los organismos presentan un crecimiento más lento en comparación con aquellos que no presentan la duplicación, lo que sugiere que las duplicaciones no son favorecedoras para la adecuación de los organismos. Ésto implica que un gen duplicado sólo puede ser mantenido en el genoma cuando, existe un aumento en la dosis génica ó es favorecedor durante condiciones selectivas (Reams *et al.* 2010). El incremento en la dosis génica, juega un papel importante para la neo-funcionalización ya que, la retención del duplicado sin costo metabólico abre una ventana de oportunidad para la adquisición de una nueva función *i.e.* por la baja divergencia evolutiva entre el duplicado y el original tal cual propone Ohno, el gen duplicado puede fijarse y subsecuentemente neo-funcionalizarse ó sub-funcionalizarse<sup>2</sup>, es decir adquiere una función diferente a la ancestral ó pierde una de las funciones ancestrales, respectivamente.

---

<sup>2</sup> El modelo de sub-funcionalización parte de la existencia de una enzima con más de una función, que después de un proceso de duplicación, cada copia se especializa para realizar sólo una de las actividades ancestrales. Durante este proceso la selección actúa positivamente.

Si bien, hasta el momento se ha propuesto que la duplicación, seguida por la sub o neofuncionalización, es uno de los principales mecanismos por los cuales las enzimas pueden cambiar ó adquirir una nueva función, dichos han sido poco explorados y caracterizados, (Xe y Zhang 2005). A pesar que tampoco se ha explorado el impacto que tiene la inserción de un gen duplicado en la topología de la red metabólica a la cual se integran, se han utilizado simulaciones *in silico* de las interacciones proteína-proteína de genes los duplicados y, se ha observado que, después del proceso de duplicación, las interacciones originales se mantienen, a medida que el gen se sub o neo-funcionaliza, se van perdiendo dichas interacciones ancestrales y se ganan nuevas interacciones (Evlampiev y Isambert 2008). De modo interesante, a pesar de la frecuencia de estos eventos, los mecanismos por los cuales ocurren las duplicaciones génicas están poco entendidas.

#### 1.1.1.2 TRANSFERENCIA HORIZONTAL DE GENES

Otro evento que permite la adquisición e inserción de material genético radicalmente nuevo a un genoma es la THG, los genes adquiridos por este tipo de eventos se conocen como genes xenólogos. Contrario a la duplicación, los mecanismos por los cuales ocurre una transferencia horizontal se conocen ampliamente y son la transducción, conjugación y transformación (Thomas y Nielsen 2005), dichos mecanismos no se explicarán a detalle dado que no son el objetivo central de esta tesis.

En bacterias la THG es diversificadora ya que un gen que es adquirido por este mecanismo aumenta la probabilidad de alterar la estructura del genoma, la red regulatoria del organismo receptor, en algunos casos dirige la expansión de familias de proteínas dando cómo resultado el surgimiento de rutas metabólicas no existentes en el organismo receptor (Iwasaki & Takagi 2009; Treangen & Rocha 2011). En este contexto, es plausible suponer que la selección fija el gen transferido cuando se produce un fenotipo ventajoso, en donde el gen adquirido por THG causa alteraciones mínimas a la red genética a la cual se integra, y lo elimina cuando el gen *per se* provoca muchas alteraciones a la red genética y metabólica a la cual se integra (Klassen 2009).

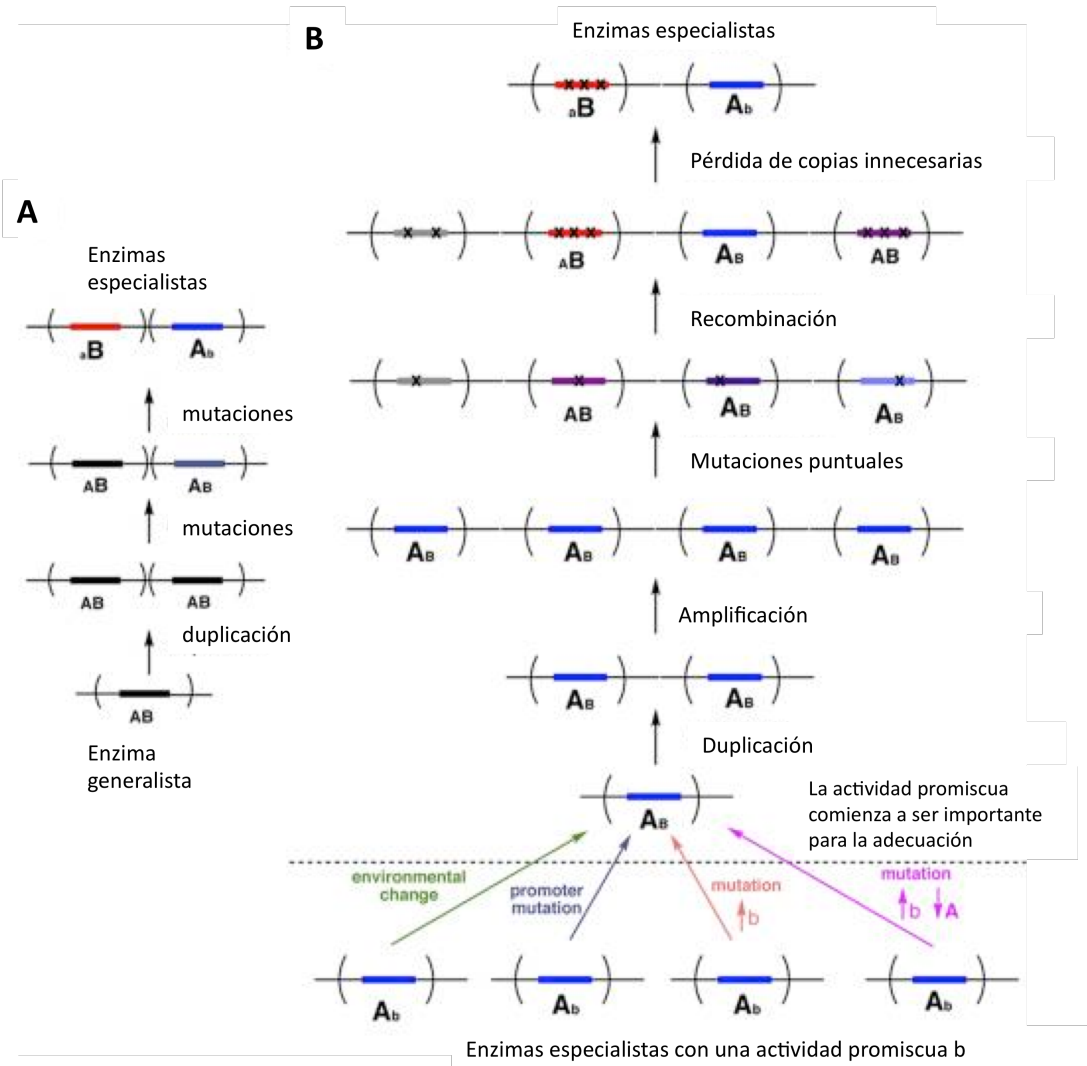
Hasta este punto es claro que la THG y la selección son las fuerzas que dirigen la evolución del metabolismo bacteriano y el surgimiento de nuevas funciones enzimáticas (Treangen & Rocha 2011). Sin embargo, son pocos los estudios que han tomado en cuenta los tipos de presión selectiva que operan para que se retenga un gen adquirido por THG y el efecto que pudiera tener esta inserción sobre las enzimas que forman parte de la ruta metabólica a la cual se integra dicho gen. Un estudio filogenético reciente identificó que durante un evento de THG, algunas enzimas del organismo receptor están bajo un régimen de relajación de la selección, lo que les permite explorar el espacio de secuencia para cambiar su especificidad por sustrato y con ello catalizar reacciones que involucran sustratos análogos que se relacionan con la inserción del gen nuevo a una ruta (Klassen 2009). Este tipo de estudios, dejan claro que para poder entender un proceso evolutivo es necesario analizar los diferentes tipos de selección que operan sobre los componentes de la ruta, para determinar si la existencia de enzimas promiscuas al sustrato en podrían ayudar a la inserción del gen nuevo. Desde este panorama queda clara la importancia de la promiscuidad enzimática como materia prima para la evolución de las rutas metabólicas y de la función enzimática.

#### 1.1.1.3 MODELO EVOLUTIVO PARA EL SURGIMIENTO DE FUNCIONES MEDIANTE DUPLICACIÓN GÉNICA

Como se describió en la primera sección, el surgimiento de nuevas funciones, puede ocurrir en bacterias mediante THG o a partir de actividades promiscuas de enzimas especialistas que se duplican y posteriormente divergen. Durante la duplicación génica, la selección natural actúa manteniendo la copia para después enaltecer una función nueva. Sin embargo, como ya se discutió mantener una copia tiene un costo en la adecuación, por lo que se han propuesto modelos que tratan de explicar el mantenimiento de un gen duplicado. Uno de ellos es el modelo IAD (de las siglas en inglés Innovation, amplification and Divergence), el cual postula que el gen ancestral tiene una actividad secundaria “débil” (innovación), es decir es promiscuo, cuando ocurre un cambio en las



condiciones, la actividad promiscua es benéfica entonces, la selección favorece un incremento en la dosis génica (amplificación), lo que resulta en dos copias del gen ancestral. El incremento en el número de copias provee de materia prima para mutaciones benéficas, las cuales podrían acumularse (divergencia) en las copias. Cualquier copia mejorada se puede amplificar aún más, mientras que las copias menos funcionales, incluyendo el gen parental, se pueden perder. En última instancia, esto resulta en una duplicación de genes en el que una copia del gen codifica la actividad ancestral y otra copia codifica para una nueva actividad (Nasvall *et al.* 2012)(Figura 1). De modo interesante, este modelo ha sido probado utilizando a la enzima ancestral HisA (EC 5.3.1.16), *N*-[5'-fosforribosil]-formimino-5-aminoimidazol carboxiamida ribonucleotido (ProFAR) isomerasa, involucrada en la biosíntesis de histidina, la cual se sabe tiene una actividad promiscua para TrpF (EC 5.3.1.24), *N*-[5'-fosforribosil]-antranilato (PRA) isomerasa, actividad involucrada en la biosíntesis de triptófano (Planch *et al.* 2016).



**Figura 1.** Evolución de la función enzimática medida por duplicación y divergencia ancestral del gen. A) El punto de inicio es una enzima generalista en la cual ambas actividades son importantes para la función. B) El punto de inicio para la optimización de una función es una enzima promiscua (Tomado de Copley 2012).

### 1.1.2 MECANISMOS QUE PROMUEVEN LA PÉRDIDA DE GENES

La pérdida de genes preexistentes durante la evolución de los bacterianos, es un evento que permite la especialización de dichos genomas y el surgimiento de nuevos linajes evolutivos (McCutcheon y Moran 2012; Albalat y Cañestro 2016). Si bien, se conocen dos mecanismos por los cuales pueden ocurrir las pérdidas de genes, el tipo de presión de selección que opera sobre estos eventos no esta

muy claro. En el primer mecanismo, la pérdida puede deberse a un evento mutacional abrupto, como un re-arreglo dentro del cromosoma bacteriano el cual es regulado por elementos de inserción. En el segundo, la pérdida del gen puede ser consecuencia de un proceso lento debido a la acumulación de mutaciones durante un proceso de pseudogenización (Cooper y Lenski 2000, Albalat y Cañestro 2016).

#### 1.1.2.1 RE-ARREGLOS DEL GENOMA, PÉRDIDA E INACTIVACIÓN DE GENES

Durante la evolución del genoma, algunas pérdidas e inactivación de genes, ocurren mediante re-arreglos en ciertas regiones del genoma en donde se encuentran algunos elementos de expropiación. Algunos elementos de expropiación son los llamados elementos-mediadores ó de inserción IS (por sus siglas en inglés de, Insertion). La inactivación de un gen, se produce cuando un elemento de expropiación es insertado en alguna parte del genoma, si esta inactivación favorece la adecuación del organismo positivamente, dicha inserción se mantiene inactivando o des-regulando al gen. Si la pérdida de la actividad es benéfica bajo condiciones de selección, se puede considerar restaurar la actividad mediante una reversión de la inserción. Tales revertantes tendrían una ventaja y la selección dependiente de la frecuencia, podría establecer una fase o estado meta-estable. Sin embargo, si a priori la reversión comienza a acumular mutaciones (como segundo mecanismo de pérdida) la deriva genética podría inactivar permanentemente el gen (Zinser et al. 2003).

#### 1.1.2.2 MUTACIÓN, COMO MECANISMO DE INACTIVACIÓN DE GENES Y SU EFECTO EN ORGANISMOS CON GENOMAS PEQUEÑOS

Como se mencionó anteriormente, la acumulación de mutaciones es el segundo mecanismo por el cual se puede llegar a perder un gen. La mayoría de las mutaciones son deletéreas o tienen un efecto neutral en la adecuación de los organismos y son comúnmente contra-seleccionadas y eliminadas rápidamente de las población. Sin embargo, en bacterias simbiotes y patógenas que

constantemente experimentan cuellos de botella, debido a la deriva genética que experimentan (McCutcheon y Moran 2012), se ha propuesto que el efecto de las mutaciones es menos nocivo. Puesto que, se ha encontrado que la selección purificante es menos severa en las proteínas por lo que podrían ser más susceptibles al desplegamiento, los organismos que presentan este tipo de proteínas suelen expresar más genes implicados en el plegamiento para así preservar la estabilidad de muchas de estas proteínas y contrarrestar el efecto de la relajación en la presión de selección (McCutcheon y Moran 2012). Así mismo, se ha propuesto que un incremento en la relajación de la selección purificante, da como resultado un incremento en la *robustness* mutacional<sup>3</sup> *i.e.* aumenta la tolerancia a las mutaciones deletéreas en proteínas (Wagner 2005).

Si bien, la inactivación de una función casi siempre se debe a procesos de especialización y adaptación de los organismos, poco se sabe sobre el efecto de la acumulación de estas mutaciones sobre la especificidad de sustrato, y la estabilidad de las proteínas y es que, recientemente, se ha puesto en duda el nulo efecto de la deriva genética y la acumulación de mutaciones sobre la estructura-función de las proteínas. Mediante experimentos de evolución dirigida, se observó que la mayoría de las mutaciones “neutras” tienen un efecto biofísico en la estructura de las proteínas, lo que sugiere que las restricciones estructurales son componentes que limitan la evolución. Este tema requiere de un estudio más detallado con polimorfismos<sup>4</sup> dentro de una población para determinar cómo la pérdida de la estabilidad afecta la adecuación de los organismos (De Pristo et al. 2005, Liberics et al. 2012). Además un buen modelo para determinar el efecto de ésta aparente acumulación de mutaciones neutras podría ser el estudio de proteínas que provienen de organismos que están experimentando una relajación en la presión de selección como aquellos que están en proceso de reducción del genoma.

---

<sup>3</sup> Describe el grado en el cual el fenotipo de un organismo se mantiene constante a pesar de la acumulación de mutaciones.

Finalmente, podemos decir que tal cual hemos escrito en ésta sección, la sub o neo funcionalización mediada por eventos de ganancia de genes como la duplicación o la THG tienen un efecto negativo en la adecuación de los organismos. Mientras que no existen ejemplos hasta el momento de subfuncionalización o neo-funcionalización mediante pérdida de genes. Este panorama deja claro que:

- a) sería más plausible que los eventos de sub o neo-funcionalización ocurran sin un evento de duplicación o THG previo. Si esto es así, es altamente probable que se puedan observar eventos de sub o neo funcionalización durante eventos de pérdidas de genes partiendo de enzimas promiscuas,
- b) en estos casos, la sub o neo funcionalización tendrá un efecto negativo en la adecuación del organismo? y c) ¿la presión de selección que actúa en eventos de sub o neo-funcionalización será la misma?

## 1.2 EVOLUCIÓN DEL METABOLISMO BACTERIANO

La secuenciación de genomas además de permitir observar cómo evolucionan las rutas metabólicas, también ha permitido el análisis de las funciones a nivel genómico y evolutivo. Con ello, se ha podido deducir que el último ancestro común universal (LUCA, por sus siglas en inglés Last Universal Common Ancestor) tenía muy pocos genes que codificaban para la mayoría de las funciones biológicas celulares (Ouzounis *et al.* 2006). A medida que los organismos fueron evolucionando, el repertorio de proteínas presentes en el último ancestro se expandió por duplicación, nuevos plegamientos fueron surgiendo, la fusión de dominios dió origen a estructuras más complejas y emergieron familias y subfamilias por medio de la divergencia de los genes ancestrales en donde ocurrió una neo o sub funcionalización., y así es como se explica la adaptación y evolución de los organismos (Copley 2012).

Por muchos años se ha especulado que las enzimas más antiguas fueron enzimas generalistas capaces de catalizar transformaciones químicas particulares debido a su amplia especificidad por sustrato. Así la duplicación de genes que codificaban para enzimas generalistas permitió la división de las funciones ancestrales mediante el surgimiento de enzimas altamente especializadas. La

evolución de enzimas generalistas a enzimas especialistas promovió una regulación más precisa de las rutas metabólicas (Jensen 1976). Recientemente, se ha propuesto que el punto de surgimiento de una nueva actividad en las enzimas modernas ha sido el resultado de las actividades promiscuas presentes en las enzimas especialistas. En algunos casos, una enzima nueva no puede contribuir con la adecuación, así que múltiples enzimas promiscuas son reclutadas para formar una ruta nueva y llevar a cabo un nuevo proceso metabólico distinto al ancestral (Copley 2012).

### 1.2.1 EVOLUCIÓN DEL METABOLISMO BACTERIANO MEDIADO POR PÉRDIDA Y GANANCIA DE GENES

Como se mencionó en la sección 1.1, el genoma bacteriano evoluciona a través de varios mecanismos como son la duplicación génica, la THG, las inserciones y la acumulación de mutaciones, dichos mecanismos permiten ganar o perder genes. Actualmente, existe un enorme esfuerzo por tratar de entender cómo estas funciones que se ganan y pierden se inter-relacionan con las funciones pre-existentes en el organismo (sección 1.1.1.2). Algunos acercamientos para determinar esto han sido el uso de experimentos de alto rendimiento, *high throughput*, y la construcción de modelos a escala genómica, GSMR del inglés, Genome-Scale Metabolic Reconstructions. El objetivo de las GSMR ha sido describir el metabolismo completo de un organismo partiendo de las funciones celulares y en el conocimiento bioquímico de las enzimas que participan en las rutas metabólicas descritas (Edwards y Palsson 1999). Generalmente, una GSMR incluye anotación semiautomática, la cual puede ser curada manualmente a partir de datos tomados de la literatura. Posteriormente, se hace una auto-completación de genes que no están anotados pero que se supone son necesarios para que los organismos pueden crecer en ciertas condiciones ambientales (González *et al.* 2008; Oberhardt *et al.* 2008; Chavali *et al.* 2008). Los análisis de las rutas y de auto-completación además de elucidar la estequiometría de ciertas reacciones, han permitido determinar rutas más eficientes para la producción de ciertos metabolitos. A este nivel las

GSMRs pueden ser útiles para comparar el contenido metabólico entre genomas de especies relacionadas y determinar el efecto de ganar o perder una función metabólica sobre la topología de la red metabólica. Recientemente, las GSMR han sido utilizadas para tratar de entender la evolución del metabolismo bacteriano mediado por pérdida y ganancia de genes. Este tipo de estudios, usualmente toman en cuenta los estados de reconstrucción ancestral (ASR, siglas en inglés de Ancestral State Reconstruction). Generalmente estas inferencias se basan en la construcción de filogenias y en la construcción de una matriz de caracteres (funciones), así se obtienen los nodos ancestrales y se reconstruyen o se buscan los caracteres presentes en los organismos actuales, en este caso se buscan las familias de genes. Mediante este tipo de estudios se ha encontrado que la evolución del metabolismo bacteriano está íntimamente ligada con los estilos de vida de cada organismo más que con la ancestría. La adaptación de las bacterias a diferentes condiciones desencadena como ya se ha planteado una pérdida y ganancia considerable de genes, lo que influye sobre las funciones y reacciones presentes en el organismo (Bryant *et al.* 2013). Por ejemplo, la aplicación de GSMRs al estudio adaptativo de *Escherichia coli* a sus diferentes nichos, ha permitido identificar el *core* metabólico ancestral de *E. coli*, y las modificaciones metabólicas relacionadas a la adaptación del uso de carbono bajo condiciones limitantes de oxígeno (Baumler *et al.* 2011). Existen muchos ejemplos de microarreglos de genes que han sido utilizados en conjunto con las GSMR para proveer un entendimiento de porque ciertos cambios en la expresión génica ocurren en ambientes diferentes. Una comparación *in silico* de los flujos metabólicos contra datos de expresión de microarreglos en *E.coli* y *Sacharomyces cereviceae* revelaron que aquellos genes metabólicos cuyos flujos son acoplados, generalmente muestran patrones de expresión similares, comparten reguladores transcripcionales, y residen en el mismo operón al menos en *E .coli* (Notebaart *et al.* 2008). Así, las GSMR en organismos endosimbiontes pueden ser útiles para: establecer una relación entre la expresión de un gen y su función metabólica, probar hipótesis sobre las adaptaciones metabólicas relacionadas con la ausencia de alguna ruta metabólica y la re-direcciona el flujo metabólico para una eficiencia en la producción de la biomasa o a algún metabolito esencial para la

interacción con el huésped (Chechik *et al.* 2008). Por ejemplo, la utilización de GSMR en *Buchnera aphidicola*, una bacteria endosimbionte, comprobó que muchas de las rutas esenciales que se encuentran ausentes en el organismo, como resultado de la relajación de la selección purificante y la deriva genética, son provistas por el huésped (Ochman y Moran 2001, Yizhak *et al.* 2011). De forma interesante, también, se encontró que existe una pérdida preferencial para las rutas de biosíntesis de aminoácidos, lo que sugiere fuertemente que una característica de los organismos en proceso de reducción de su genoma es la pérdida en las rutas de biosíntesis de ciertos aminoácidos. Además, parece ser que esta forma de adaptación ha sido adoptada por muchas bacterias durante su especialización a ambientes ricos en nutrientes. Así mismo, éstos organismos sólo conservan aquellas rutas que les proveen de alguna ventaja adaptativa como es el caso de *B. aphidicola*, en donde la selección ha actuado manteniendo la duplicación de los genes *trpEG*, sobre-produciendo triptófano para el huésped (Lai *et al.* 1994). Cabe destacar que éstos estudios han asumido que éstas enzimas TrpEG involucradas en la biosíntesis de L-triptófano en *B. aphidicola* poseen una buena eficiencia catalítica. Sin embargo, hasta la fecha no existen reportes de sus parámetros cinéticos  $k_{cat}/K_M$  para determinar si en verdad son eficientes para la producción de L-triptófano, y es que, seguramente por el estilo de vida que presentan dichos organismos los genes ancestrales como las copias *trpEG* presenten mutaciones importantes ya sea para la función o estructura de dichas enzimas, de aquí la importancia de realizar estudios que integren el uso de GSMR con el efecto de la pérdida de rutas metabólicas, genes y funciones para tratar de entender la evolución de una función enzimática durante la adaptación de los organismos.

#### 1.2.1.1 EVOLUCIÓN DE LAS RUTAS DE BIOSÍNTESIS DE AMINOÁCIDOS A TRAVÉS DE GANANCIA DE GENES

La secuenciación genómica también, ha permitido observar que cada paso en las rutas de biosíntesis de aminoácidos está altamente conservado (Kanehisa y Goto 2000; Tatusov *et al.* 2000). Así mismo, dado que por años se ha sugerido que las rutas metabólicas evolucionaron a partir del



reclutamiento de enzimas con amplia especificidad por sustrato y su subsecuente especialización (Horowitz 1965; Jensen 1976; Cruz Morales *et al.* 2016), es frecuente encontrar enzimas con amplia especificidad por sustrato en los genomas actuales, al menos en *E.coli*, se ha encontrado que 22 de las 82 enzimas involucradas en la biosíntesis de los 20 aminoácidos (reportados hasta hoy) catalizan más de dos pasos en una ruta de biosíntesis de aminoácidos y de éstas, las siguientes 14 enzimas: ArgE, ArgH, AspA, CysK, CysM, DapE, HisA, HisC, HisF, IlvA, MetB, MetC, TcdB y YfdZ tienen un homólogo muy cercano que cataliza un sustrato diferente a su homólogo (Nishida 2001, Nam *et al.* 2012). Tal es el caso de la enzima HisA que tiene alta similitud con la ciclasa HisF (43% de identidad), ambas enzimas están involucradas en la ruta de biosíntesis de L-histidina. Se ha postulado que éstas enzimas evolucionaron por duplicación y subsecuente especialización, y se sugiere que la enzima ancestral HisA pudo haber sido una enzima con una amplia especificidad por sustrato (Fani *et al.* 1994).

#### 1.2.1.2 EVOLUCIÓN DE LAS RUTAS DE BIOSÍNTESIS DE AMINOÁCIDOS A TRAVÉS DE PÉRDIDA DE GENES

Durante los procesos de reducción del genoma debido a la adaptación a los diferentes estilos de vida como son: endosimbiótico, intracelular o patogénico, se ha observado que algunas de las rutas para la biosíntesis de aminoácidos están ausentes. Una de las posibles explicaciones a ésta pérdida de rutas ha sido la adaptación a hábitats en donde los aminoácidos se encuentran disponibles y son adquiridos por los organismos por medio de transportadores. Aunque, a la fecha no existe un estudio que revele si existe una tendencia hacia perder ciertas rutas de biosíntesis de aminoácidos (Tabla 1), se sabe que organismos cuyo tamaño del genoma oscila entre 2-4 Mpb han perdido rutas de biosíntesis de aminoácidos (Yu *et al.* 2011). Todo esto sugiere que durante la evolución del genoma bacteriano hacia un estado patogénico ó endosimbiótico ocurre mediante la pérdida de rutas

de biosíntesis de aminoácidos. Además, es altamente probable que éstas pérdidas de rutas metabólicas han tenido un efecto en la evolución de la red metabólica y de las enzimas.

<b>Estilos de vida</b>	<b>Organismo</b>	<b>Huésped</b>	<b>Rutas pérdidas</b>	
<b>Saprófitas</b>	<i>Bacillus anthracis</i>	Suelo, animales	Ninguna	
	<i>Burkholderia pseudomallei</i>	Suelo, animales	Ninguna	
	<i>Clostridium botulinum</i>	Suelo, animales	F,H,I, K, L,M, W,V	
	<i>Clostridium perfringens</i>	Suelo, animales	H,I, K, L,M, T, W,V	
	<i>Clostridium tetani</i>	Suelo, animales	F,H,I, K, L, M, W, V	
<b>Intracelulares</b>	<i>Pseudomonas aeruginosa</i>	Suelo, Plantas, Humano	Ninguna	
	<i>Anaplasma phagocytophilum</i>	Pulgas	F,H,I,K,L,M,R,T,W,V	
	<i>Bartonella quintana</i>	Humano	F,H,I,K,L,M,R,T,W,V	
	<i>Bartonella henselae</i>	Gato	F,H,K,L,M,R,W	
	<i>Brucella abortus</i>	Animales	Ninguna	
	<i>Chlamydia pneumoniae</i>	Humano	F,H,I,K,L,M,R,T,W,V	
	<i>Coxiella burnetii</i>	Vaca	F,H,I,K,L,M,R,T,W,V	
	<i>Ehrlichia chaffeensis</i>	Pulgas	F,H,I,K,M,R,T,W,V	
	<i>Francisella tularensis</i>	Ratón	H, M, R	
	<i>Legionella pneumophila</i>	Amiba	F,I,K,L,M,T,V	
	<i>Mycobacterium tuberculosis</i>	Humano	H	
	<i>Rickettsia typhi</i>	Pulgas	F,H,I,K,L,M,R,T,W,V	
	<b>Tracto respiratorio y mucosas</b>	<i>Bordetella pertussis</i>	Humano	M
		<i>Corynebacterium diphtheriae</i>	Humano	F, K
<i>Haemophilus influenzae</i>		Humano	K, M, R	
<i>Mycoplasma pneumoniae</i>		Humano	F,H,I,K,L,M,R,T,W,V	
<i>Neisseria meningitidis</i>		Humano	Ninguna	
<i>Staphylococcus aureus</i>		Humano	K, M	
<i>Staphylococcus haemolyticus</i>		Humano	F, H, K, W, V	
<i>Streptococcus pneumoniae</i>		Humano	H, K, R	
<i>Streptococcus pyogenes</i>		Humano	F,H,I,K,L,M,R,T,W,V	

<b>Membrana del tracto gastrointestinal</b>	Campylobacter jejuni	Humano	K, R
	Eenterococcus faecalis	Humano	A,F,H,I,L,V,W
	Escherichia coli K12	Humano	Ninguna
	Escherichia coli 0157	Humano	Ninguna
	Helicobacter pylori	Humano	F, H, I, L, M, R, V
	Salmonella typhi	Humano	Ninguna
	Shigella flexneri	Humano	Ninguna
	Streptococcus agalactiae	Humano	A,F,H,I,K,L,M,V,W,
<b>Otros</b>	Borrelia burgdorferi	Tick	F,H,I,K,L,M,R,T,W,V
	Burkholderia mallei	Caballo	Ninguna
	Neisseria gonorrhoeae	Uretra humana	Ninguna
	Yersinia pestis	Ratón	Ninguna
	Vibro cholerae	Zooplankton	Ninguna
	Vibrio vulnificus	Peces	Ninguna

**Tabla 1** Pérdidas de rutas de biosíntesis de aminoácidos en bacterias con diferentes estilos de vida\*

\*Tabla tomada de Yu *et al.* 2011

### 1.2.1.3 EVOLUCIÓN DE LA BIOSÍNTESIS DE L-HISTIDINA Y L-TRIPTÓFANO

En la mayoría de las bacterias la biosíntesis de L-Histidina requiere de un total de 10 pasos enzimáticos (asociados a 10 genes), tres de ellos *hisD*, *hisB* y *hisI* codifican para enzimas bifuncionales y dos de ellos *hisH* y *hisF* codifican para un solo polipéptido el cual forma una sola enzima que cataliza sólo un paso de las 10 reacciones. En bacterias Gram + , los genes de la biosíntesis de histidina están acomodados en un operón, el orden de los genes es *hisHAFI*, en donde los otros genes *hisDCB* están codificados al principio del operón (Fani *et al.* 2005). Estudios que se han hecho referentes a la evolución esta ruta biosintética como filogenéticos de síntesis química, han postulado que: los genes que forman parte de la ruta de biosíntesis de histidina han estado presente desde el último ancestro común y además los metabolitos con un grupo imidazol estuvieron presentes desde la sopa prebiótica (Alifano *et al.* 1996; Lazcano *et al.* 1992).

El triptófano es el aminoácido con la síntesis más costosa a nivel celular (Bentley *et al.* 1990) por lo que su biosíntesis esta altamente controlada (Yanofsky *et al.* 1981). Los genes que conforman la ruta de biosíntesis de triptófano generalmente se encuentran formando un operon, que involucra la organización de 5-7 genes (*trpEGDFCBA*). Estos genes han experimentado fusiones, lo que les ha permitido facilitar su transferencia en un solo paso. Por sus características, esta ruta ha sido utilizada para el estudio del impacto de la THG y la inserción de genes externos al operón (Xie *et al.* 2003; Priya *et al.* 2014).

Cómo ya se mencionó, bioquímicamente, el triptófano y la histidina son dos de los aminoácidos más costosos de producir, por lo que se ha observado que en bacterias patógenas o endosimbiontes, las cuales suelen obtener nutrientes del medio ambiente, dichas rutas están ausentes, salvo algunas excepciones como en el caso de *B. aphidicola* que conserva la ruta de biosíntesis de L-triptófano (Ochman & Moran 2001, Lai *et al.* 1994). Otros organismos se caracterizan por haber experimentado una dinámica de pérdidas y ganancias de algunos de los genes *trp*. Se sabe que, algunas especies de *Corynebacterium* adquirieron el operon completo de la biosíntesis de L-Triptófano, mediante una THG de algún miembro de las Gammaproteobacterias, esta THG pudo ratearse debido a que en gammaproteobacterias, los organismos tienen la característica de tener fusionados los genes *trpC-trpF* (Xie *et al.* 2003 y 2004; Noda-García *et al.* 2013). Cabe resaltar que, la mayoría de las especies clasificadas dentro del *filo* Actinobacteria han perdido el gen *trpF* necesario para la biosíntesis de L-triptófano, sin embargo, pueden producir el aminoácido (Barona-Gómez y Hodgson 2003). Tanto la biosíntesis de L-histidina como L-triptófano tienen en común dos conversiones enzimáticas de isomerización, las cuales son generalmente catalizadas por dos enzimas específicas por un solo sustrato: TrpF y HisA. Ambas enzimas adoptan un plegamiento tipo  $(\beta\alpha)_8$  y pertenecen a la misma superfamilia de proteínas (E.C 5.3.1). Sorpresivamente, estudiando estas rutas de biosíntesis en *Streptomyces coelicolor*, se encontró que este organismo no poseía un gen homólogo a *trpF*, y una enzima bi-específica homóloga a HisA, llamada PriA (del

inglés Phosphoribosyl Isomerase A), cataliza ambas reacciones (Barona-Gómez y Hodgson 2003). Posteriormente, mediante estudios de genómica comparativa entre especies de Actinobacteria se ha encontrado que aproximadamente el 70% de las Actinobacterias pierden el gen *trpF* (Verduzco-Castro *et al.* 2015). Los estudios sobre éstas rutas de biosíntesis se retomarán a detalle en la sección de Antecedentes.

En resumen, hemos descrito cómo la evolución del metabolismo en bacterias está modulada por la pérdida y ganancia de genes. Así mismo, los organismos que están en proceso de reducción del genoma tienden a perder rutas de biosíntesis de aminoácidos, en algunos casos éstas se conservan porque tienen una función adaptativa. También, hemos mencionado que dos rutas que han sido ampliamente utilizadas para entender cómo evolucionan las rutas metabólicas son la de L-histidina y L-triptófano, éstas son dos rutas ancestrales que son muy costosas de producir. De manera interesante, en Actinobacteria existe una enzima bi-funcional que cataliza dos pasos en estas rutas. Dicha enzima puede ser utilizada como un modelo para tratar de entender el efecto de la pérdida de genes sobre la evolución de la función dando origen a eventos de neo o sub funcionalización enzimática independientes de la duplicación génica.

### 1.3. ESTUDIO DE LA EVOLUCIÓN DE PROTEÍNAS DESDE UNA PERSPECTIVA POSTGENÓMICA

Zuckerkindl y Pauling observaron que las diferencias en aminoácidos entre dos proteínas ortólogas son proporcionales al tiempo de su divergencia desde el ancestro común. Este aparente ritmo constante de la evolución de proteínas es conocido como reloj molecular (Zuckerkindl y Pauling 1962). El reloj molecular provee las bases para estimar el tiempo de divergencia de una población y especies, detectar selección natural a escala genómica y entender el origen de las variaciones en las secuencias de proteínas y genes (Kumar *et al.* 2000; Yang *et al.* 1995). Tradicionalmente, la tasa de sustituciones no-sinónimas sobre las sinónimas  $d_N/d_S$  ha sido usado para detectar parámetros de selección en la evolución molecular (Kimura 1977; Yang y Bielawski 2000). Se considera que una

proteína esta bajo selección positiva cuando la tasa de sustituciones no sinónimas  $d_N$  excede a las sinónimas  $d_S$ . Un valor de  $d_N/d_S < 1$  es interpretado usualmente interpretado como que la proteína evoluciona lentamente bajo una selección negativa o purificante (esta altamente conservada), muchas de las mutaciones no-sinónimas son perjudiciales para el *fitness* y por ende tienen una probabilidad muy baja de ser fijadas. Cuando una proteína tiene un  $d_N/d_S \sim 1$ , se considera que la proteína evoluciona bajo la neutralidad (Kimura 1977). La estimación de los valores de  $d_N/d_S$  en la práctica requiere de modelos estadísticos de sustituciones así como de cadenas ocultas de Markov; específicamente los métodos de máxima verosimilitud y Bayesianos determinan las probabilidades de sustitución entre secuencias de genes ortólogos utilizando diferentes modelos de sustitución (Felsenstein y Churchill 1996; Yang *et al.* 2005). Es altamente posible examinar varia hipótesis biológicas con énfasis en la variación  $d_N/d_S$  a lo largo de varios sitios en una proteína y a lo largo de las ramas y clados de los árboles filogenéticos y distinguir cual es la hipótesis alterna usando una prueba de máxima verosimilitud (LRT, por sus siglas en inglés de likelihood Ratio Test).

### 1.3.1 ESTUDIOS SOBRE EVOLUCIÓN DE UNA FUNCIÓN Y CONECTIVIDAD METABÓLICA

Una estrategia que ha sido utilizada para relacionar la evolución de una función enzimática, después de un proceso de duplicación (ganancia) ó mutación (pérdida), dentro de la red metabólica, es la construcción de matrices estequiométricas que describen las reacciones químicas. Se asume que una enzima A y una enzima B están conectadas si comparten el mismo metabolito ya sea como producto o sustrato. En base a este supuesto de conectividad y al cálculo de las sustituciones no-sinónimas sobre las sinónimas ( $d_N/d_S$ ), se ha logrado establecer una relación entre la presión de selección que opera sobre ciertas enzimas y su conectividad en la red metabólica. Enzimas altamente conectadas que presentan una diferencia  $d_N/d_S < 1$ , son importantes para la adecuación del organismo y por ello están bajo un régimen de selección purificante, mientras que aquellas que presenten valores de  $d_N/d_S$

altos son enzimas que tienen poca conectividad en la red (Vitkup *et al.* 2006, Yizhak *et al.* 2011). Si bien este estudio es un inicio para tratar de entender la evolución de una función enzimática y su integración en la red metabólica es posible que los resultados a los que llegaron estén sesgados puesto que no tomaron en cuenta la presencia de enzimas bi-funcionales o con más de una función (promiscuas), que pueden tener más de una conexión en la ruta debido a que pueden catalizar una reacción sobre más de un sustrato diferente. Basados en este panorama la subsección siguiente describe la importancia del estudio de la promiscuidad como punto de partida para la evolución de una función enzimática.

### 1.3.2 EVOLUCIÓN DE LA FUNCIÓN MEDIADA POR LA PROMISCUIDAD ENZIMÁTICA

La promiscuidad enzimática puede definirse como la capacidad que tienen algunas enzimas para catalizar otra reacción química, adicional a la que se han especializado, dichas actividades promiscuas no son fisiológicamente relevante y se asume están sujetas a la neutralidad (Khersonsky y Tawfik 2010). Partiendo de este concepto de promiscuidad, es posible que estas actividades promiscuas sean el punto de partida para la evolución de una función enzimática como se ha comprobado con experimentos de evolución dirigida. Interesantemente el estudio de la enzima promiscua citocromo P450, ha permitido observar que a medida que se incrementaba el número de mutaciones al azar, se enaltecen las actividades promiscuas permitiendo aumentar su especificidad por sustrato (Bloom *et al.* 2007). Esta capacidad que tienen las enzimas para adoptar nuevas funciones rápidamente con los plegamientos ya existentes (confórmeros) o mediante un plegamiento nuevo, se conoce como evolvabilidad y ha sido detectada en enzimas promiscuas (Tokuriki y Tawfik 2009). También se ha observado que cuando se enaltece una actividad promiscua, se compromete la actividad original (Aharoni *et al.* 2005), por lo que las enzimas promiscuas se mantienen siendo promiscuas hasta que la selección favorece cierta actividad (Copley 2003; Khersonsky *et al.* 2006).

A pesar de la importancia de la acumulación de mutaciones al azar, el efecto que tiene esta acumulación de mutaciones neutrales sobre las propiedades biofísicas de las proteínas ha sido poco explorado. Por lo que, resultaría interesante determinar cómo estas actividades promiscuas que promueven el surgimiento de nuevas actividades se ve afectada o mejorada por el efecto en las características biofísicas de las proteínas.

Estas observaciones sugieren que, las enzimas que están sujetas a la neutralidad como el caso de este tipo de enzimas podrían ser buenos candidatos para el estudio del efecto de la pérdida de genes sobre la evolución de una actividad.

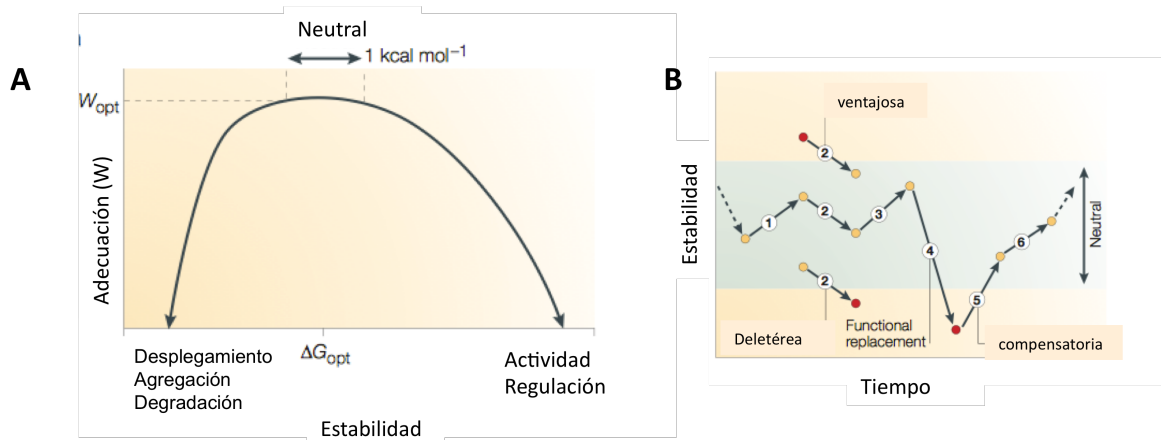
### 1.3.3 RESTRICCIONES BIOFÍSICAS SOBRE LA EVOLUCIÓN DE LA FUNCIÓN

Si bien, la acumulación de mutaciones neutrales, puede favorecer el surgimiento de una actividad o enaltecer una actividad promiscua, también puede afectar la estabilidad de las proteínas y es que, se ha observado que la acumulación de éste tipo de mutaciones tienen un efecto desestabilizador en la estructura de ciertas proteínas, éste efecto depende de la estabilidad primaria de las proteínas; es decir, en aquellas proteínas que son altamente estables las mutaciones casi siempre son neutras y no tienen un efecto en la estabilidad. Por otro lado, cuando las proteínas son propensas al desplegamiento, es decir aquellas proteínas que provienen de organismos en proceso de reducción del genoma, la selección actúa en dos niveles. Primero, la selección purificante elimina las mutaciones desestabilizadoras ( $d_N/d_S < 1$ ), y segundo, la selección positiva actúa fijando las mutaciones que estabilizan a la proteína ( $d_N/d_S > 1$ ) (Dasmeh *et al.* 2014). Además, algunas proteínas que se ven afectadas en su estabilidad tienden a formar agregados para estabilizarse sin embargo, muchas veces éstas proteínas son degradadas por la maquinaria celular (DePristo *et al.* 2005). Por lo que se ha propuesto que existe un costo-benéfico (función-estructura) cuando se gana una actividad, se afecta la estabilidad. Dado éste panorama se ha propuesto un modelo que tratan de explicar éste costo-beneficio, el cual se describe en la siguiente subsección.



### 1.3.3.1 MODELOS DE EVOLUCIÓN DE LA FUNCIÓN QUE INTEGRAN ADECUACIÓN Y ESTABILIDAD

El entendimiento de esta relación costo-beneficio, ha sido utilizada para llevar a cabo experimentos de evolución dirigida, en donde se crea un *screening* para detectar mutaciones relacionadas con el aumento en una función y después se hace un *screening* para detectar mutaciones compensatorias que restauran la estabilidad (Cuñago *et al.* 2006). Estos estudios han sido tomados en cuenta para proponer un modelo de cómo la evolución de la función, mediada por la acumulación de aparentes mutaciones neutrales, tiene un efecto en la estabilidad, seguido por la función y finalmente afectando la adecuación de los organismos dependiendo del tamaño de la población (DePristo *et al.* 2005). Así este modelo propone la existencia de un umbral en donde pueden acumularse las mutaciones neutrales sin afectar la estabilidad, manteniendo una estabilidad óptima, se mantiene la función y la adecuación también se mantiene (Figura 2A). A medida que va pasando el tiempo más mutaciones se van acumulando y existirán algunas mutaciones ventajosas que compensen la pérdida gradual de la estabilidad (Figura 2B). Sin embargo si se sobrepasa dicho umbral de estabilidad, se afectará la función y con ello la adecuación, llevando a las proteínas al desplegamiento, agregación y a la degradación (Figura 2A) (DePristo *et al.* 2005). Todo esto depende del fondo genético del organismo y de la estabilidad a priori de las proteínas en cuestión. Basados en este modelo surgen algunas preguntas sobre el efecto de estas mutaciones sin sentido en un fondo genético en donde la deriva genética es la que actúa, es decir proteínas que están sujetas a un régimen de selección relajada, provenientes de organismos en proceso de reducción del genoma, ¿se adecuarán a éste modelo? ¿pueden ser más tolerantes a la desestabilización y a la agregación, es decir el umbral de lo óptimo que se ha propuesto puede ser más grande?. Para contestar éstas preguntas se necesita de un modelo que nos permita estudiar dichos efectos.



**Figura 2.** Efecto de las mutaciones neutrales en la estabilidad, la función y la adecuación (W). A) Relación entre la estabilidad ( $\Delta G$ ) y adecuación (W). La adecuación disminuye cuando se observan mutaciones desestabilizantes que afectan el desplegamiento, promueven la agregación y aumentan la resistencia a la degradación. La zona de neutralidad de  $1 \text{ kcal mol}^{-1}$  se basa en la observación de que las proteínas operan sobre un rango de temperaturas ambientales. B) La evolución de la función esta restringida a mantener el umbral de estabilidad. La acumulación de mutaciones neutrales a través del espacio de secuencia puede sacar a las proteínas del umbral de la estabilidad. El número de mutaciones acumuladas (trayectoria) se representa con números, los círculos amarillos indican mutaciones neutrales, mientras que los círculos rojos indican mutaciones compensatorias necesarias para mantener dicho umbral (Tomado de DePristo *et al.* 2005).

### 1.3.4 EL PAPEL DE LA EPISTÁSIS EN LA EVOLUCIÓN DE LA FUNCIÓN

La epistásis se refiere a la interacción entre *loci* genéticos, puede ocurrir entre nucleótidos en un gen y entre genes en un genoma o entre mutaciones en residuos de aminoácidos en la secuencia de una proteína. En éste último punto, la epistaxis tiene un efecto positivo en el potencial evolutivo de ciertas actividades promiscuas; no todas las enzimas promiscuas ni los organismos de los que provienen son iguales en términos de evolvabilidad, es decir una mutación puede verse favorecida por la secuencia de la proteína en la que cae, efecto epistático (Palmer *et al.* 1999; Amitai *et al.* 2007). Enzimas con <30% de identidad a nivel de secuencia usualmente tienen estructuras y parámetros cinéticos parecidos pero sus actividades promiscuas difieren debido a la evolvabilidad o al efecto de la epistaxis, ya que ambas ejercen una influencia sobre las trayectorias evolutivas y de la acumulación de mutaciones necesarias para incrementar las actividades (Palmer *et al.* 1999, Salvedra *et al.* 2011; Miller *et al.* 2010). Si bien, la deriva genética puede incrementar o disminuir

la estabilidad de las proteínas esta desestabilización puede permitir la exploración de nuevas actividades promiscuas (Bloom *et al.* 2007; Bershtein *et al.* 2008). En resumen, algunos organismos tienen la capacidad para evolucionar nuevas funciones partiendo de actividades promiscuas, y otros no. La epistaxis afecta la evolución de la función que se enaltece a partir de enzimas promiscuas (Copley 2012). Por lo que es importante tomar en cuenta el efecto de las mutaciones en la secuencia completa de las proteínas.

En resumen, el estudio de todos los factores que restringen la evolución de la función sugieren que la evolución funcional se debe principalmente a la selección que actúa en enzimas con actividades promiscuas (Khersonsky *et al.* 2006; Khersonsky & Tawfik 2010). Por lo que, actualmente existe un gran interés en el estudio de la relación entre la promiscuidad, la estabilidad proteica y la diversidad conformacional (James & Tawfik, 2003), destacando ésta última como una de las más importantes. Se ha propuesto que dada la propiedad inherente de las proteínas para poseer diversidad conformacional (diferentes variantes de plegamiento), es posible que estos conformeros ó estructuras variables que se encuentran en equilibrio con su “estado nativo” puedan llevar a cabo actividades promiscuas (Tokuriki y Tawfik 2009). Además, la disponibilidad de secuencias nucleotídicas de proteínas y los métodos para cuantificar los diferentes tipos de selección, han demostrado que sólo una proporción muy baja de las mutaciones que se acumulan están sujetas a la selección para una nueva función, una alternativa a esto es que muchas de las mutaciones neutras que permiten la divergencia de secuencia no cambian la función biológica primaria o ancestral de la enzima *per se*, sino que producen cambios sutiles en la estructura de la proteína sin o alterando poco la función. Aquellas mutaciones neutras que se acumulan bajo las restricciones que les impone el mantenimiento de la actividad ancestral podrían sustancialmente alterar otras actividades promiscuas. Como muchos de estos cambios no afectan la adecuación de los organismos es muy difícil poder identificar enzimas con actividades promiscuas mediante una señal de selección. Sin embargo, recientemente se ha demostrado que utilizando información de secuencia e información

estructural cómo la diversidad conformacional de familias de proteínas es posible identificar promiscuidad enzimática (Noda-García *et al.* 2015, Zou *et al.* 2014). Lo que es un buen punto de partida para los estudios de evolución de la función.

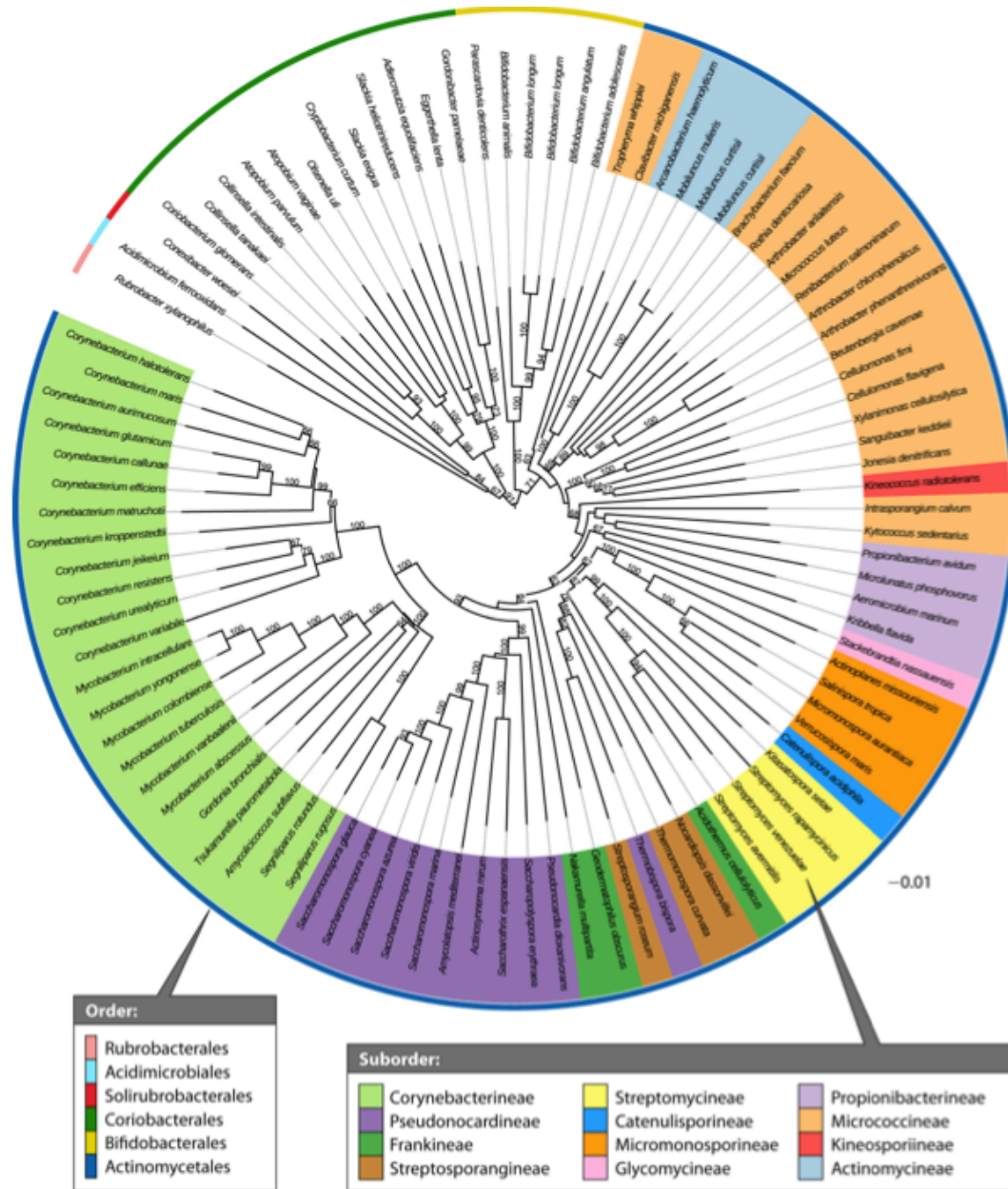
## 2. ANTECEDENTES

### 2.1 HISTORIA EVOLUTIVA DEL FILO ACTINOBACTERIA

El filo Actinobacteria, está compuesto por bacterias positivas para la tinción de Gram, alto contenido-(G+C) [más del 55 % y hasta 80% en sus genomas] y una gran diversidad en términos de su fisiología, morfología y capacidades metabólicas (Barka *et al.* 2016). Estudios recientes de las relaciones taxonómicas entre los miembros de este *filo* se han enfocado en resolver las relaciones taxonómicas de actinobacterias de importancia comercial, agrícola y médica como las del género *Streptomyces*, *Clavibacter* y *Mycobacterium*. Sólo recientemente, se ha hecho un esfuerzo por resolver las relaciones taxonómicas de todas las Actinobacterias que se encuentran secuenciadas, en estos estudios se ha utilizado una matriz de genes concatenados para obtener filogenias robustas que reflejen la historia evolutiva de las especies, dos grandes trabajos son el de Gupta *et al.* 2012; quien propone el uso de las firmas evolutivas, es decir secuencias de proteínas con inserciones o deleciones que son únicas para ciertos géneros, para la construcción de filogenias de cada una de las familias que componen al grupo de las Actinobacterias. El segundo trabajo de importancia es el de Barka *et al.* 2016; quien usa una serie de marcadores filogenéticos, es decir proteínas que se encuentran presentes en todas las actinobacterias para resolver las relaciones evolutivas de ciertos taxa (Figura 3). En ambos trabajos se puede observar que algunos comensales y patógenos como los pertenecientes a las familias: *Bifidobacteriaceae*, *Actinomyceaceae*, *Brevibacteriaceae*, *Coriobacteriales*, *Micrococcaceae*, *Nocardiodaceae*, entre otras se separan temprano del ancestro común (Silva *et al.* 2014; Gupta *et al.* 2012; Barka *et al.* 2016). De forma interesante, muchos de estos organismos tienen genomas reducidos debido a las adaptaciones nicho-específicas, por si fuera poco, la resolución taxonómica en esta parte de la filogenia es pobre, en parte porque no se cuenta con una secuencia del genoma debido a que muchos de ellos son comensales o patógenos

por lo que son difíciles de cultivar en el laboratorio y sólo se tienen secuencias parciales de sus 16s provenientes del uso de iTAGs.

Por lo que es evidente un estudio que tome en cuenta el mayor número de secuencias genómicas disponibles de Actinobacterias que se encuentran en la parte basal de las filogenias ya reportadas. Además, debido a las características de éstos organismos, parece ser que podrían estar perdiendo genes, y dado que uno de los objetivos de ésta tesis es entender cómo evoluciona la función enzimática durante la pérdida de genes, resulta interesante el estudio de éstos a detalle.



**Figura 3.** Filgenómica basada en 97 genomas de Actinobacterias, la filogenia fue construida con una matriz de secuencias de proteínas (Tomado de Barka *et al.* 2016).

### 2.1.1 ACTINOBACTERIAS BAJO DINÁMICA GENÓMICA DE PÉRDIDA DE GENES: LA FAMILIA ACTINOMYCETACEAE

Como se mencionó en la sección anterior, el estudio de organismos con una dinámica genómica de pérdida de genes es esencial para la realización de ésta tesis. Una de las familias que se ha propuesto esta bajo un proceso de pérdida de genes es la familia *Actinomycetaceae* (Zhao *et al.*

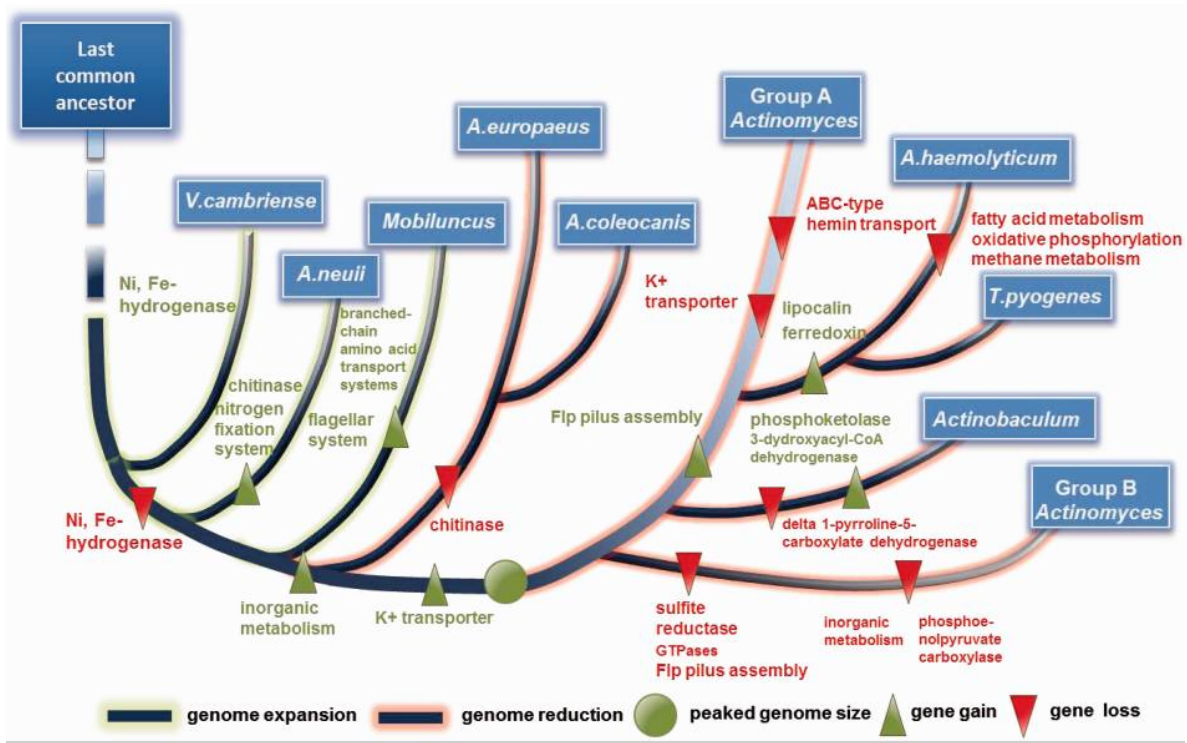
2014). Dicha familia fue identificada en 1918, actualmente se ha propuesto que ésta familia comprende seis géneros incluyendo *Actinomyces*, *Actinobaculum*, *Varibaculum*, *Mobiluncus*, *Arcanobacterium* y *Trueperella* especies (Schaal *et al.* 2006; Zhao *et al.* 2014). Muchos de los miembros de esta familia han sido encontrados exclusivamente como comensales y patógenos de humanos y animales de sangre caliente. Recientemente se encontró que miembros de esta familia también pueden colonizar el intestino de los pingüinos (Dewar *et al.* 2013). Lo que sugiere que esta familia co-evolucionó junto con los vertebrados y que posiblemente esta co-evolución surgió antes de la divergencia entre mamíferos y aves hace aproximadamente 315 millones de años.

A medida que se han secuenciado más genomas de organismos pertenecientes a ésta familia, la mayoría del género *Actinomyces*, se han hecho mejores estudios para tratar esclarecer las relaciones taxonómicas entre sus miembros (Schaal *et al.* 2006; Zhao *et al.* 2014). Sin embargo, dado que esta familia esta bajo una dinámica genómica importante, ha sido difícil resolver sus relaciones taxonómicas. A pesar de ello, se ha hecho un enorme esfuerzo para esclarecer las relaciones taxonómicas partiendo de la concatenación de 186 genes compartidos entre los miembros de esta familia. Hasta el momento, las relaciones taxonómicas sugieren que los géneros *Varibaculum*, *Actinobaculum*, y *Mobiluncus* son taxones hermanos del género *Actinomyces* (el más representado). Notablemente las especies que conforman este género se agrupan en tres sub-clados, los cuales difieren en el contenido- (G+C).

Otra característica importante de esta familia es la dinámica genómica que presenta la cual parece que esta marcada por episodios de pérdidas de genes, los cuales se ha propuesto como el mecanismo principal por el cual nuevos linajes pueden surgir por especialización al nicho que ocupan (Zhao *et al.* 2014). A pesar de que se reconoce una fuerte dinámica genómica en esta familia, el análisis reportado no da cuenta de la evolución de las rutas de biosíntesis de aminoácidos y debido a que nos interesa el estudio de la evolución de la función enzimática asociada a la dinámica genómica de pérdida de genes, en este estudio se plantea el análisis de esta dinámica en



las rutas de biosíntesis de L-Histidina y L-Triptófano que son dos rutas ancestrales y su efecto en la evolución de la función enzimática.

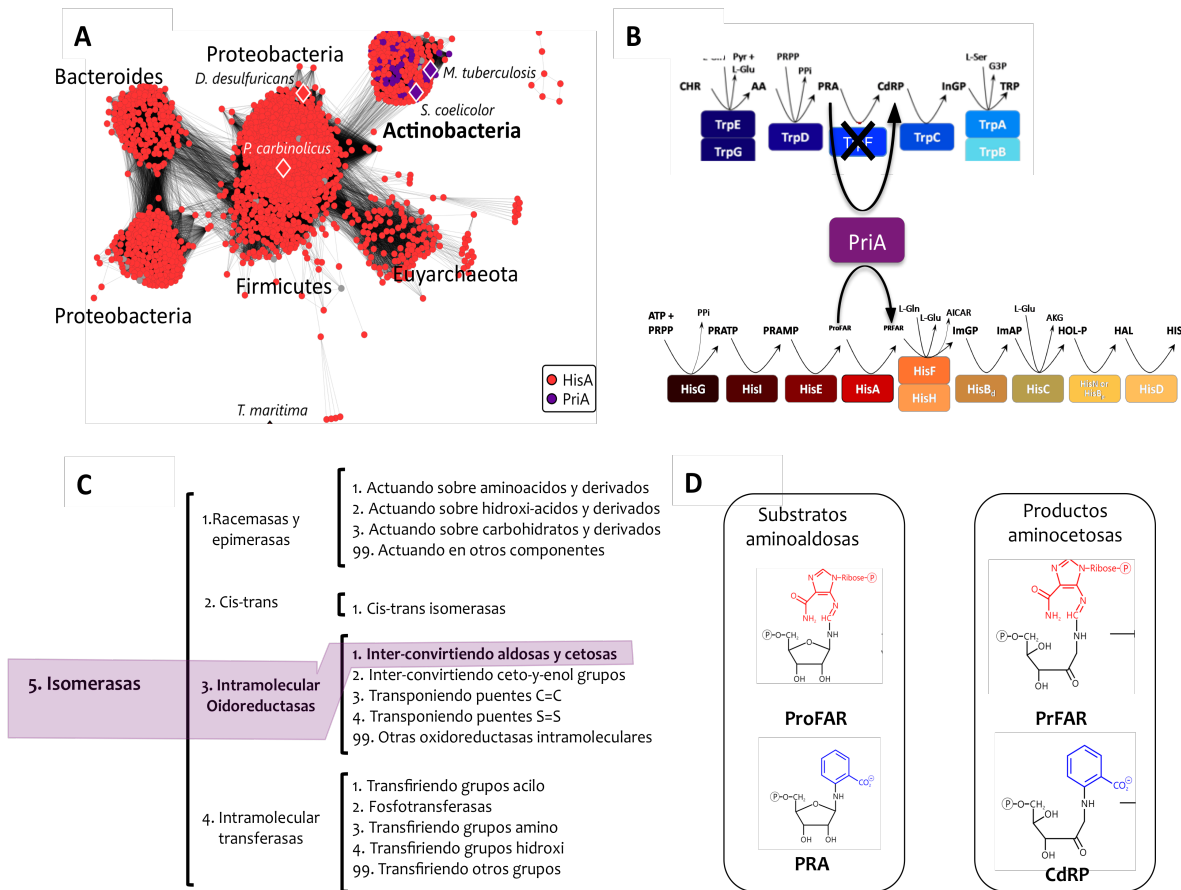


**Figura 4.** la Historia evolutiva de la familia *Actinomyceatceae* esta marcada por episodios de perdidas y ganancias de genes, lo que les ha permitido adaptarse mediante la especialización de sus genomas (Figura tomada de Zhao *et al.* 2014).

## 2.2 CONVERGENCIA DE LAS RUTAS DE BIOSÍNTESIS DE L-HISTIDINA Y L-TRIPTÓFANO EN ACTINOBACTERIA

Como se explicó en la sección 1.2.1.3, en Actinobacteria, incluida la familia *Actinomyceatceae* (Figura 5A), la biosíntesis de L-Histidina y L-Triptófano convergen en la enzima bi-funcional PriA (del inglés Phosphoribosyl Isomerase A), (Figura 5B). PriA cataliza dos reacciones de isomerización, las cuales son generalmente catalizadas por dos enzimas análogas y específicas por un solo sustrato: TrpF (PRA) isomerasa y HisA (ProFAR) isomerasa. Ambas enzimas adoptan un plegamiento tipo  $(\beta\alpha)_8$  y pertenecen a la misma superfamilia de proteínas (E.C 5.3.1) (Figura 5C-

D). Actualmente se sabe que el gen *hisA* esta presente en la mayoría de los grupos filogenéticos, no así el gen *trpF*, que esta ausente en la mayoría de las Actinobcaterias y poseen un homólogo de *hisA* llamado *priA* que cataliza ambas reacciones HisA y TrpF (Barona-Gómez & Hodson 2003). Como ya se mencionó PriA y HisA son homólogos, las dos enzimas son muy similares en cuanto a estructura y secuencia (Due *et al.* 2011; Kuper *et al.* 2005, Noda-García *et al.* 2015, Plach *et al.* 2016). Así mismo, mediante la reconstrucción ancestral y caracterización *in vitro* e *in vivo* de enzimas HisAs selectas, se encontró que la enzima HisA es bi-funcional desde hace más de dos billones de años (Figura 5A) (Plach *et al.* 2016). Sin embargo, los datos cinéticos e *in vivo* de las reconstrucciones ancestrales de estas HisAs sugieren que más que ser una enzima bi-funcional es una enzima promiscua hacia el sustrato análogo PRA.



**Figura 5.** Descubrimiento de PriA en Actinobacteria. A. Distribución de PriA en los diferentes fila. En morado se representa la existencia de una PriA mediante análisis de secuencia y en rojo la existencia de una HisA. B) Convergencia de la biosíntesis de L-Histidina y L-Triptófano en Actinobacteria. Las enzimas involucradas en ambas biosíntesis se muestran en cuadros de gradientes de rojo y azul respectivamente. El paso común entre estas dos rutas lo lleva a cabo PriA. C) Clasificación actual de la reacción que cataliza PriA. D) Mecanismo de reacción de PriA, PriA reconoce sustratos tipo aminoaldosas (ProFAR y PRA) y los convierte a aminocetosas mediante una isomerización.

## 2.2.1 INTERMEDIARIOS EVOLUTIVOS ENTRE HISA Y PRIA

Si bien, ya explicamos que PriA evolucionó de una enzima HisA promiscua por el sustrato PRA, el camino evolutivo a nivel molecular que se dio durante este proceso comienza a ser revelado, como parte de ésta tesis, los resultados obtenidos de este análisis fueron publicados y a continuación se explicarán a detalle (Noda-García y Juárez-Vázquez. 2015). Con la finalidad de explorar el posible estado ancestral de PriA, se diseñó una estrategia basada en el análisis de secuencias disponibles en

bases de datos públicas utilizando Modelos Ocultos de MARKOV (HMM). Para, identificar secuencias que estuvieran en un posible estado de transición entre HisA y PriA. Aunque la finalidad de este trabajo no era encontrar una enzima HisA promiscua hacia PRA, el resultado arrojó un conjunto de candidatos los cuales fueron promiscuos hacia PRA. Para la construcción de dichos perfiles se utilizó como semilla la secuencia de PriA de *S. coelicolor*. Así se desarrolló el perfil de PriA y subsecuentes secuencias fueron incorporadas en los siguientes perfiles durante cada iteración. Para la construcción del perfil se minó la base de datos de Uniprot; mediante BLASTp y se estableció un punto de corte, la redundancia a nivel de secuencia fue eliminada. Se obtuvo un total de siete perfiles, en donde se tomaron como perfiles óptimos aquellos que contenían una gran cantidad de secuencias tipo PriA con una baja varianza entre los valores esperados<sup>5</sup> asignados, y que tenían valores de E significativamente más bajos para PriAs que para HisAs. Así podemos ver que los perfiles 5 y 6 marcan una separación clara entre HisAs (círculos rojos) y PriAs (círculos azules), sin embargo no son capaces de distinguir entre la sub-familia de PriA subHisA (Noda García *et al.* 2013). Entre estos perfiles encontramos una región que representa lo que hemos llamado zona de transición (círculos verdes), que es una zona en la cual encontramos secuencias que el perfil no puede agrupar ni como HisA o PriA, por lo que suponemos que podrían ser intermediarios evolutivos (Figura 5A). Dichas secuencias provienen de i) Actinobacterias que se separaron temprano el ancestro común como *Actinomyces odontolyticus* y ii) una gamaproteobacteria *Photorhabdus luminescens*. Interesantemente, ambas secuencias provienen de organismos que se sabe pueden estar bajo un proceso de reducción del genoma, por lo que la divergencia que se observa en estas secuencias puede ser consecuencia de la selección relajada que experimentan (Moran *et al.* 2012). Estas secuencias además representan un buen escenario en donde se puede estudiar el efecto de la acumulación de mutaciones sobre la estructura-función de PriA.

---

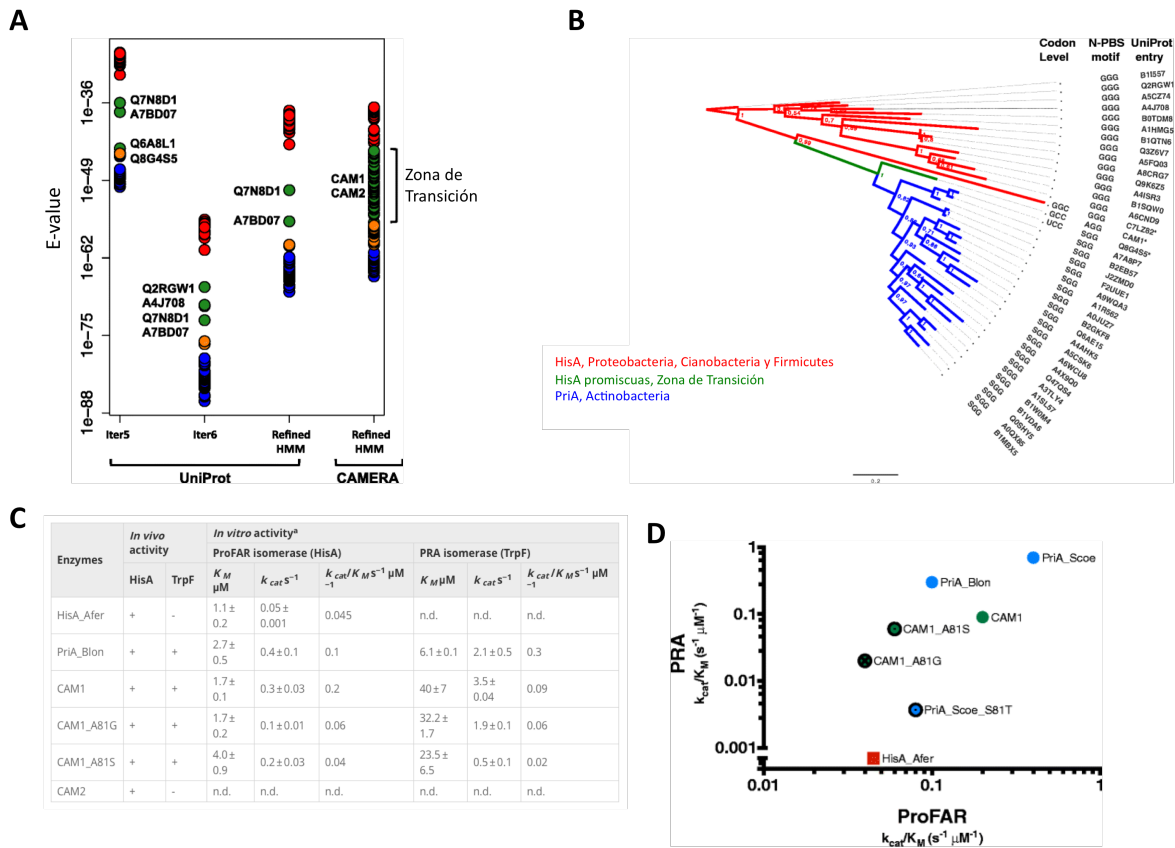
<sup>5</sup> El E-value o valor esperado es un parámetro que describe el número de hits que uno puede esperar ver por cambio, cuando lleva a cabo un alineamiento o una búsqueda en una base de datos de tamaño definido.

Después de esta validación del perfil con la base de datos de la Uniprot, el mismo perfil fue utilizado para buscar intermediarios evolutivos promiscuos en la base de datos de metagenomas marinos CAMERA. Así se lograron identificar PriAs, HisAs y secuencias intermedias (Figura 6A círculos verdes de la derecha). Dentro de estas secuencias intermediarias se encontraron dos que tienen una alanina en la posición 81, cabe resaltar que las secuencias de PriA tienen una Serina en la posición 81 y las HisAs poseen una Alanina en esta posición dicho residuo forma parte de uno de los sitios de unión a fosfato (Noda-García *et al* 2008; 2013). Por lo que se propuso un estado de transición a nivel mutacional para evolucionar de una HisA a una PriA GGC(Gly)→GCC(Ala)→UCC (Ser) Figura 5B. Por medio de un análisis filogenético conteniendo secuencias selectas se observó que algunas proteínas agrupadas dentro de las HisAs con el perfil provienen de Proteobacterias, Firmicutes y Cianobacterias. En la transición se encontró a la enzima HisA de *Acidimicrobium ferrooxidans* (llamada HisA\_Afer), un organismo que se agrupa con los Gram positivos con alto contenido-(G+C), y que pertenece al orden de los Acidimicrobiales. También dentro de las PriAs que se separan temprano del ancestro común están las provenientes de organismos como Bifidobacteriales y *Actinomyces*, lo cual se relacionado con la historia evolutiva de dichos organismos y soporta que PriA es única a Actinobacterias (Figure 5A y 6B) (Barka *et al.* 2016).

La caracterización bioquímica de algunas de las secuencias que se encontraron en la zona de transición mostraron que las secuencias denominadas como los intermediarios evolutivos (CAM), arrojaron parámetros catalíticos en el orden de magnitud parecidas a las PriAs bi-funcionales para la actividad de ProFAR isomerasa no así para la actividad de PRA isomerasa, lo que sugiere que el ancestro de PriA fue una enzima HisA con actividad promiscua hacia PRA (Figura 6C-D).

De manera interesante, detectamos que PriAs provenientes del género *Actinomyces* ocupan una posición diferentes, mientras una secuencia PriA de *A. naeslundii* se agrupa con las secuencias tipo PriA, la otra proveniente de *A. odontolyticus* se ubica como posible intermediario evolutivo, lo que

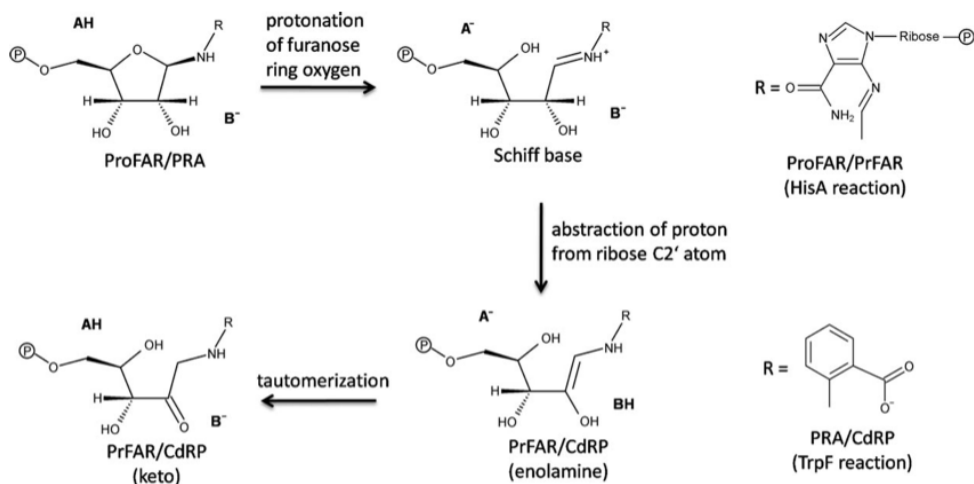
sugiere que ésta última podría estar evolucionando su especificidad por sustrato y que tal vez este escenario podría estar relacionado con una dinámica de pérdida de genes, por lo que el estudio detallado de enzimas provenientes de éste género son el objeto de estudio del presente trabajo.



**Figura 6.** Descubrimiento de secuencias intermedias entre HisA y PriA. A) Clasificación de enzimas en HisA y PriA usando el perfil construido y las bases de datos de la Uniprot y de metagenomas marinos, CAMERA. B) Análisis filogenético de las secuencias clasificadas por el perfil como HisA, PriAs e intermedios evolutivos. También se muestra el estado de transición a nivel de nucleótidos. C) Parámetros cinéticos obtenidos de algunas secuencias tipo HisA, PriA y de intermedios evolutivos. D) Caracterización funcional de enzimas selectas tipo HisA, PriA e intermedios evolutivos, secuencias CAM y mutantes.

### 2.2.2 RELACIÓN ESTRUCTURA FUNCIÓN DE PRIA

La reacción que cataliza PriA consiste en la transformación de una aminoaldosa a una aminocetosa a través de una catálisis ácido-base que involucra un rearrreglo del tipo de Amadori. PriA cataliza la reacción de conversión de PRA a 1-(*o*-carboxi-fenilamino)-1-deoxirribulosa-5'-fosfato (CdRP) para la actividad de PRA isomerasa (TrpF), mientras que también cataliza la conversión ProFAR en PrFAR, *N*-(5'-fosforribosil)-formimino-5-aminoimidazol-4-carboxamida ribonucleotido en la actividad de ProFAR isomerasa (HisA). Aunque los sustratos PRA y ProFAR presentan un grupo ribosa-5-fosfato, sus grupos funcionales (R) difieren por completo: ProFAR es dos veces más grande que PRA, y presenta un grupo ribosa-5-fosfato adicional; por ende, el sitio activo de HisA y TrpF también difiere sustancialmente (Henn-Sax *et al.* 2002). HisA presenta dos sitios de unión a fosfato (PBS, de sus siglas en inglés Phosphate Binding Site), mientras que TrpF presenta sólo uno. La enzima PriA (homóloga a HisA) une al sustrato ProFAR mediante los dos grupos fosfato (los cuales se ubican en posiciones simétricas a lo largo del sitio activo); y la unión con el sustrato PRA, que es más pequeño, ocurre mediante la unión a un sólo grupo fosfato. Las diferencias entre las uniones con diferentes sustratos sugieren que la enzima PriA desarrolló una estrategia estructural novedosa para acomodar dos sustratos análogos, topológicamente disimilares con el mismo sitio activo. Esto es posible gracias a cambios conformacionales principalmente de las asas que tienen un papel central acomodando a los residuos funcionales (Figura 7) (Wright *et al.* 2008. Noda-García *et al.* 2010; Düe *et al.* 2011, Verduzco-Castro *et al.* 2016).



**Figura 7.** Mecanismo de reacción de PriA. Esquema de la isomerización de ProFAR y PRA, reacciones catalizadas por las enzimas TrpF y HisA. La reacción comienza con la protonación del oxígeno en el anillo de furanosa de la aminoaldosa (ProFAR y PRA) por el ácido general (HA), y su subsecuente corte del esqueleto C-O. Así, resulta la base intermedia de shiff, la cual es desprotonada por la base general (B<sup>-</sup>) en el átomo de C2 de esta ribosa abierta. Los productos de la reacción son la enolamina de PrFAR y CdRP, los cuales se tautomerizan espontáneamente a sus productos finales (Tomado de List *et al.* 2011).

La bi-funcionalidad de PriA ha sido demostrada *in vitro* midiendo los parámetros cinéticos para ambas actividades PRA y ProFAR de varias enzimas provenientes principalmente de algunas especies de *Streptomyces* y *M. tuberculosis* (Noda-García *et al.* 2009; Verduzco-Castro *et al.* 2016; Due *et al.* 2011). Desafortunadamente, las actividades catalíticas de PriA de *M. tuberculosis* que han sido reportadas han utilizado un pH no óptimo para la reacción y por ende los parámetros catalíticos de  $k_{cat}/K_M$  no son comparables con los de *Streptomyces* (Camacho-Zarco Tesis de Maestría).

Hasta el momento, 10 estructuras cristalográficas de los homólogos de PriA han sido reportadas, la mayoría provienen del género *Streptomyces*. Cuatro de ellas corresponden a PriAs de *S. coelicolor* (5DN1, 2VEP, 2X30 y 1VZW), dos a *S. sp* MG-1 (4X9S, 4W9T) y dos a *S. svicensis* (4TX9 y 4U28) sólo dos PriAs provienen de *M. tuberculosis* (2Y85, 3ZS4)(Tabla 3). Las estructuras de *M. tuberculosis* son las que presentan análogos de los sustratos PRA (2Y85) y ProFAR (3ZS4)



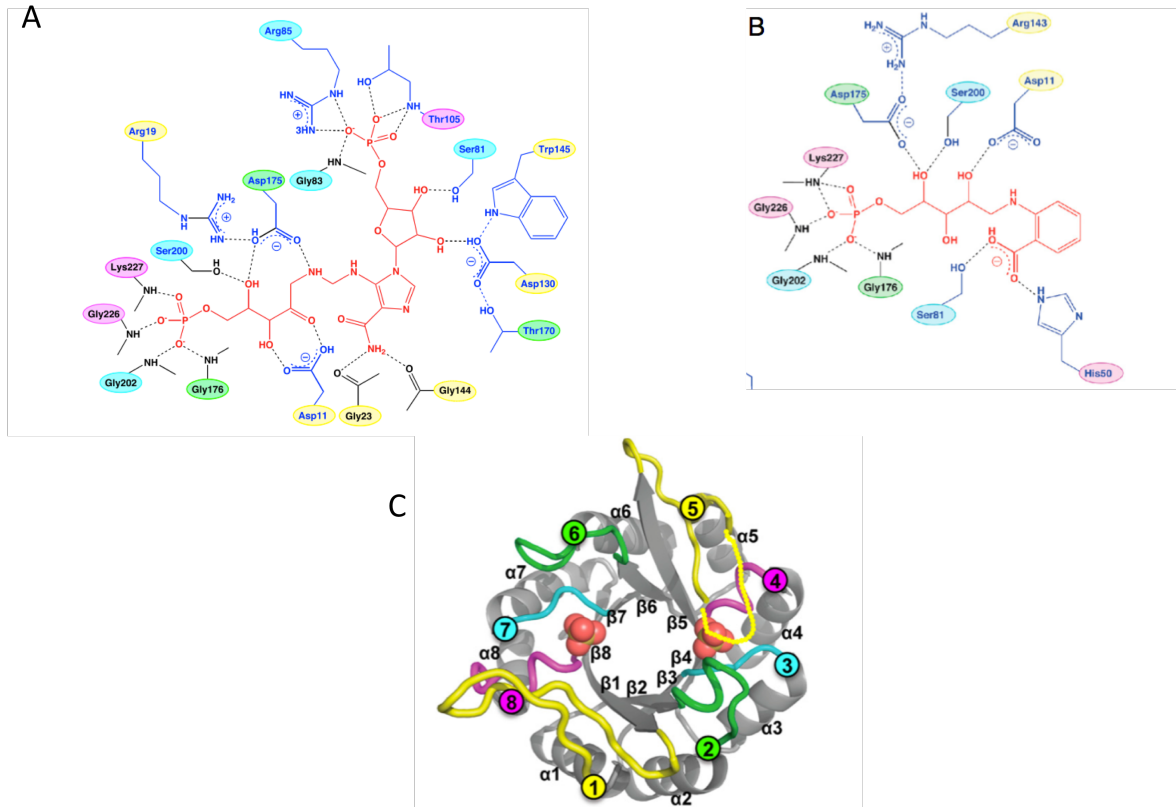
localizados en el sitio activo de la proteína. A partir de estos resultados y de la caracterización funcional de algunas mutantes puntuales, se han logrado obtener algunas conclusiones acerca de la relación estructura-función del sitio activo que le da la bi-funcional a PriA (Wright *et al.* 2008 y Due *et al.* 2011)(Tabla 2).

PDB	Organismo	Resolución Å	Características	$k_{cat}/K_M$		Referencia PDB
				( $\mu\text{M}^{-1} \text{s}^{-1}$ ) ProFAR	( $\mu\text{M}^{-1} \text{s}^{-1}$ ) PRA	
5DN1_PriA	<i>S. coelicolor</i>	1.95	sin sustratos	0.4	0.7	Chang et al. 2015
4X9S_PriA	<i>S. sp.MGI</i>	1.6	sin sustratos	0.069	0.0022	Michalska et al. 2014
4W9T_HisAp	<i>S. sp.MGI</i>	1.57	sin sustratos	0.069	0.0022	Michalska et al. 2014
ATX9_HisAp	<i>S. sviveus</i>	1.6	ProFAR	0.18	0.67	Michalska et al. 2014
4U28_PriA	<i>S. sviveus</i>	1.33	sin sustratos	0.18	0.67	Chang et al. 2014
3ZS4_PriA	<i>M. tuberculosis</i>	1.9	PrFAR	0.1		Due et al. 2012
2y85_PriA	<i>M. tuberculosis</i>	2.4	CdRP		0.01	Kuper et al. 2011
2X30_PriA	<i>S. coelicolor</i>	1.95	mutant R139N	0.07	ND	Noda-García et al. 2010
2VEP_PriA	<i>S. coelicolor</i>	1.8	sin sustratos	0.4	0.7	Wright et al. 2008
1VZW_PriA	<i>S. coelicolor</i>	1.8	sin sustratos	0.4	0.7	Kuper y Wilmanns 2005

**Tabla 2** Estructura función de homólogos de PriA.

La enzima PriA presenta un plegamiento tipo barril ( $\beta\alpha$ )<sub>8</sub>; el cual es un arreglo de repetidos con asas 1-4 y 5-8 en el sitio activo. La comparación con los dos complejos CdRP y PrFAR unidos a PriA\_ *M. tuberculosis* en cuanto a orientación con los sustratos es similar a la reacción de ProFAR isomerasa que ocurre con estereoquímica sobre una de las ribosas fosforiladas. La unión del fosfato de esta ribosa y del fosfato de PRA se lleva a cabo en el PBS C-terminal, conformado por los

residuos Gly176, Gly202, Gly226 y Lys227 (asas 7 y 8). La unión del fosfato de la ribosa de ProFAR que no se isomeriza, se lleva a cabo en el PBS N-terminal, formado por los residuos Gly83, Arg85 y Thr105 (asas 3 y 4). Los residuos Asp11 y Asp171 actúan como ácido/base generales en la isomerización de ambos sustratos. Los residuos Arg19, Arg143 y Trp145 se localizan en las asas 1 y 5 y tienen un papel importante durante la conformación del sitio activo cuando ocurre la isomerización. La isomerización de ProFAR se ve afectada cuando alguno de los residuos anteriores es mutado; mientras que la isomerización de PRA solo se ve afectada cuando se muta el residuo Arg143. Así la Arg19 (asa 1) es reclutado junto con el Asp175 en el sitio activo para la isomerización de ProFAR. La importancia de la Arg143 recae en que este residuo forma un puente salino con la Glu109, la cual es aparentemente un residuo crítico que se ha observado en la conformación del asa 5, esto provee la base del cambio en la conformación de Arg143/Trp145. La mutación del residuo Trp145 mata la actividad de ProFAR isomerasa. Por el contrario para la isomerización de PRA el cambio en los residuos Thr105, Asp130 y Thr170 matan la actividad de PRA isomerasa (Due *et al.* 2011) (Figura 8).



**Figura 8.** Estructura del sitio activo de PriA de *M. tuberculosis* A) Se muestran las interacciones de residuos que participan en la isomerización de ProFAR. B) Se resaltan los residuos y las interacciones de PriA con PRA y C) Estructura tridimensional de PriA en donde se muestran las asas que interactúan con los sustratos. Los colores con que se marcan los residuos en los incisos A y B están referenciados con su posición en las asas que se muestran en el inciso C (Modificado de Due *et al.* 2011).

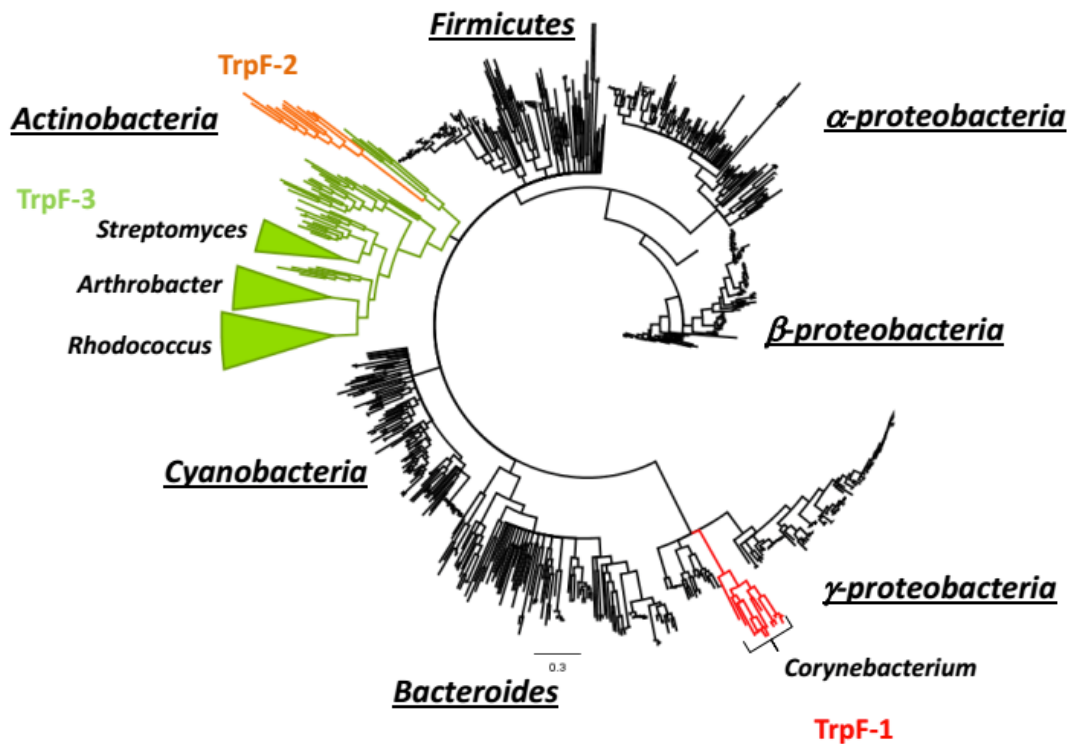
De forma interesante, en las enzimas HisAp de *S. svicensis* y *S. sp.* MG1, la Arg141 conservada en las PriAs de *S. coelicolor* y *M. tuberculosis* como Arg143 es reemplazada con una Lys141. Esta Lys141 forma un puente salino con la Glu148 del asa 5. A pesar de este reemplazo, la Gly140 puede interactuar con el Trp143 de forma débil, sin embargo no es capaz de estabilizar totalmente vuelta- $\beta$  entre  $\beta 5a$  y la  $\beta 5b$ . Por lo que este re-arreglo podría no ser compatible con el sustrato PRA (Verduzco-Castro *et al.* 2016).

### 2.2.3 DINÁMICA GENÓMICA Y LA EVOLUCIÓN DE SUBFAMILIAS DE PRIA

En un estudio previo en donde se analizó exhaustivamente la presencia y organización de los genes *his/trp* en 101 genomas de Actinobacteria se encontró que la mayoría de los genes de biosíntesis de L-Histidina y L-Triptófano están conservados en Actinobacteria y se encuentran en copia sencilla, se encontraron algunas excepciones, las cuales se ha propuesto pueden servir como marcadores para resolver relaciones taxonómicas (Tesis Noda-García 2013). Basados en este análisis se propusieron dos grandes escenarios en donde PriA ha sido retenida. En el primero, PriA co-evolucionó ante la presencia del gen *trpF* y en el segundo escenario, PriA evoluciona después de haber adquirido el operón completo para la biosíntesis de L-Triptófano. Dichos escenarios han dado lugar a la formación de subfamilias las cuales se describen a continuación.

#### 2.2.3.1 CO-EVOLUCIÓN DE PRIA Y EL GEN TRPF, SUBFAMILIA HISAP

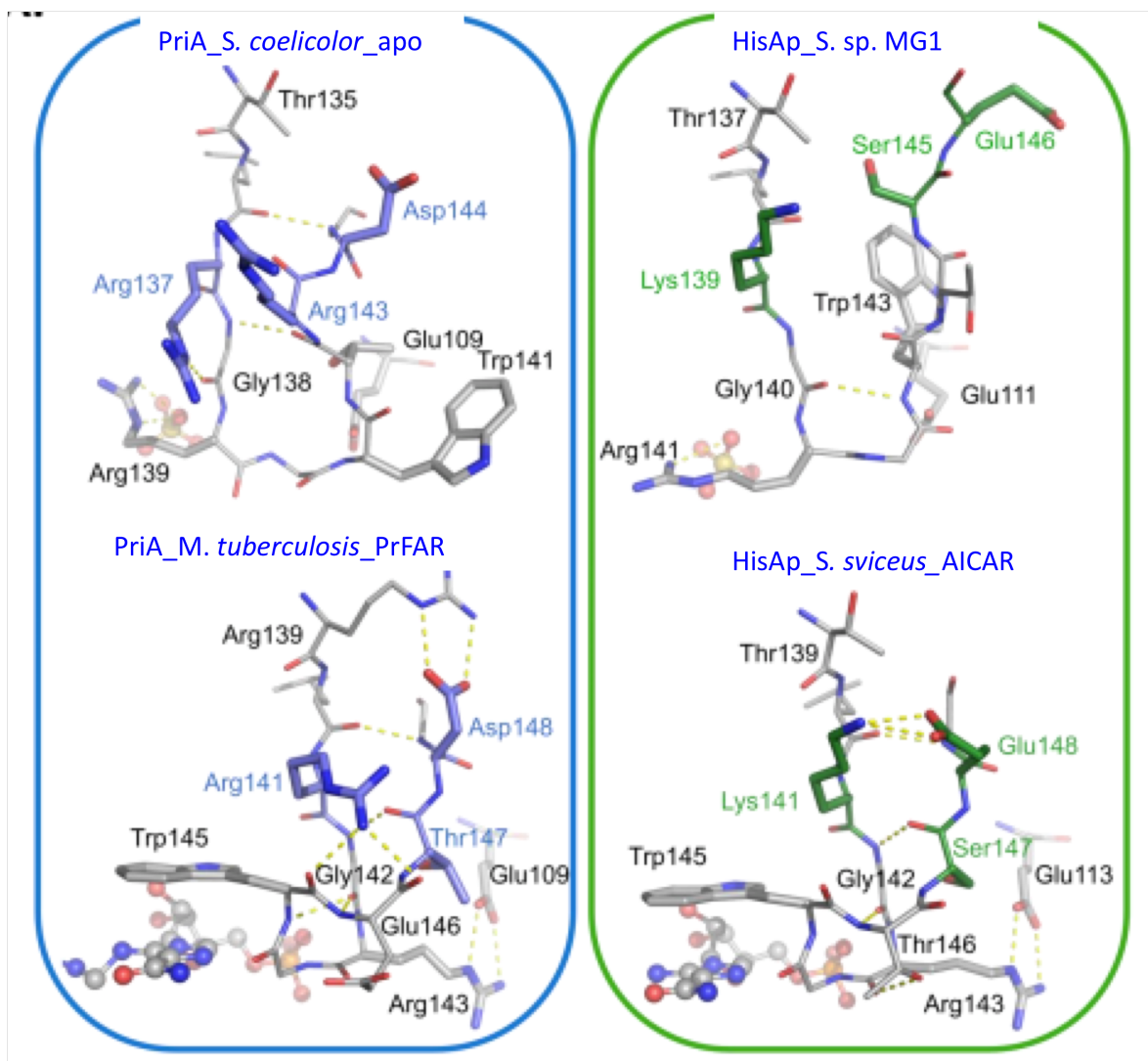
Mediante el uso de análisis filogenéticos de los homólogos de *trpF* presentes en Actinobacterias, se encontró que existen tres clases de homólogos de TrpF en Actinobacterias: TrpF1, TrpF2 y TrpF3. Estas sub-clases son género específicas y cada una de estas clases contiene varias secuencias de corynebacterias, las cuales han adquirido el gen *trpF* mediante transferencia horizontal de las gammaproteobacterias (Noda- García *et al.* 2013). Los homólogos de *priA* que se encuentran en este grupo han sido caracterizados bioquímicamente y se ha encontrado que han pasado por un proceso de sub-funcionalización que ha favorecido a la actividad ProFAR isomerasa. Los otros sub-clados (TrpF2 y TrpF3) en los cuales se agrupan los homólogos de TrpF provenientes de organismos como *Nocardia*, *Actinokineospora* y *Streptomyces*. Interesantemente, estos homólogos de TrpF muestran sintenia (Figura 9). Los genes que se encuentran asociados a su contexto genómico son *trpD*, *trpA*, *trpC* así como el gen *ectC*, el cual codifica para una L-ectoina sintasa, estos genes se encuentran formando un operón asociado con la producción de productos naturales al menos en *Micromonospora sp.* (Zhou *et al.* 2013). El otro subclado contiene genes *trpF* provenientes de especies de *Rhodococcus*, *Arthrobacter* or *Streptomyces*, grupo TrpF3.



**Figura 9.** Árbol filogenético construido con secuencias de la proteína TrpF presente en todos los grupos de bacterias. Se observa que géneros del filo Actinobacteria conservan el gen *trpF*, pero este gen no tiene un origen común (Tomado de Verduzco-Castro *et al.* 2016).

La caracterización de estas enzimas *in vivo e in vitro* demostró que estas tienen actividad de TrpF isomerasa. Interesantemente, los homólogos de PriA que se encuentran en este escenario donde está presente TrpF, presentan una especificidad por sustrato para ambos sustratos PRA y ProFAR. Sin embargo, la actividad de PRA isomerasa está disminuida dos órdenes de magnitud con respecto a las PriAs bonafide que se han caracterizado, estas actividades son comparables con las actividades de PRA isomerasa detectadas en los intermediarios evolutivos CAM. Estas actividades bajas se han asociado con cambios en residuos altamente conservados en las PriAs como la Arg143 por una Lys143 en estos homólogos, los cuales se han denominado como enzimas tipo HisAp, para hacer alusión a aquellas enzimas que evolucionan en un contexto en donde se retiene la enzima TrpF ancestral (Verduzco-Castro *et al.* 2016) (Figura 10 y 11A). Este cambio de la Arg143 por una

Lys143 también se ha encontrado en *Thermotoga maritima* (Plach *et al.* 2016). Finalmente, cabe resaltar que la influencia de esta actividad promiscua en las HisAp tiene una relación con la tasa de evolución acelerada detectada en estas enzimas tipo TrpFs, lo que les permite explorar el espacio de secuencia y tal vez ganar una actividad nueva como es el caso de la enzima TrpF de *Chlamydia*, la cual se ha observado tiene actividad de GTP-hidrolasa. (Adams *et al.* 2015) (Ver sección 11 de resumen de publicaciones relacionadas y contribuciones de ésta tesis).



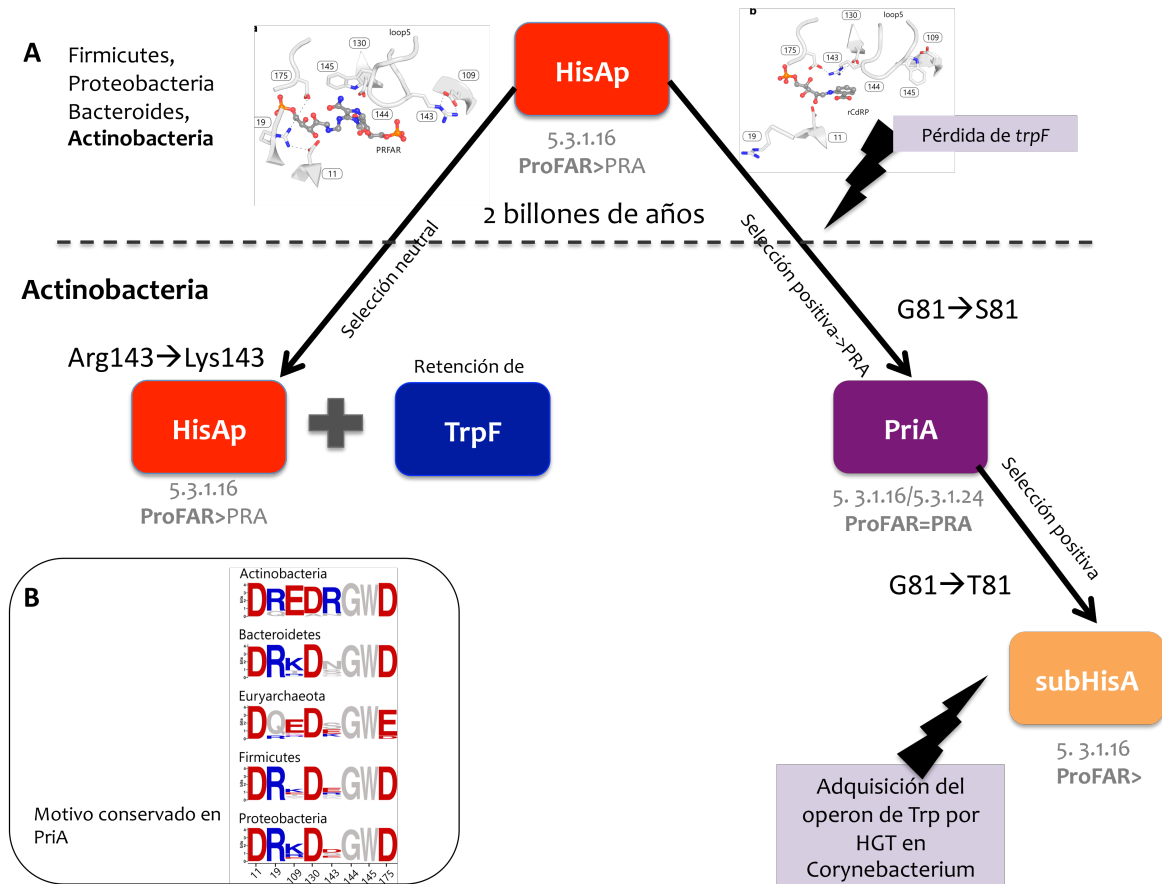
**Figura 10.** Diferencias a nivel estructural entre enzimas tipo PriAs y HisAp. En el primer recuadro se muestran las enzimas tipo PriA de *S. coelicolor* y *M. tuberculosis* en la conformación apo y unida a ProFAR, se muestra el acomodo de la Arg139 y la Arg143 para conseguir una unión eficiente a ProFAR. En las HisAp, el cambio de la Arg139 por la Lys139 no permite el correcto acomodo del sustrato, por lo que se reduce la actividad detectada en estas enzimas (Tomado de Verduzco-Castro *et al.* 2016).

#### 2.2.3.2. EVOLUCIÓN DE LA ESPECIFICIDAD POR SUSTRATO DE PRIA PROMOVIDA POR THG: SUBFAMILIA SUBHISA

Basados en la filogenia de los grupos de congruencia, se encontró que miembros del género *Corynebacterium* habían adquirido por transferencia horizontal el operon completo para la biosíntesis de L-triptófano (WPTO), esta adquisición proviene del operon de las Gammaproteobacteria (Xie *et al.* 2004). Se sabe además, que las corynebacterias son un grupo hermano de las Mycobacterias que son organismos que se comparten un ancestro común, pero cuya divergencia es reciente (Barka *et al.* 2016). Por lo que se estudió si la adquisición de este operon tenía una influencia sobre la especificidad por sustrato de las PriAs en este grupo de organismos.

El estudio filogenómico hecho con marcadores como RpoB y la comparación del contexto genómico de los genes *his/trp* en corynebacterias dio como resultado la existencia de dos grandes grupos. En un clado se agrupaban todas aquellas corynebacterias que habían retenido el operon WPTO, y en el otro clado aquellas que habían perdido el operon. Algunas PriAs provenientes de ambos clados fueron caracterizadas *in vivo* e *in vitro*, y se encontró que aquellas que provenían del contexto en donde se encontraba presente el operon WPTO habían perdido la capacidad para catalizar a PRA como sustrato, por lo que se les llamo subHisAs, debido a la sub-funcionalización hacia la actividad de ProFAR isomerasa. Mediante un análisis de secuencia con PriAs caracterizadas se encontraron cambios en residuos involucrados en la catálisis como son Arg143, His50 y Ser81 (Figura 11A) (Due *et al.* 2013, Noda-García *et al.* 2015, Plach *et al.* 2016). Dichos residuos están reemplazados en las subHisAs como Asn142, Thr81. Estos cambios junto con otros que no están conservados participan en este cambio de actividad de PriA → subHisA.

Interesantemente, entre estas enzimas se encontró que la diversidad conformacional media el cambio de especificidad. Por un lado, las PriAs adoptan más estados conformacionales en comparación con las subHisAs, las cuales solo adoptan un estado conformacional.



**Figura 11.** Evolución de PriA en Actinobacteria. A) PriA (morado) evolucionó hace aproximadamente 2 billones de años de una enzima tipo HisA promiscua para la actividad PRA isomerasa, en algunas actinobacterias se pierde el gen *TrpF* y HisAp evoluciona por medio de la acumulación de mutaciones específicamente de una Gly81 a una Ser81, lo que le permite ser eficiente para catalizar ambos sustratos ProFAR y PRA. Algunas *Corynebacterias* adquieren un operon completo de Triptófano y PriA se subfuncionaliza hacia una SubHisA (naranja, sólo cataliza a ProFAR como sustrato) mediante la mutación Ser81 a Thr81. Algunas Actinobacterias conservan el gen *trpF* y HisAp evoluciona en este contexto. B) Motivos conservados en PriAs.



### 3. JUSTIFICACIÓN

Dado que se ha observado que la pérdida de genes es un mecanismo que promueve la evolución y adaptación de los genomas bacterianos, poco se sabe del efecto que pueden llegar a tener éstas pérdidas de genes sobre la evolución de la especificidad por sustrato de ciertas enzimas. Muchas preguntas relacionadas a éste tema, siguen estando sin responder. Ésta tesis postula el uso de la filogenómica bacteriana, como marco conceptual, para analizar una trampa evolutiva, en la que caen las enzimas con más de una función. El análisis de estos episodios evolutivos nos dará la oportunidad de entender el efecto a nivel de la función enzimática de la pérdida de genes, lo que es caracterizable a nivel experimental, y además esto puede ser de interés biotecnológico para el diseño de compuestos que inhiban dichas enzimas, provenientes de organismos que son patógenos de humano.

## 4. HIPÓTESIS

El uso de la filogenómica en algunas especies bajo deterioro genómico, nos permitirá detectar episodios de pérdidas diferenciales de genes *his* y *trp*, que impactarán la evolución de la especificidad por sustrato de aquellas enzimas bifuncionales o promiscuas que se retengan.

### 4.1. Hipótesis nula

La pérdida diferencial de los genes *his* y/o *trp* no tendrá efecto en la evolución de la función de las enzimas bifuncionales que se retengan.

## 5. OBJETIVOS

### 5.1. OBJETIVO GENERAL

Detectar por medio de análisis filogenómicos episodios de pérdida de los genes *his* y *trp* en *Actinomycetaceae* con la finalidad de analizar el efecto de estas pérdidas sobre la evolución de la función de PriA mediante análisis computacionales y experimentales en el laboratorio.

#### 5.1.1. OBJETIVOS ESPECÍFICOS

1. Detectar familias de Actinobacterias que estén bajo una dinámica de pérdida diferencial de los genes *his* y *trp* mediante análisis filogenómicos y de características genómicas como: tamaño del genoma, contenido de G+C, número de genes que codifican para proteínas, etc., para seleccionar una familia, la cual será utilizada como modelo de estudio.
2. Obtener más organismos pertenecientes a la familia seleccionada (*Actinomycetaceae*) por medio de un muestreo microbiológico de las cavidades orales y/o búsqueda de grupos que trabajen con dichos organismos para obtener DNA genómico y secuenciarlo.
3. Resolución de las relaciones taxonómicas de la familia selecta para llevar a cabo análisis de pérdidas de genes centrándose en las rutas de biosíntesis de L-Histidina y L-Triptófano, que son las rutas en las que converge PriA, y determinar si las pérdidas en los genes *his/trp* son clado específicas.
4. Cuantificar la tasas de sustitución nucleotídica de los homólogos de *priA* para establecer el tipo de presión de selección que opera en los distintos escenarios en donde evoluciona PriA,

por ejemplo, determinar qué presión de selección opera en PriA durante la pérdida de genes.

5. Establecer la influencia de la pérdida de los genes *his/trp* en la especificidad por sustrato de PriA mediante la determinación de los parámetros cinéticos ( $k_{cat}/K_M$ ) de PriAs selectas y el papel funcional de PriA in vivo en la cepa *A. oris* MG-1.
6. Conocer los determinantes estructurales que median la especificidad de PriA por los sustratos ProFAR y PRA en la familia *Actinomycetaceae*, mediante análisis de secuencia, cristalografía de rayos X y modelado por homología de PriAs selectas.

## 6. MATERIALES Y MÉTODOS

### 6. 1. MÉTODOS BIOINFORMÁTICOS

#### 6.1.1 SECUENCIACIÓN, ENSAMBLE Y ANOTACIÓN DE GENOMAS

##### 6.1.1.1. SECUENCIACIÓN Y ANÁLISIS DE LECTURAS (READS)

Los genomas de las cepas *A. oris* MG1, *A. oris* MG1\_Δ*priA* y UTB1-8 fueron secuenciados usando Miseq de illumina servicio disponible en Argonne Cass Ave, Lemont, IL 60439, United States Chicago y la University of New South Wales en Australia. La secuenciación de genomas por Miseq genera más de 24 millones de lecturas o reads con una longitud de 150 pb hasta 300 pb cada uno, lo que genera de 3.7 hasta 7.2 GB de datos disponibles por corrida (esto es por una sola línea). Se utilizaron dos líneas de secuenciación por genoma con una cobertura de aproximadamente de 200x.

Los reads obtenidos de cada línea en la secuenciación fueron filtrados utilizando el programa Trimmomatic (Bolger *et al.* 2014). Este programa se ha utilizado para recortar los reads obtenidos de los métodos de secuenciación de última generación. Debido a que este tipo de secuenciación genera datos pareados (pair-end) que conservan los adaptadores utilizados, es necesario eliminar dichos adaptadores sin alterar las secuencias obtenidas (Li *et al.* 2013). Trimmomatic fue ejecutado en línea de comandos utilizando los parámetros por default y especificando un LEADING:26 TRAILING:26 MINLEN:150.

##### 6.1.1.2. ENSAMBLE DE GENOMAS

Los genomas se ensamblaron de novo utilizando dos algoritmos: Velvet y SPAdes. Velvet es un paquete de algoritmos que ha sido diseñado para ensamblar genomas de novo con reads cortos. El algoritmo se basa en el uso de los gráficos de Brujin para ensamblar reads cortos. Específicamente Velvet representa cada *k-mer* diferente que se obtiene de cada read por un único nodo en la

gráfica. Los nodos se conectan si el *k-mer* tiene un solapamiento de *k-1* (Zerbino & Birney 2008). Todos los genomas se ensamblaron con Velvet utilizando kmers desde 40-180. Una vez que se obtuvieron los datos se analizó la calidad del ensamblado con FastQC (Andrews 2010). Y se seleccionaron aquellos genomas ensamblados en donde el número de contigs, el número de reads y el tamaño del ensamblado fueran los más grandes. Para verificar que seleccionamos el mejor ensamblado, utilizamos un segundo algoritmo llamado SPAdes. Éste usa los k-mers para construir también gráficos de Bruijn pero, posteriormente desarrolla operaciones teórico-gráficas, las cuales se basan en la estructura del gráfico, cobertura y tamaño de las secuencias. Así mismo ajusta los errores iterativamente (Bankevich *et al.* 2012). De igual forma los resultados obtenidos fueron analizados utilizando FastQC.

#### 6.1.1.3. ANOTACIÓN DE GENOMAS

Los genomas secuenciados en este trabajo y los que se tomaron de la base de datos del NCBI, fueron anotados automáticamente utilizando una anotación rápida que usa la tecnología basada en subsistemas llamada RASTtk (Brettin *et al.* 2015). RAST es un anotador de genomas bacterianos (2008). Su base de datos contempla genomas que han sido manualmente curados y que se encuentran disponibles en SEED (Overbeek *et al.* 2005). La consistencia y precisión de una anotación hecha con RAST es el resultado de que cada anotación disponible en SEED ha sido cuidadosamente estructurada y organizada dentro de un subsistema (conjunto de roles funcionales relacionados lógicamente).

#### 6.1.2 OBTENCIÓN DEL GENOMA NÚCLEO

Para obtener el genoma núcleo o grupos de genes ortólogos entre los organismos seleccionados como Actinobacterias basales (Tabla 3) así como los provenientes de la familia *Actinomycetaceae* (33 cepas) y el grupo externo conformado por ocho cepas de Bifidobacteriales y una de *Beutenbergia cavernae* (Tabla 4), se utilizó la anotación hecha en RASTtk (Brettin *et al.* 2015).

Una vez que los genomas fueron anotados en esta plataforma, se descargaron los archivos en formato .fasta conteniendo sólo las secuencias de aminoácidos. Posteriormente, para obtener los mejores Hits Bidireccionales (BBH, por sus siglas en inglés Best Bidirectional Hits) se utilizó el algoritmo de BLASTp para hacer un Blastp de todas los genomas contra todos los genomas. Para correr el análisis se utilizó un *e-value* de 0.001. Una vez obtenidos los resultados del BLAST all-versus-all, se extrajeron los BBH utilizando un algoritmo desarrollado por Selém *et al.* (sin publicar).

Familia	Organismos	Contenido de G+C	Tamaño del genoma
<i>Acidothermaceae</i>	<i>Acidothermus cellulolyticus</i> 11B ATCC 43068 NC_008578.1	66.82213896	2.184312
<i>Actinomycetaceae</i>	<i>Actinomyces cardiffensis</i> F0333 AQHZ01	61.88293283	1.840926
<i>Actinomycetaceae</i>	<i>Actinomyces dentalis</i> DSM 19115 AUBL01	73.65597279	2.907084
<i>Actinomycetaceae</i>	<i>Actinomyces massiliensis</i> 4401292 AKIO01	68.54350816	2.826504
<i>Actinomycetaceae</i>	<i>Actinomyces neuii</i> BVS029A5	56.6	2.14371
<i>Actinomycetaceae</i>	<i>Actinomyces odontolyticus</i> ATCC 17982 AAYI02	66.285791	2.08572
<i>Actinomycetaceae</i>	<i>Actinomyces oris</i> MGI	68	2.9
<i>Actinomycetaceae</i>	<i>Actinomyces sp. oral taxon 848 str.</i> F0332	64.4	2.52042
<i>Actinomycetaceae</i>	<i>Actinomyces viscosus</i> C505 ACRE02	69.4836183	2.579952
<i>Actinomycetaceae</i>	<i>Mobiluncus mulieris</i> ATCC 35243 ACKW01	55.904049	2.094444
<i>Actinomycetaceae</i>	<i>Trueperella pyogenes</i> MS249 JALQ01	60.56750957	1.897131
<i>Actinomycetaceae</i>	<i>Varibaculum cambriense</i> DSM 15806 ATUF01	53.96752331	1.798767
<i>Beutenbergiaceae</i>	<i>Beutenbergia cavernae</i> DSM 12333 NC_012669.1	73.2195077	4.309047
<i>Bifidobacteriaceae</i>	<i>Bifidobacterium adolescentis</i> DSM	60.38452799	1.80429

	20087 JNKM01		
<b>Bifidobacteriaceae</b>	<i>Bifidobacterium bifidum</i> NCIMB 41171 AKCA01	63.72154989	1.883253
<b>Bifidobacteriaceae</b>	<i>Bifidobacterium dentium</i> ATCC 27678 ABIX02	59.2829391	2.304351
<b>Bifidobacteriaceae</b>	<i>Bifidobacterium longum</i> subsp. <i>infantis</i> CDPT01	60.92630147	2.128767
<b>Brevibacteriaceae</b>	<i>Brevibacterium album</i> DSM 18261 AUFJ01	71.25341382	3.655728
<b>Brevibacteriaceae</b>	<i>Brevibacterium casei</i> S18 AMSP01	68.6138938	3.262434
<b>Brevibacteriaceae</b>	<i>Brevibacterium linens</i> JTJZ01	64.83417638	3.418572
<b>Brevibacteriaceae</b>	<i>Brevibacterium massiliense</i> 5401308 CAJD01	62.75778861	2.081211
<b>Cellulomonadaceae</b>	<i>Cellulomonas</i> sp. HZM JEOE01	73.1080039	3.212097
<b>Cellulomonadaceae</b>	<i>Cellulomonas</i> sp. KRMCY2 JAGF01	71.91227699	3.999954
<b>Cellulomonadaceae</b>	<i>Cellulomonas</i> sp. URHD0024 AUEW01	71.56677783	4.03899
<b>Cellulomonadaceae</b>	<i>Cellulomonas</i> sp. URHE0023 JIAN01	71.23946295	3.817839
<b>Cellulomonadaceae</b>	<i>Oerskovia turbata</i> JOFV01	72.39904689	3.57567
<b>Cellulomonadaceae</b>	<i>Paraoerskovia marina</i> DSM 21750 JNIY01	71.18709919	2.687088
<b>Coriobacteriales</b>	<i>Adlercreutzia equolifaciens</i> DSM 19450 NC_022567.1	64.51624125	2.472255
<b>Coriobacteriales</b>	<i>Atopobium minutum</i> 10063974 AGXC01	49.53765877	1.521279
<b>Coriobacteriales</b>	<i>Atopobium parvulum</i> DNF00906 NC_0132031	46.31059482	1.378542
<b>Coriobacteriales</b>	<i>Atopobium vaginae</i> DSM 15829 ACGK02	43.25505229	1.225584
<b>Coriobacteriales</b>	<i>Coriobacterium glomerans</i> PW2 NC_015389.1	60.99164179	1.865472
<b>Coriobacteriales</b>	<i>Cryptobacterium curtum</i> DSM 15641 NC_0131701	51.32089463	1.410597
<b>Coriobacteriales</b>	<i>Eggerthella lenta</i> DSM 2243 NC_0132041	64.93758935	3.185514
<b>Coriobacteriales</b>	<i>Enterorhabdus caecimuris</i> B7	65.1759921	2.479884



	<i>ASSY01</i>		
<b><i>Coriobacteriales</i></b>	<i>Gordonibacter NC_0210211</i>	65.78316002	2.524184
<b><i>Coriobacteriales</i></b>	<i>Olsenella uli DSM 7084 NC_0143631</i>	65.35579889	1.76715
<b><i>Coriobacteriales</i></b>	<i>Slackia heliotrinireducens DSM 20476 NC_0131651</i>	61.06498695	2.722362
<b><i>Demequinaceae</i></b>	<i>Lysinimicrobium mangrovi BBLU01</i>	71.96748206	2.696604
<b><i>Dermabacteraceae</i></b>	<i>Brachybacterium muris UCD-AY4 AORC01</i>	70.35559396	2.876708
<b><i>Dermabacteraceae</i></b>	<i>Brachybacterium paraconglomeratum LC44 AGSO01</i>	72.4850615	3.344714
<b><i>Dermabacteraceae</i></b>	<i>Brachybacterium phenoliresistens W13A50 JDYK01</i>	73.191617	3.760611
<b><i>Dermabacteraceae</i></b>	<i>Brachybacterium squillarum M-6-3 AGBX01</i>	73.00723979	2.803674
<b><i>Dermabacteraceae</i></b>	<i>Dermabacter hominis 1368 JDRS01</i>	63.14967995	2.217438
<b><i>Dermabacteraceae</i></b>	<i>Dermabacter sp. HFH0086 ATFO01</i>	63.18108963	2.10255
<b><i>Dermacoccaceae</i></b>	<i>Demetria terragena DSM 11295 AQXW01</i>	64.88767474	3.301083
<b><i>Dermacoccaceae</i></b>	<i>Dermacoccus nishinomiyaensis strain M25 NZ_CP008889.1</i>	69.46030483	2.639175
<b><i>Dermacoccaceae</i></b>	<i>Kytococcus sedentarius DSM 20547 NC_013169.1</i>	71.88410522	2.465961
<b><i>Dermatophilaceae</i></b>	<i>Austwickia chelonae NBRC 105200 BAGZ01</i>	66.67688213	3.063981
<b><i>Dermatophilaceae</i></b>	<i>Dermatophilus congolensis DSM 44180 = NBRC 105199 AUCS01</i>	59.75798355	2.314884
<b><i>Dermatophilaceae</i></b>	<i>Mobilicoccus sp. SIT2 CDGT01</i>	70.62669556	3.387147
<b><i>Intrasporangiaceae</i></b>	<i>Arsenicococcus bolidensis DSM 15745 AUF01</i>	72.16620659	3.507057
<b><i>Intrasporangiaceae</i></b>	<i>Intrasporangium oryzae NRRL B-24470 AWSA01</i>	71.55868437	4.071408
<b><i>Intrasporangiaceae</i></b>	<i>Janibacter sp. HTCC2649 AAMN01</i>	68.74705135	3.859815
<b><i>Intrasporangiaceae</i></b>	<i>Knoellia subterranea KCTC 19937 AVPK01</i>	69.31698902	3.224853
<b><i>Intrasporangiaceae</i></b>	<i>Ornithinimicrobium pekingense</i>	73.27574053	3.475434

	<i>DSM 21552 ATWJ01</i>		
<b><i>Intrasporangiaceae</i></b>	<i>Phycoccus jejuensis JOEE01</i>	73.94143071	3.375216
<b><i>Intrasporangiaceae</i></b>	<i>Serinicoccus marinus DSM 15273 ATWM01</i>	72.6668952	3.028068
<b><i>Intrasporangiaceae</i></b>	<i>Terrabacter sp. 28 JYOE01</i>	71.77049545	4.351844
<b><i>Intrasporangiaceae</i></b>	<i>Terracoccus sp. 273MFTsu31 ARTT01</i>	71.74330307	4.426356
<b><i>Jonesiaceae</i></b>	<i>Jonesia quinghaiensis DSM 15701 AUHN01</i>	57.29348181	2.689182
<b><i>Kineosporaceae</i></b>	<i>Kineococcus radiotolerans SRS30216 = ATCC BAA-149 NC_0096642</i>	74.63445585	4.225413
<b><i>Kineosporaceae</i></b>	<i>Kineosporia aurantiaca JCM 3230 JODP01</i>	70.11536876	7.976856
<b><i>Microbacteriaceae</i></b>	<i>Agreia bicolorata JYFC01</i>	65.69361799	3.541755
<b><i>Microbacteriaceae</i></b>	<i>Agrococcus lahaulensis DSM 17612 AULD01</i>	72.81290391	2.331882
<b><i>Microbacteriaceae</i></b>	<i>Agromyces italicus DSM 16388 ATXF01</i>	70.50402528	3.368958
<b><i>Microbacteriaceae</i></b>	<i>Agromyces subbeticus DSM 16689 ATXG01</i>	69.46262292	3.906177
<b><i>Microbacteriaceae</i></b>	<i>Candidatus Aquiluna sp. IMCC13023 AJKR01</i>	51.87663116	1.266738
<b><i>Microbacteriaceae</i></b>	<i>Candidatus Rhodoluna lacicola NZ_CP0074901</i>	51.80033271	1.339308
<b><i>Microbacteriaceae</i></b>	<i>Clavibacter michiganensis subsp. michiganensis NCPPB 382 NC_0094801</i>	73.01272627	2.881284
<b><i>Microbacteriaceae</i></b>	<i>Clavibacter michiganensis subsp. nebraskensis NCPPB 2581 NC_0208911</i>	73.30883647	2.552163
<b><i>Microbacteriaceae</i></b>	<i>Clavibacter michiganensis subsp. sepedonicus NC_010407.1</i>	72.91362222	2.843289
<b><i>Microbacteriaceae</i></b>	<i>Cryocola sp. 340MFSHa31 ARTB01</i>	70.24340667	3.727137
<b><i>Microbacteriaceae</i></b>	<i>Glaciibacter superstes DSM 21135 ATWH01</i>	64.75622663	4.302726
<b><i>Microbacteriaceae</i></b>	<i>Gryllotalpicola ginsengisoli DSM 22003 AULK01</i>	71.25357249	2.677404

<i>Microbacteriaceae</i>	<i>Herbiconiux</i> sp. YR403 JUGL01	62.23288944	3.254832
<i>Microbacteriaceae</i>	<i>Humibacter albus</i> DSM 18994 ATXT01	68.25086635	3.28677
<i>Microbacteriaceae</i>	<i>Knoellia subterranea</i> KCTC 19937 AVPK01	69.31698902	3.224853
<i>Microbacteriaceae</i>	<i>Leifsonia</i> sp. 109 AQYK01	70.16174021	3.676575
<i>Microbacteriaceae</i>	<i>Leucobacter chironomi</i> DSM 19883 ATXU01	70.36147305	2.633751
<i>Microbacteriaceae</i>	<i>Leucobacter salsicius</i> M1-8 AOCN01	65.21294127	2.799199
<i>Microbacteriaceae</i>	<i>Leucobacter</i> sp. UCD-THU APJM01	70.69582116	2.954193
<i>Microbacteriaceae</i>	<i>Microbacterium hominis</i> JWSZ01	69.68040455	2.624349
<i>Microbacteriaceae</i>	<i>Microbacterium paraoxydans</i> 77MFTsu32 AQYI01	69.84543066	3.166734
<i>Microbacteriaceae</i>	<i>Microbacterium profundum</i> JPSY01	66.89826535	3.051399
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. 11MF ARTJ01	70.10402739	2.696982
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. 292MF AQYD01	70.89822291	3.750795
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. CF335 JUGQ01	71.00474943	2.755065
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. CH12i JHET01	64.09304956	3.009966
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. G3 CDAR01	69.53931953	3.332657
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. KROC2 JAGG01	70.80606829	2.435811
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. TS-1 BASQ01	70.39780757	3.145365
<i>Microbacteriaceae</i>	<i>Microbacterium</i> sp. UCD-TDU AOSO01	68.7308725	3.414259
<i>Microbacteriaceae</i>	<i>Mycetocola saprophilus</i> JOEB01	67.02311668	2.967597
<i>Microbacteriaceae</i>	<i>Mycetocola saprophilus</i> JOEC01	67.01211875	2.972337
<i>Microbacteriaceae</i>	<i>Pseudoclavibacter soli</i> DSM 23366 AUIC01	64.66375787	1.99395
<i>Microbacteriaceae</i>	<i>Salinibacterium</i> sp. PAMC 21357 AHWA01	60.38724534	2.833191
<i>Micrococcaceae</i>	<i>Arthrobacter gangotriensis</i> Lzly AOCK01	63.71592034	3.647484

<i>Micrococcaceae</i>	<i>Arthrobacter sanguinis</i> DSM 21259 JIAG01	62.37476655	2.608728
<i>Micrococcaceae</i>	<i>Arthrobacter</i> sp. MA-N2 AQR101	63.54786944	4.193187
<i>Micrococcaceae</i>	<i>Arthrobacter</i> sp. W1 JWMD01	63.00218684	3.281448
<i>Micrococcaceae</i>	<i>Citricoccus</i> sp. CH26A AFXQ01	71.27744814	3.047748
<i>Micrococcaceae</i>	<i>Micrococcus luteus</i> JFYQ01	73.23104374	2.285472
<i>Micrococcaceae</i>	<i>Nesterenkonia</i> sp. AN1 JEMO01	68.04932928	2.626189
<i>Micrococcaceae</i>	<i>Renibacterium salmoninarum</i> NC_010168.1	56.64448844	2.823039
<i>Micrococcaceae</i>	<i>Rothia dentocariosa</i> M567 ADDW01	55.01445785	2.122722
<i>Micrococcaceae</i>	<i>Yaniella halotolerans</i> DSM 15476 ATVQ01	56.11262916	2.502123
<i>Nocardiodaceae</i>	<i>Aeromicrobium marinum</i> DSM 15272 ACLF03	71.40321427	2.791548
<i>Nocardiodaceae</i>	<i>Aeromicrobium massiliense</i> JC14 CAHG01	72.83243017	3.042624
<i>Nocardiodaceae</i>	<i>Marmoricola</i> sp. URHB0036 AUEP01	70.22313572	4.387554
<i>Nocardiodaceae</i>	<i>Nocardioides alkalitolerans</i> DSM 16699 AUFN01	73.47474634	4.313889
<i>Nocardiodaceae</i>	<i>Nocardioides halotolerans</i> DSM 19273 AUGT01	72.32941356	4.388259
<i>Nocardiodaceae</i>	<i>Nocardioides insulae</i> DSM 17944 AUGU01	70.90570255	3.647787
<i>Nocardiodaceae</i>	<i>Nocardioides</i> sp. Iso805N AQOZ01	70.29653432	4.031304
<i>Nocardiodaceae</i>	<i>Nocardioides</i> sp. UNC345MFTsu51 JQJL01	72.39836564	4.690023
<i>Nocardiodaceae</i>	<i>Nocardioides</i> sp. URHA0020 JIAR01	72.02298497	4.210056
<i>Nocardiodaceae</i>	<i>Nocardioides</i> sp. URHA0032 JIAY01	71.97571656	4.15839
<i>Promicromonospora</i>	<i>Cellulosimicrobium cellulans</i> J36 JAGJ01	74.88868474	3.637866
<i>Promicromonospora</i>	<i>Isoptericola variabilis</i> 225 NC_015588.1	74.12391496	2.965437
<i>Promicromonospora</i>	<i>Promicromonospora sukumoe</i> 327MFSha3.1 ARQM01	73.26164159	5.647512

<b><i>Promicromonospora</i></b>	<i>Xylanimonas cellulositytica</i> DSM 15894 NC_013530.1	72.80375945	3.387516
<b><i>Propionibacteriaceae</i></b>	<i>Aestuariimicrobium kwangyangense</i> DSM 21549 ATXE01	68.65666539	2.720052
<b><i>Propionibacteriaceae</i></b>	<i>Microlunatus phosphovorius</i> NM-1 NC_0156351	67.71145435	5.019393
<b><i>Propionibacteriaceae</i></b>	<i>Propionibacterium acidipropionici</i> ATCC 4875 NC_0193951	69.19193702	3.228882
<b><i>Propionibacteriaceae</i></b>	<i>Propionibacterium acnes</i> KPA171202 NC_0060851	60.33744109	2.267418
<b><i>Propionibacteriaceae</i></b>	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i> CIRM-BIAI NC_0142151	67.48279525	2.262021
<b><i>Propionibacteriaceae</i></b>	<i>Propionibacterium humerusii</i> P08 AFAM01	60.34880375	2.283003
<b><i>Propionibacteriaceae</i></b>	<i>Propionibacterium propionicum</i> F0230a NC_0181421	66.7547847	3.014139
<b><i>Propionibacteriaceae</i></b>	<i>Propionimicrobium lymphophilum</i> ACS-093-V-SCH5 AGZR01	56.7354316	2.016144
<b><i>Ruaniaceae</i></b>	<i>Ruania albidiflava</i> DSM 18029 ATWL01	70.40018658	3.721689
<b><i>Sanguibacteraceae</i></b>	<i>Sanguibacter keddieii</i> DSM 10542 NC_0135211	72.15891421	3.81363
<b><i>Timonella</i></b>	<i>Timonella senegalensis</i> JC301 CAHH01	61.97217071	2.549688
<b><i>Unclassified Micrococcaceae</i></b>	<i>Tropheryma whipplei</i> str. DigMusc17 CAVA01	46.42506237	0.751118

**Tabla 3.** Genomas de *Actinomycetaceae* utilizados en este estudio.

Identificador en RAST	Taxon ID	Organismos	Tamaño del genoma	Contenido de G+C
471853.12	Org_1	<i>Beutenbergia cavernae</i> DSM 12333	4.669183	73
561180.12	Org_2	<i>Bifidobacterium gallicum</i> DSM 20093 = LMG 11596	2.0198	57.5
1207543.7	Org_3	<i>Bifidobacterium bifidum</i> IPLA 20015	2.13833	62.7
391904.18	Org_4	<i>Bifidobacterium longum</i> subsp. <i>infantis</i> ATCC 15697 = JCM 1222 = DSM 20088	2.83	59.9
518634.16	Org_5	<i>Bifidobacterium breve</i> DSM 20213 = JCM 1192	2.33139	58.5
473819.17	Org_6	<i>Bifidobacterium dentium</i> ATCC 27678	2.64208	58.5
411481.14	Org_7	<i>Bifidobacterium adolescentis</i> L2-32	2.38911	59.3
566552.15	Org_8	<i>Bifidobacterium catenulatum</i> DSM 16992 = JCM 1194 = LMG 11043	2.05843	56.1
547043.15	Org_9	<i>Bifidobacterium pseudocatenulatum</i> DSM 20438 = JCM 1200	2.28	56.4
666666.811	Org_10	<i>Trueperella pyogenes</i>	2.33839	59.5
649743.7	Org_11*	<i>Actinomyces</i> sp. oral taxon 848 str. F0332	2.52042	64.4
883067.5	Org_12	<i>Actinobaculum schaalii</i>	2.09	61.1
883066.5	Org_13	<i>Actinobaculum massiliae</i> ACS-171-V-Col2	2.02	60.2
1321817.5	Org_14	<i>Actinomyces graevenitzii</i> F0530	2.09095	57.7
521393.6	Org_15	<i>Actinomyces timonensis</i> DSM 23838	2.91	71.2
525246.7	Org_16*	<i>Actinomyces urogenitalis</i> DSM 15434	2.60442	68.4
1120941.6	Org_17	<i>Actinomyces dentalis</i> DSM 19115	3.53	73.1
888056.7	Org_18	<i>Actinomyces</i> sp. oral taxon 448 str. F0400	2.82817	69.7
1120943.5	Org_19	<i>Actinomyces gerencseriae</i> DSM 6844	3.42	70.7
1167628.6	Org_20	<i>Actinomyces massiliensis</i> 4401292	3.37103	67.8
1278298.5	Org_21	<i>Actinomyces slackii</i>	3.17	70.1
666666.814	Org_22*	<i>Actinomyces oris</i> MGI	2.9	68
706439.7	Org_23	<i>Actinomyces</i> sp. oral taxon 171 str. F0337	3.00625	68
653386.7	Org_24	<i>Actinomyces</i> sp. oral taxon 849 str. F0330	3.45937	67.1
762963.7	Org_25	<i>Actinomyces</i> sp. oral taxon 170 str. F0386	3.13514	67.8
585198.6	Org_26	<i>Mobiluncus curtisii</i> subsp. <i>curtisii</i> ATCC 35241	2.14	55.7

<b>871571.6</b>	Org_27	<i>Mobiluncus mulieris ATCC 35239</i>	2.43	55
<b>1120945.5</b>	Org_28	<i>Actinomyces neuui BVS029A5</i>	2.14371	56.6
<b>1123488.6</b>	Org_29	<i>Varibaculum cambriense DSM 15806</i>	2.02	53.4
<b>525245.7</b>	Org_30	<i>Actinomyces coleocanis</i>	1.72	49.6
<b>883069.5</b>	Org_31	<i>Actinomyces europaeus</i>	1.91	56.6
<b>1120946.5</b>	Org_32	<i>Actinomyces suimastitidis DSM 15538</i>	2.29	56.4
<b>1118058.5</b>	Org_33	<i>Actinomyces sp. ph3</i>	1.86418	56
<b>1127690.6</b>	Org_34	<i>Actinomyces sp. oral taxon 181 str. F0379</i>	1.90672	56
<b>1120947.6</b>	Org_35	<i>Actinomyces vaccimaxillae</i>	2.34	57.6
<b>883077.6</b>	Org_36	<i>Actinomyces turicensis ACS-279-V-Col4</i>	1.95	57.2
<b>888050.7</b>	Org_37	<i>Actinomyces cardiffensis F0333</i>	2.19	61.5
<b>1125717.6</b>	Org_38	<i>Actinomyces georgiae F0490</i>	2.47736	69.8
<b>888051.7</b>	Org_39	<i>Actinomyces sp. oral taxon 178 str. F0338</i>	2.81711	70.7
<b>888052.6</b>	Org_40	<i>Actinomyces sp. oral taxon 180 str. F0310</i>	2.37266	66.3
<b>1105029.6</b>	Org_41	<i>Actinomyces sp. ICM39</i>	2.26766	65.5
<b>411466.12</b>	Org_42	<i>Actinomyces odontolyticus ATCC 17982</i>	2.39396	65.4

**Tabla 4.** Genomas del grupo Actinobacteria seleccionados para los análisis filogenómicos del grupo de las Actinobacterias basales.

### 6.1.3 ESTIMACIÓN DE TASAS DE SUSTITUCIÓN EN SECUENCIAS DE PROTEÍNAS

#### 6.1.3.1 ALINEAMIENTO DE NUCLEÓTIDOS

Para estimar las tasas de sustitución de las proteínas que componen el genoma núcleo y algunas proteínas seleccionadas como PriA se extrajeron las secuencias de nucleótidos de los genes de la plataforma de RAST en formato fasta. Una vez que se obtuvieron las secuencias de nucleótidos, se alinearon mediante el uso del programa RevTrans v. 1.4. Este programa construye múltiples alineamientos de ADN mediante tres pasos: i) las secuencias de DNA son traducidas; ii) se hace un alineamiento con las secuencias traducidas; iii) se construye un múltiple alineamiento de ADN

mediante una traducción inversa de la proteína alineada (Wernersson y Gorm 2003). Los alineamientos fueron guardados en formato .phy (Apéndice 1).

#### 6.1.3.2. CÁLCULO DE LAS TASAS DE SUSTITUCIÓN UTILIZANDO CODEML

Las tasas de sustitución para cada conjunto de genes  $dN/dS$  se calculó utilizando el paquete de programas disponibles en PAML (para Análisis filogenéticos por máxima verosimilitud)(Yang 2004). Utilizamos el programa codeml. Este programa hace un análisis de máxima verosimilitud de secuencias de nucleótidos basándose en la implementación de los modelos de sustitución de codones de Golman y Yang 1994. Para cada conjunto de genes utilizamos el modelo de ramas, en donde a partir de una filogenia dada marcamos las ramas que se consideran están bajo una tasa de sustitución diferente. Codeml fue ejecutado en línea de comandos cuatro veces, la primera vez para estimar el valor de  $kappa$ , que corresponde a las tasas de transversiones/transiciones para cada una de las tres posiciones que contiene un codon. En la segunda se fijó el valor de  $kappa$  y se supuso una tasa de sustitución igual para todas las ramas ( $H_0$ ). La tercera corrida consistió en marcar las ramas con diferentes tasas de sustitución ( $H_1$ ). En este caso el marcaje de las ramas se hizo basándose en los linajes propuestos en este trabajo. Los parámetros y árboles utilizados se presentan a continuación, en este caso tomamos como ejemplo el gen 1, pero para los 205 genes restantes se utilizaron los mismos parámetros (Apéndice 2).

#### 6.1.3.3. ANÁLISIS DE MÁXIMA VEROSIMILITUD (LRT)

Una vez obtenidas las tasas de sustitución para cada rama (valores de omega), se hizo una prueba de hipótesis de Chi2 con un intervalo de confianza del 95% para decidir si se aceptaba o rechazaba la  $H_0$ .

Las hipótesis que manejamos son las siguientes:

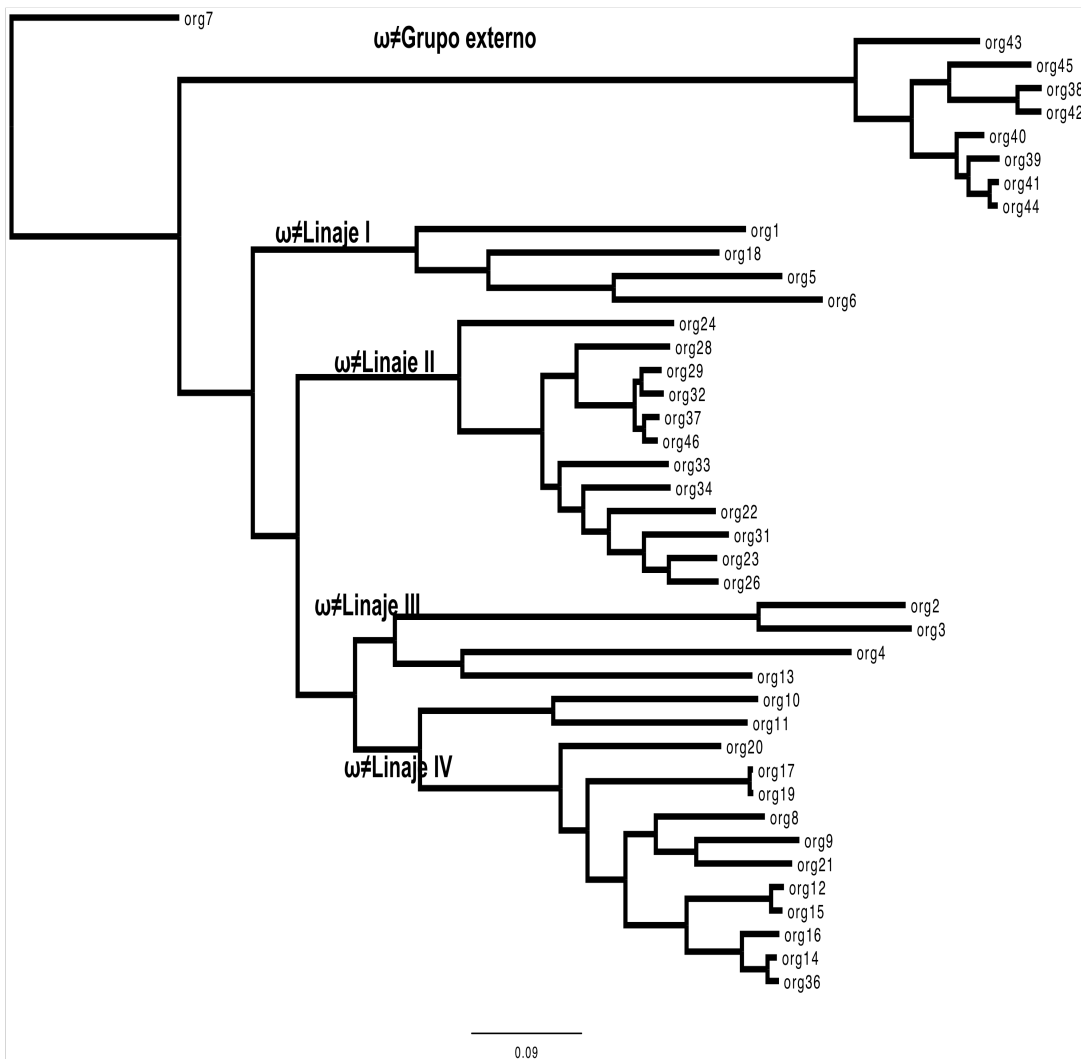


### HIPÓTESIS NULA

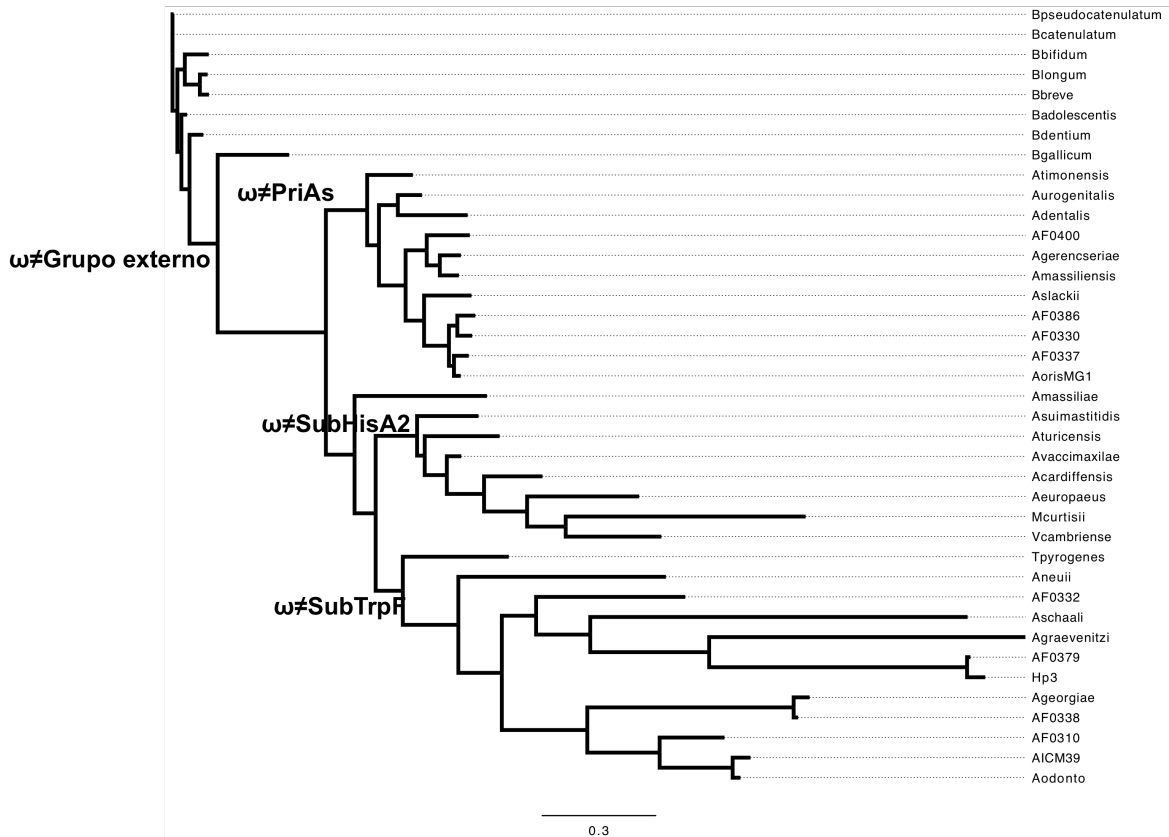
H0: Todas las ramas acumularon las mismas tasas de sustitución, por lo que la presión de selección es similar entre los diferentes linajes evolutivos.

### HIPÓTESIS ALTERNATIVA

H1: Cada rama que dio origen a los diferentes linajes presenta tasas de sustitución diferentes, por lo que diferentes presiones de selección dieron origen a los diferentes linajes detectados dentro de la familia *Actinomycetaceae*.



**Figura 12.** Filogenia utilizada para el cálculo de las tasas de sustitución en cada rama de los 205 genes que forman parte del genoma núcleo. Las marcas en cada rama muestran las particiones que se utilizaron en la cuarta corrida (H1) dentro del programa codeml.



**Figura 13.** Filogenia utilizada para el cálculo de las tasas de sustitución en cada rama del gen *priA* y sus homólogos. Las marcas en cada rama muestran las particiones que se utilizaron en la cuarta corrida (H1) dentro del programa codeml.

#### 6.1.4 ANÁLISIS FILOGENÓMICOS

Se llevaron a cabo dos tipos de análisis filogenéticos, uno con la secuencia de las proteínas concatenadas y otro basado en grupos de ortólogos individuales. Para todos los casos se utilizó el algoritmo Bayesiano y el de máxima verosimilitud implementado en MrBayes v. 3.2.1 y RaxML v.8 respectivamente (Ronquist *et al.* 2012; Stamatakis *et al.* 2014). En ambos casos se seleccionó el modelo por el ajuste más alto a los criterios BIC o de Akaike dependiendo del algoritmo. Utilizando ProtTest v1.4 (Abascal *et al.* 2005) se seleccionó el mejor modelo de ajuste y este se utilizó para construir el análisis. ProtTest es un programa que permite identificar de entre 97 modelos de sustitución diferentes el más adecuado para las secuencias de aminoácidos basándose ya sea en el criterio de información de Akaike (para ML) o BIC (para Mr. Bayes). A partir de estos resultados se

seleccionó el mejor modelo para los genes individuales y para el árbol concatenado se eligió el modelo GTR+I+G (Tabla 5). La probabilidad posterior para cada rama utilizando el algoritmo de Mr. Bayes fue estimada usando 1 000 000 de generaciones en cada análisis. En el caso del algoritmo de ML el soporte estadístico de cada rama fue calculado utilizando un bootstrap de 1000 réplicas.

#### 6.1.4.1 ANÁLISIS FILOGENÓMICO DE SECUENCIAS CONCATENADAS

Todas las secuencias de proteínas obtenidas del uso de los BBH se alinearon individualmente utilizando el programa MUSCLE 3.8.31 (Edgar 2004). Posteriormente, dado que las secuencias serían utilizadas para hacer el análisis filogenómico y de genes individuales, se optó por editar los alineamientos utilizando el programa GBLOCKS (Castresana 2000). Este programa elimina posiciones que no se alinean bien así como posibles bloques de regiones divergentes, estas posiciones podrían no ser homólogas o podrían haber llegado al límite de sustituciones múltiples y es conveniente eliminarlas antes de los análisis filogenéticos (Talavera & Castresana 2007).

#### 6.1.5. ANÁLISIS DEL CONTEXTO GENÓMICO

Para obtener el contexto genómico de los genes *his/trp* de la familia *Actinomycetaceae* se hizo una búsqueda por homología utilizando el recurso Blastp que se encuentra disponible dentro de RAST. Se utilizó como semilla la secuencia de los genes *his/trp* incluyendo *priA* del organismo *A. oris*. Con la información de la localización genómica se hizo un esquema.

#### 6.1.6 PRUEBAS ESTADÍSTICAS PARA DETECTAR DIFERENCIAS ENTRE LINAJES

Para encontrar diferencias entre el contenido-(G + C), el tamaño del genoma, el número de secuencias codificantes (CDS) y el contenido de subsistemas entre los linajes sugeridos de la familia *Actinomycetaceae* se utilizó la prueba de rangos con signo de Wilcoxon en el software R. La prueba de rangos de Wilcoxon es una prueba no paramétrica para la comparación de la mediana

entre dos muestras relacionadas y determinar si existen diferencias significativas entre ellas (Wilcoxon 1945). La utilizamos como una alternativa a la prueba de T-test puesto que no pudimos asumir una normalidad en dichas muestras. El conjunto de datos se obtuvo de la plataforma RAST y se lista en la Tabla 4.

### 6.1.7. ANÁLISIS DE PÉRDIDAS Y GANANCIAS DE GENES

Para el análisis de pérdidas y ganancias de genes, se utilizó el algoritmo What Changed disponible en la plataforma RAST. Este algoritmo se basa en la construcción de una filogenia por máxima verosimilitud construida con el programa FastTree (Price *at al.* 2009). A partir de esta filogenia se extraen los genes únicos y comunes a todas las cepas (genoma núcleo) y se mapean las presencias ausencias de los genes no-core genes en cada nodo desde el más ancestral hasta el más reciente. Una vez que se obtuvieron las ganancias y pérdidas de cada nodo, se hizo un análisis para cada una de las familias de proteínas.

## 6.2 MÉTODOS DE BIOLOGÍA MOLECULAR Y BIOQUÍMICA

### 6.2.1 CLONACIÓN DE GENES HOMÓLOGOS DE *PRIA*

Los genes *priA* y *subTrpF* de los organismos *A. urogenitalis*, *A. sp. oral taxon* 848 str. F0332, *A. odontolyticus* ATCC 17982, fueron sintetizados por la compañía GeneArt (Thermo Fisher Scientific, 81 Wyman street Waltham, MA. USA). Los genes *priA*, *subHisAs* y *subTrpFs* de los organismos *A. sp. oral taxon* 171 str. F0337, *A. sp. oral taxon* 180 str. F0310, *A. viscosus*, *A. cardiffensis* F0330, *A. vaccimaxillae* DSM 15804 y *A. graevenetzi* fueron sintetizados por la compañía GeneScript (860 Centennial Ave. Piscataway, USA). En todos los casos se optó por una síntesis optimizada en el uso de codones de *E.coli* para facilitar la expresión heteróloga en este fondo genético. A todos los genes sintéticos se les añadió el sitio de restricción *NdeI* en el extremo 5' y *XhoI* y *HindIII* en el extremo 3', esto con la finalidad de facilitar la subsecuente subclonación en los vectores de expresión pASK y pET. *priA* del organismo *A. oris* MG1 fue amplificado por

medio de PCR utilizando los oligonucleótidos Ao-Rev y Ao-Fw (Tabla 5). Los genes *priA* de los organismos *B. longum*, *B. gallicum* y *B. adolescentis* fueron donados por El laboratorio Nacional Argonne. Todos los genes fueron clonados en pASK y pET22b (Novagen) utilizando los sitios *NdeI* y *HindIII*, y los oligos listados en la tabla 5, en el caso de pET28a (Novagen) se optó por utilizar los sitios *NdeI* y *XhoI* para eliminar el His6-Tag en el extremo C-Terminal.

Amplificar genes	Tm	secuencia (5'-3')
<b>priA_Org2</b> Reverse	64°C	TTTTCAAAGCTTGCCGCGCGCCA
<b>priA_Org2</b> Forward	64°C	AAAAACATATGAACGTGCTCCAAGCG
<b>priA_Org6</b> Reverse	65°C	TTTTTTCAAAGCTTGCGTGCGACTTC
<b>priA_Org6</b> Forward	65°C	AAAAACATATGTCTTGACGCTGCTTC
<b>priA_Org22</b> Reverse	65°C	CCATGGTTA AAGCTTGGAACGAGCCTGCGAAGG
<b>priA_Org22</b> Forward	65°C	CCGAATTCCATATGCTCACTCTACTTCCC
<b>Construcción de la mutante en priA</b>		
<b>priA_Org22</b> up-F	55°C	GGCGCATATGTAGAGTGAGCATGGGTGTCT GGC GGA ATT CGC CGT AGG AGT CCA GAA
<b>priA_Org22</b> up-R	55°C	GGA TA
<b>priA_Org22</b> down-F	55°C	GGCGTCTAGACGCGTCGACTCCCTGCTCCTGGTC
<b>priA_Org22</b> down-R	55°C	GGCCATATGGCTCGTTCCTGACCTTTGCC
<b>Confirmar la mutación en priA</b>		
<b>priA_Org22</b> Forward	55°C	CCGAATTCCATATGCTCACTCTACTTCCC
<b>priA_Org22</b> Reverse	55°C	CCATGGTTAAAGCTTGGAACGAGCCTGCGAAGG
<b>Sequence</b>		
<b>pasK_ACP*</b> pASKNFor	50°C	AAATGATCAATTCAAGGCC
<b>pasK_ACP*</b> pASKNRev	50°C	GCGGATTAGAAAAACAAC
<b>16s</b> R1492	50°	TAC GGY TAC CTT GTT ACG ACT T
<b>16s</b> F27	50°	AGA GTT TGA TCM TGG CTC AG

**Tabla 5. Oligos utilizados en este estudio**

\*Vector reportado en Noda-García *et al.* 2015

## 6.2.2 EXPRESIÓN HETERÓLOGA DE HOMÓLOGOS DE PRIA

Para sobre-producir las proteínas PriA\_*A. urogenitalis*, PriA\_*A.oris*, PriA\_*A. sp. oral taxon 337*, PriA\_*A. viscosus*, PriA\_*B. longum*, PriA\_*B. gallicum*, PriA\_*B. adolescentis*, SubTrpF\_*A.odontolyticus*, SubTrpF\_*A. sp. oral taxon* , SubTrpF\_*A. sp. oral taxon 332*, SubTrpF\_*A. sp. oral taxon 310*, SubTrpF\_*A. graevenitzi*, SubHisA2\_*A. cardiffensis*, SubHisA2\_*A. vacxmaxiiliae*; células BL21Magic, BL21\*, R2, c41/43 fueron transformadas con los vectores: en los vectores pET22b, pET28a y pMCSG68 conteniendo las proteínas de interés. Las células se crecieron a 37°C a 200 revoluciones por minuto en medio LB hasta alcanzar una Densidad Óptica de 600 nanómetros. Una vez que alcanzaron la densidad el cultivo fue inducido con 1 mM de Isopropil β-D-1-trigalactidopiranosido (IPTG) a una temperatura de 20°C e incubadas toda la noche. Después las células se centrifugaron a 13000 rpm durante 30 min. Una vez que se obtuvo el pellet se resuspendió en Buffer de Lisis (50 mM Tris pH 7.5, 0.5 mM de EDTA pH 8.0, 50 mM de NaCl y 10% de Glicerol).

## 6.2.3 PURIFICACIÓN DE HOMÓLOGOS DE PRIA

### 6.2.3.1 CROMATOGRAFÍA DE AFINIDAD A NÍQUEL

Una vez que las células se resuspendieron en buffer de Lisis, estas se lisaron mediante sonicación (Vibracell, Sonics) con una frecuencia de 70 kHz y duración de 20 segundos e intervalos de 20 segundos de descanso en hielo. Para obtener la fracción soluble de las proteínas de interés, el lisado se centrifugó a 13 000 rpm durante 45 min. La proteína contenida en la fracción soluble se filtró utilizando un filtro Amicon de 0.45 μM. Para la purificación de todas las proteínas se utilizaron columnas pre empacadas y cargadas con níquel His6-TrapVivapure Maxiprep-Sartorius. La cola de Histidinas provista por los vectores pET que poseen las proteínas de interés fue utilizada. La columna de Níquel fue equilibrada con Buffer 50 mM HEPES pH 7.5 y 100 mM de NaCl. La

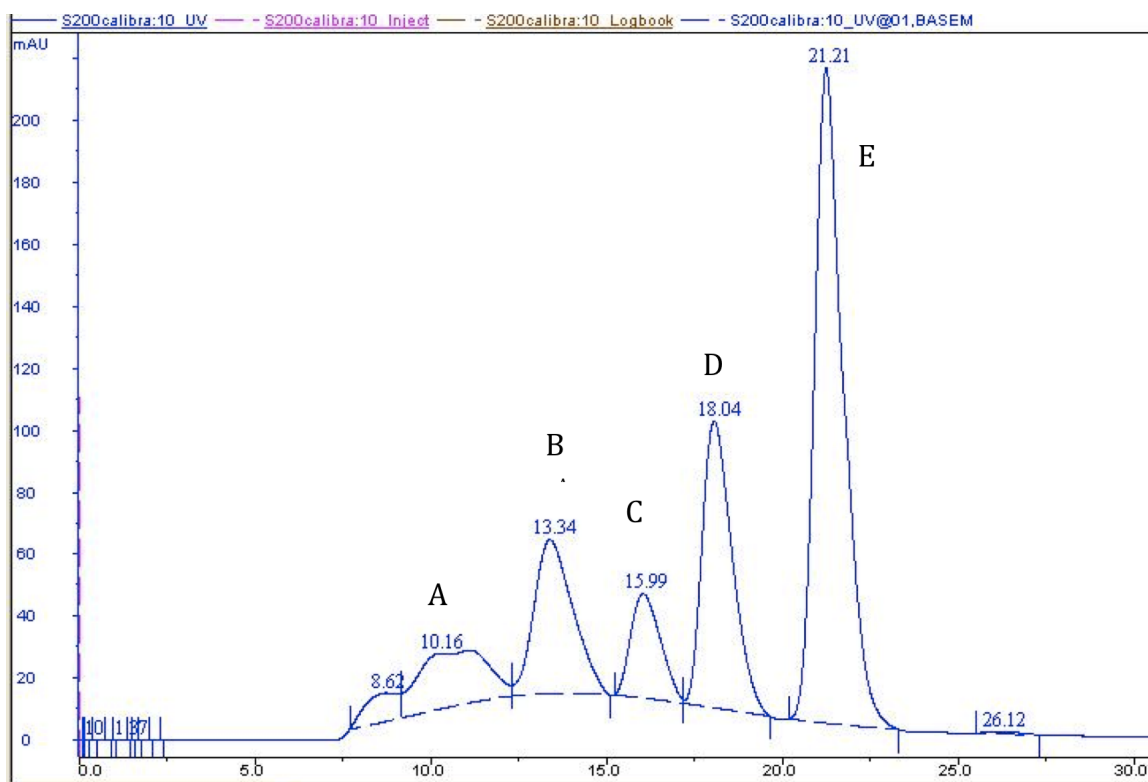
fracción que contenía la proteína de interés, fue pasada por la columna a 200 g durante 10 min. después, se hicieron lavados con del buffer que contiene 50mM HEPES pH 7.5, 100 mM NaCl y 50 mM Imidazol. La proteína se eluyó con 300 mM de Imidazol. Una vez que la proteína se eluyó, se cuantificó utilizando el método de Bradford. Finalmente en algunos casos la proteína se concentró utilizando tubos Amicon con un corte de tamaño de poro de 10 kD.

#### 6.2.3.2 PURIFICACIÓN POR EXCLUSIÓN MOLECULAR

Para determinar la estructura cuaternaria de aquellas proteínas solubles, se utilizó la purificación por exclusión molecular. Antes de pasar la proteína ésta fue purificada utilizando una columna His-Trap de 5 ml. Una vez colectada la fracción en la cual estaba contenida la proteína, se concentró por ultracentrifugación (Amicon Ultra, Millipore) hasta llegar a una concentración de 15mg/ml como mínimo. Para la purificación por exclusión molecular utilizamos una columna Superdex 200 ó S200 (GE). La columna fue equilibrada con dos volúmenes de columna del buffer (HEPES 50 mM pH 7.5, NaCl 200 mM) a un flujo de 2 ml/ min. Después la proteína fue inyectada utilizando un loop de 500 µl manteniendo un flujo constante de 0.4 ml/min, de igual forma se pasaron dos volúmenes de columna. Para determinar el peso molecular y la posible formación de estructura cuaternaria de las proteínas se utilizó como marcador de peso molecular el concentrado de Bio-RAD 151-1901 Gel Filtración Standard. La curva de calibración se muestra a continuación en la Tabla 6 y Figura 14.

	Proteína	Peso molecular	mL	lg PM
A	Thyroglobulin	670 000	10.16	5.82607
B	Bovine gamma-globulin	158 000	13.34	5.1986
C	Chicken ovalbumin	44 000	15.99	4.6434
D	Equine mioglobina	17 000	18.04	4.23044
E	Vitamin B12	1350	21.21	3.1303

**Tabla 6.** Calibradores Utilizados para la columna S200.



**Figura 14.** Perfil obtenido de la calibración de la columna S200. En el eje de las “x” se representa el volumen de elución y en el eje de las “y” la absorbancia en miliunidades. Las letras corresponden a las proteínas de la tabla 6.



## 6.2.4 ENSAYOS ENZIMÁTICOS ACOPLADOS PARA DETECTAR ACTIVIDAD DE PRA Y PROFAR ISOMERASA EN LOS HOMÓLOGOS DE PRIA

### 6.2.4.1 PURIFICACIÓN DE LAS ENZIMAS HISF, TRPC Y TRPD

El plásmido pET11c-HisF que contiene el gen que codifica para la enzima HisF de *Thermotoga marítima* fue donado por la Dra. Birte Höcker (Max Planck Institute). El gen hisF fue subclonado en el vector de expresión pET15b (Novagen) utilizando las enzimas NdeI y HindIII (NEB). La enzima TrpC también de *T. marítima* fue clonada en el plásmido pQE1. Mientras que la enzima TrpD fue clonada en el vector pET28a. Para la Expresión de la enzima HisF se utilizó la cepa de expresión *E. coli* BL21\* (DE3). Para la expresión de las enzimas TrpC y TrpD se utilizó la cepa de expresión *E. coli* C41 (Miroux y Walter 1996) transformada con el vector pGro7 (Takara). A partir de una colonia sencilla de estas diferentes cepas de expresión transformadas con los plásmidos que contienen las proteínas de interés se puso un preinóculo en 10 ml de medio LB, con el antibiótico indicado para mantener la resistencia, en el caso de las células c41\_pGro7 se le adiciona arabinosa al 2% p/v incubado a 37°C con agitación de 200 rpm durante una noche. Posteriormente con el preinóculo fue puesto a crecer en 1 litro de medio LB a 37°C y 200 rpm hasta alcanzar una OD<sub>600nm</sub>. La expresión de las proteínas se indujo con 0.5 mM de IPTG a 20°C toda la noche. Posteriormente las células fueron recolectadas por centrifugación a 5000 rpm durante 30 min. Las células obtenidas fueron lisadas por sonicación (REF) en 20 ml de buffer de lisis que contiene: 50 mM Tris-NaOH pH 7.5, 0.5 mM de EDTA pH 7.7, 50 mM de NaCl y 10% de Glicerol v/v. La sonicación se llevó a cabo incubando la muestra en hielo con tiempos de sonicación de 20 segundos con una amplitud de 70 y un tiempo de recuperación de 20 segundos. Esto se repitió 20 veces. Para obtener las proteínas HisF, TrpC y TrpD en la fracción soluble, el lisado se centrifugó a 13 000 rpm durante 45 minutos. Una vez colectada la fracción soluble, ésta se filtró utilizando filtros Millipore de 45 µm. Las proteínas fueron purificadas por afinidad a columna de Níquel, aprovechando el His6-tag en el extremo N-terminal. Para la purificación se utilizó una columna de Níquel His-Trap de 5ml (GE) y

el FPLC AKTA (GE). La columna fue equilibrada con 5 volúmenes de columna de la solución: 50 mM de HEPES pH 7.5 y 100 mM de NaCl. A un flujo constante de 2ml/ min. Se cargo la columna con la proteína de interés y para eluir la proteína se utilizó un gradiente de imidazol de 0 a 500 mM de imidazol en 40 min, utilizando el mismo amortiguador antes especificado. Todas estas enzimas eluyeron a una concentración no mayor a 200 mM de imidazol. Las fracciones conteniendo a la proteína fueron concentradas por ultracentrifugación utilizando tubos Amicon de corte de 10 kD. La concentración final de las proteínas casi siempre oscila entre 60  $\mu$ M y 80  $\mu$ M. La concentración de enzimas fue calculada por el método de Bradford. Todas las enzimas fueron puestas en 10% de Glicerol v/v y almacenadas a -80°C para su posterior utilización en los ensayos enzimáticos. Así mismo verificamos por medio de un Gel-SDS page que las fracciones colectadas correspondiera a las proteínas de interés.

#### 6.2.4.2 DETERMINACIÓN DE LOS PARÁMETROS CINÉTICOS PARA LA ACTIVIDAD DE PROFAR ISOMERASA

Para determinar la actividad de ProFAR isomerasa en los homólogos de PriA, se utilizó la metodología descrita por Sterner (1996) adaptada por el Dr. Aldo Camacho Zarco (2009). El ensayo consiste en detectar espectroscópicamente la conversión de ProFAR en en ImGP y AICAR a 300 nm de longitud de onda ( $\Delta\epsilon_{300}$  ( ProFAR - AICAR) = 5.637 mM<sup>-1</sup>cm<sup>-1</sup>) (Jurgens *et al.* 2000). Tomando en cuenta esto la actividad de ProFAR isomerasa fue determinada a 300 nm utilizando el lector TECAN infinite M1000. La reacción se llevó a cabo utilizando placas de 96 pozos “Half-area” transparentes (Corning). El amortiguador utilizado fue TES 50mM pH 7.5, NH<sub>4</sub>HCO<sub>3</sub> 200mM y la enzima HisF 2mM. En este ensayo la enzima HisF es añadida en exceso debido a que convierte PrFAR a AICAR, un paso limitante en la reacción. Siguiendo las recomendaciones del proveedor de las placas, el volumen final de reacción siempre correspondió a 170  $\mu$ l, esta cantidad equivale a la distancia de 1cm de paso de luz, por lo cual el coeficiente de extinción molar previamente reportado no tuvo que ser re-calculado. Antes de llevar a cabo el ensayo calculamos la concentración de la enzima que fuera la más adecuada para el ensayo. Para esto se varió la concentración final de

enzima ProFAR isomerasa entre 20 nM y 1 mM. La concentración óptima fue estipulada como aquella que se encontrase en la fase lineal de la relación “Concentración de enzima / Velocidad inicial”. Una vez que se obtuvo la concentración adecuada de la enzima para realizar los ensayos, se varió la concentración del sustrato (ProFAR) entre 0- 100 mM, debido a que estas enzimas presentan valores de  $K_M$  muy pobres, decidimos monitorear la reacción durante 40 min. A partir de la sección lineal de la curva del progreso de la reacción, se obtuvo una relación entre la absorbancia y el tiempo, misma que fue convertida a velocidad inicial ( $\text{mM}_{\text{ProFAR}} \text{S}^{-1}$ ) utilizando el coeficiente de extinción molar previamente reportado. Para calcular los parámetros catalíticos se graficaron las velocidades iniciales contra las respectivas concentraciones de sustrato, y se obtuvieron los ajustes no lineales a la ecuación de Michaelis – Menten usando el programa Prism v.6 (GraphPad).

#### 6.2.4.3 DETERMINACIÓN DE LOS PARÁMETROS CINÉTICOS PARA LA ACTIVIDAD DE PRA ISOMERASA

Para medir la actividad de PRA isomerasa de las enzimas homólogas a PriA se siguió la metodología descrita por el grupo del Dr. Kirschner (REF Hommel 1995), y posteriormente adaptada por la M. en C. Karina Verdel Aranda (Tesis Verdel Aranda 2009). El ensayo para detectar la actividad de PRA isomerasa de una enzima consiste en cuantificar fluorómetricamente la conversión de PRA en CdRP (Hommel 1995). Debido a que PRA no se puede obtener comercialmente y es lábil, se llevó a cabo un ensayo enzimático acoplado utilizando la enzima TrpD que es la enzima que cataliza la conversión de ácido antranílico y PRPP a PR-antranilato (PRA). Además dado que se ha observado una inhibición por producto en algunos homólogos de TrpF, se añadió la enzima río abajo TrpC esta enzima convierte el CdRP en IGP, eliminando así el posible efecto de CdRP sobre subTrpF, PriA.

Todos los ensayos enzimáticos de las enzimas caracterizadas en esta tesis fueron medidos a 310 nm de longitud de onda de excitación y 400 nm de longitud de onda de emisión (TECAN Infinite M1000) en placas de 96 pozos adecuadas para ensayos fluorométricos (Nunc). El amortiguador

utilizado fue Tris-HCl 50 mM pH 7.5, MgCl<sub>2</sub> 2 mM, EDTA 0.5 mM, ditioneitol 0.4 mM, TrpC 1 uM y TrpD 2 uM. Inicialmente se monitorea la conversión de ácido antranílico y PRPP en PRA llevada a cabo por la enzima TrpD, en un exceso 5 M de PRPP con respecto al ácido antranílico. Una vez que la curva de progreso de la reacción ha llegado al equilibrio aproximadamente en 15 minutos, se agregó a la reacción la enzima con actividad de PRA isomerasa. La concentración óptima fue estipulada como aquella que se encontrase en la fase lineal de la relación “Concentración de enzima / Velocidad inicial”.

Los parámetros catalíticos fueron obtenidos utilizando la concentración de enzima previamente calculada y variando la concentración de ácido antranílico entre 0 y 300 mM, en el amortiguador de reacción y condiciones previamente mencionadas. La reacción fue monitoreada por aproximadamente durante 40 min.

A partir de la sección lineal de la curva del progreso de la reacción, se obtuvo una relación entre la fluorescencia y el tiempo, misma que fue convertida a velocidad inicial ( $\text{mM}_{\text{ProFAR}} \text{s}^{-1}$ ) realizando una curva de calibración con los datos de fluorescencia correspondientes a los tiempos iniciales de cada concentración de ácido antranílico. Para calcular los parámetros catalíticos se graficaron las velocidades iniciales contra las respectivas concentraciones de sustrato, y se obtuvieron los ajustes no lineales a la ecuación de Michaelis – Menten usando el programa Prism v.6 (GraphPad).

Debido a la sensibilidad del ensayo, las velocidades iniciales utilizando concentraciones de PRA superiores a 300 mM fueron imposibles de detectar. Por esta razón, no se reportaron los datos de  $K_M$  y  $k_{cat}$  de enzimas con actividades bajas como la FoIE de *C. trachomatis*. En estos casos, en lugar de obtener una hipérbola característica de las enzimas con comportamiento michaeliano se obtuvo una recta cuya pendiente es una constante de velocidad de segundo orden ( $V_{max}/K_M$ ). Dado que la concentración de enzima es conocida, la eficiencia catalítica de estas enzimas si fue reportada.

## 6.3 MÉTODOS DE BIOLOGÍA ESTRUCTURAL

### 6.3.1 CRISTALOGRAFÍA DE RAYOS X, PARA LA ELUCIDACIÓN ESTRUCTURAL DE LA PRIA DE *A. UROGENITALIS*

#### 6.3.1.1 OBTENCIÓN DE CRISTALES

Toda la metodología para obtener los cristales de la PriA\_*A. urogenitalis* fue llevada a cabo en el laboratorio de Biología estructural del Laboratorio Nacional Argonne. Para la sobreproducción de la proteína se utilizaron células BL21 Magic transformadas con el plásmido pMCSG68: PriA\_*A. urogenitalis*. Las células se crecieron a 37°C con una agitación de 200 rpm en medio LB enriquecido con fosfato 50 mM. Una vez que se alcanzó una densidad de 600 nm, el cultivo fue transferido a 4°C por una hora. Posteriormente las células fueron incubadas toda la noche a 20°C y la inducción de la proteína se llevó a cabo con 0.5 mM de IPTG. Después se centrifugaron a 5 000 rpm durante 30 min. Se resuspendieron en buffer de lisis que contenía (500 mM NaCl, glicerol al 5% (v/v), 50 mM HEPES-NaOH pH 8.0, 20 mM de imidazol y 10 mM de  $\beta$ -mercaptoetanol. El protocolo para la purificación de PriA\_*A. urogenitalis* incluye inmovilización por cromatografía de afinidad a Níquel (IMAC) utilizando el sistema AKTExpress (IMAC-I: GE Healthcare Life Sciences) seguida por el corte de la cola de Histidinas His6-tag usando la proteína recombinante His-tagged TEV proteasa y un segundo paso IMAC (IMAC-II) para remover la proteasa. Las proteínas que no se cortaron. Las proteínas purificadas fueron concentradas utilizando filtros Amicon Ultra (Millipore, Bedford, Massachusetts, USA) en 20 mM HEPES-NaOH pH 8.0, 250 mM de NaCl, 2 mM de ditioneitol. La proteína fue cristalizada utilizando la técnica de gota colgante (sitting-drop) y difusión de vapor utilizando placas de 96 pozos CrystalQuick (Greiner Bio-One, Monroe, Carolina del Norte, USA). La finalidad de la técnica de la gota colgante, es tratar de encontrar una condición en la que la proteína pueda cristalizarse. Ésta técnica consiste en un pocillo donde estará el reservorio que es la solución donde se intenta cristalizar la proteína. En un portaobjetos especialmente diseñado para ello, se pondrá una gota de la proteína y se le añadirá

una gota de este reservorio. Posteriormente el portaobjetos se coloca arriba del pocillo y se sellará evitando el contacto con el líquido contenido en el pocillo. Lo que ocurre dentro se denomina equilibrio de vapor, que consiste en que ambas soluciones deben alcanzar la misma concentración final. Como la gota donde se encuentran el reservorio y la proteína está menos concentrada, comenzará a perderse agua mediante una reacción química de evaporación que llevará esta agua hasta el pocillo, provocando que el reservorio este más diluido y la gota más concentrada (Rhodes *et al.* 2006). Las gotas para la cristalización consistieron de 0.4  $\mu\text{l}$  de la proteína en solución y 0.4  $\mu\text{l}$  de la solución de reservorio del Screen de cristalización de MCSG (Microlytic, Woburn, Massachusetts, USA). Fueron preparadas utilizando un dispensador líquido de mosquito (TTP LabTech, Cambridge, Massachusetts, USA). La concentración de la proteína fue de 37  $\text{mg/ml}^{-1}$ . La mezcla fue equilibrada contra 135  $\mu\text{l}$  de la concentración de reservorio. Los cristales aparecieron a 289 K en las siguientes condiciones; 0.2 M  $\text{Li}_2\text{SO}_4$ , 0.1 M CAPS:NaOH ph 10.5 y 1.2 M  $\text{NaH}_2\text{PO}_4/0.8$  M  $\text{K}_2\text{HPO}_4$ .

#### 6.3.1.2 COLECCIÓN DE DATOS

Los cristales de PriA\_*A. urogenitalis* fueron crioprottegidos en 2.4 M  $\text{K}_2\text{PO}_4$  y subsecuentemente congelados de forma rápida. Los datos de difracción fueron colectados a 100 K en el centro de Biología Estructural en la Fuente Avanzada de Protones en el laboratorio Nacional Argonne, Chicago, IL. USA. Los conjuntos de datos de PriA\_*A. urogenitalis* fueron colectados a 19-ID equipado con un ADSC cuántico Q315r detector CCD a una longitud de onda de 0.979 Å. La imagen fue procesado utilizando el software HKL3000 (Minor *et al.* 2006). Las estadísticas de los datos de colección se resumen en la siguiente tabla.

#### 6.3.1.3 RESOLUCIÓN DE LA ESTRUCTURA Y AFINAMIENTO

La estructura de PriA\_*A. urogenitalis* fue resuelta con reemplazamiento molecular tomando en cuenta el homólogo de PriA 1VZW (Due *et al.* 2011) depositados en la base de datos del Protein

Data Bank (PDB), éste fue usado como prueba en Phaser (McCoy et al. 2007). El modelo inicial fue mejorado por el protocolo de reconstrucción automático en Arp/wArp. La estructura fue modificada por iteraciones de reconstrucción manual en el programa gráfico COOT (Emsley y Cowtan 2004) y completado con refinamiento anisotrópico en PHENIX (Adams *et al.* 2010) con átomos de hidrógeno en ciertas posiciones. El modelo de PriA\_*A. urogenitalis* comprende el residuo Ser-2-Arg137 y Gly143-Ala247, 305 moléculas de agua, cuatro iones fosfato y un motivo CAPS. El mapa de diferencias mFo-DFc reveló dos picos fuertes (cerca del Asp51 y Leu230) que pudieron haber sido asignados ambiguamente. La calidad del modelo refinado fue verificado por el servidor Molprobit (Chen *et al.* 2010).

### 6.3.2 CONSTRUCCIÓN DE MODELOS POR HOMOLOGÍA DE PRIAS, SUBHISA2S Y SUBTRPFS

La construcción de los modelos SubHisA2\_Org36 y SubTrpF\_Org42 se hizo por homología tomando en cuenta la estructura cristalográfica de la proteína PriA\_*A. urogenitalis* (PDB:4X2R). Cada secuencia de proteína fue alineada con la secuencia de la 4X2R. De cada secuencia se construyó una librería de entre seis y nueve fragmentos utilizando el servidor Robetta. Para cada modelo se crearon 10, 000 estructuras. Una vez generadas estas estructuras se eligió la que tenía la menor energía.

## 6.4 MÉTODOS DE MICROBIOLOGÍA

### 6.4.1 ENSAYOS DE COMPLEMENTACIÓN DE LAS ACTIVIDADES HISA Y TRPF EN MUTANTES DE E.COLI HISA Y TRPF MENOS

Los ensayos de complementación de las actividades HisA y TrpF se hicieron utilizando las cepas de *E. coli hisA-* (HFrG6) y *trpF-* (JM101, FBG\_Wf) (Francisco Barona *et al* 2002 y Noda García *et al.* 2008). Las cepas fueron transformadas con los plásmidos pASK conteniendo los diferentes genes.

Como control negativo se utilizó el vector pASK vacío. Una vez que se transformaron las cepas, bajo selección a ampicilina, se seleccionó una colonia y se puso a crecer en 2 ml de medio LB y se incubó a 37°C y a 200 rpm. Una vez que las células alcanzaron una OD<sub>600nm</sub> las células fueron centrifugadas a 3000 rpm durante 5 minutos y lavadas con 1 ml de PBS 1x tres veces. Después de los lavados, las células se resuspendieron en 1 ml de PBS, volvió a medir la densidad óptica a 600 nm y se ajustaron con una OD de 1. Se llevaron a cabo tres diluciones de 10<sup>8</sup>, 10<sup>6</sup>, 10<sup>4</sup> y 10<sup>2</sup>. Las diluciones con las células fueron sembradas por goteo de 5 µl en cajas que contenían medio mínimo M9 sin histidina ni triptófano. Adicionalmente se indujo la expresión con 100 ng/ml de anhidrotetraciclina.

#### 6.4.2 CONSTRUCCIÓN DE LA MUTANTE DE *PRI A* EN *A. ORIS* MG-1

La generación de la mutante en *priA* de *A. oris* MG-1 fue llevada a cabo por el Dr. Wu Chenggang del Instituto Nacional de Salud de la universidad de Texas (NIH por sus siglas en ingles). Para crear la mutante en *priA*, mediante PCR se amplificaron 1.5 kb de la region adyacente al gen rio arriba y rio abajo. Utilizando los oligos upstream\_1Fwd y downstream\_1Rev, y upstream\_2Fwd y downstream\_2Rev (Tabla 5). Los fragmentos obtenidos rio arriba del gen fueron digeridos con EcoRI y NdeI, mientras que el fragmento rio abajo fue digerido con NdeI y XbaI. Ambos fragmentos fueron ligados utilizando un solo paso de ligación. Dicho fragmento fue clonado en pCWU3 digerido con EcoRI y XbaI. el plásmido generado fue introducido en la cepa *A. oris* MG-1 por electroporación. La delección en el marco de lectura correcto fue seleccionada por el uso del marcador mcherry. Las colonias que fueron fluorescentes y presentaron resistencia a kanamicina son las que llevan la mutación. La mutación fue confirmada por PCR Utilizando los primers MRev y MFwd y por secuenciación.



### 6.4.3 ENSAYOS DE COMPLEMENTACIÓN DE LAS ACTIVIDADES HISA Y TRPF UTILIZANDO LA MUTANTE EN *A. ORIS* MG-1 $\Delta$ PRI A

Para los ensayos de complementación de la actividad PRA y ProFAR isomerasa de la mutante en *priA* de *A. oris* se utilizó el cósmido pJRD215, que confiere resistencia a Kanamicina y Streptomicina (donado amablemente por el Dr. Ton-That del Instituto Nacional de Salud, NIH en Texas. Tanto el cósmido como los genes de complementación fueron digeridos utilizando las enzimas NdeI y HindIII (NEB). Una vez digerido el plásmido y los genes se ligaron (T4 ligasa NEB) siguiendo las recomendaciones del proveedor a 16 °C durante 24 horas aproximadamente. Posteriormente la ligación se transformó en células T10 (invitrogen) y se extrajo el cósmido utilizando el Kit de QIAGEN para extracción de plásmidos. Para verificar que la ligación se digirió el plásmido con las enzimas NdeI y HindIII, posteriormente se observó la liberación del inserto mediante un gel de agarosa. La construcción purificada fue transformada por electroporación en células mutantes en *priA* *A. oris* MG-1. Las células transformadas se sembraron en medio BHI sólido y se incubaron a 37°C durante cinco días.

### 6.4.4 CARACTERIZACIÓN FENOTÍPICA DE LA CEPA *A. ORIS* MG1

La caracterización del metabolismo de la cepa *A. oris* MG-1 se hizo de la siguiente forma. Las cepas *A. oris* MG-1 y la mutante en *priA* *A. oris* MG-1 fueron estriadas en medio BHI sólido (Bioxon) e incubadas a 37°C durante cinco días. Una vez que se obtuvieron colonias aisladas se tomaron tres y fueron puestas a crecer en 10 ml de medio BHI líquido (Bioxon) y a 37°C con una agitación de 100 rpm durante tres días. Una vez que se observó crecimiento celular, se centrifugó el cultivo a 3000 rpm durante 5 minutos. Las células fueron lavadas con 2 ml de una solución PBS 1x. Posteriormente se cuantificó por densidad óptica a OD<sub>600nm</sub> y se ajustó a una OD de 1. Se tomaron 200  $\mu$ l de las células y se pusieron a crecer en 15 ml de medio mínimo líquido desarrollado

para *A. oris* (Wu Chenggang) (Tabla 8). De igual forma los cultivos fueron incubados durante siete días a 37°C con una agitación de 100 rpm. Como control de crecimiento utilizamos medio BHI líquido.

<b>Componente</b>	<b>Porcentaje final</b>
<b>(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub></b>	1%
<b>Na<sub>2</sub>HCO<sub>3</sub>·7H<sub>2</sub>O</b>	1%
<b>KH<sub>2</sub>PO<sub>4</sub></b>	1%
<b>Na<sub>2</sub>HPO<sub>4</sub></b>	1%
<b>MgSO<sub>4</sub>·7H<sub>2</sub>O</b>	1%
<b>FeSO<sub>4</sub>·7H<sub>2</sub>O</b>	1%
<b>Dextrosa</b>	20%
<b>Niacina</b>	0.00%
<b>Tiamine.HCL</b>	0.00%
<b>Tirosine</b>	0.01%
<b>Alanina</b>	0.01%
<b>Arginina</b>	0.01%
<b>Aspartato</b>	0.01%
<b>Cisteína</b>	0.01%
<b>Glutamato</b>	0.01%
<b>Glutamina</b>	0.01%
<b>Glicina</b>	0.01%
<b>Prolina</b>	0.01%
<b>Valina</b>	0.01%
<b>Serina</b>	0.01%
<b>Asparagina</b>	0.01%
<b>Isoleusina</b>	0.01%
<b>Leucina</b>	0.01%
<b>Lisina</b>	0.01%

<b>Metionina</b>	0.01%
<b>Phenilalanina</b>	0.01%
<b>Treonina</b>	0.01%
<b>Triptófano</b>	0.01%
<b>Histidina</b>	0.01%

**Tabla 7.** Componentes medio sintético *A.oris* MG-1.

#### 6.4.5 EXTRACCIÓN DE DNA GENÓMICO DE *A. ORIS* MG1, *A. ORIS* $\Delta$ PRI $\Delta$ Y LA CEPA UTB-8

Para la extracción del DNA se utilizó el siguiente protocolo. Partiendo del glicerol que contiene a las células de *A. oris* MGI, *A. oris*  $\Delta$ priA, UTB-8 se estrió una asada en un placa de medio BHI y se puso a incubar a 37°C y durante 96 horas aproximadamente. Una vez que se obtuvieron colonias aisladas, se inoculó una colonia de cada cepa en 10 ml de medio líquido BHI y se incubó en agitación a 100 rpm y 37°C durante 72 horas. Se centrifugaron los cultivos a 13000 rpm durante 10 minutos y se decantó el sobrenadante. La extracción de ADN se llevó a cabo mediante el método de fenol/cloroformo/alcohol isoamílico (Stulnig y Amberger 1994), con las siguientes modificaciones. La incubación de la muestra con el buffer de lisis se llevó a cabo durante 4 horas y a 65°C. Para recuperar el ADN se dejó precipitando con isopropanol aproximadamente 12 horas a -20°C.

#### 6.4.6 AISLAMIENTO DE CEPAS DE ACTINOMYCES PROVENIENTES DE MUESTRAS ORALES

Se tomaron muestras de un paciente de 43 años de edad con problemas de caries severa. las muestras fueron colectadas de la zona distal (ZD), intraocular (IN) y media profunda (MP) de la pieza dental número 46, inferior derecha. Las muestras fueron puestas en medio BHI, se incubaron durante 1 hora a temperatura ambiente. Posteriormente, los medios de BHI que contenían las muestras se dividieron en dos. La primera parte se conservó en glicerol al 50% y a -80°C. La otra parte de la muestra se puso a incubar durante 12 horas a 37°C para obtener un pre-inóculo con

suficiente biomasa de donde se partió para poner a crecer las muestra en 200 ml de medio BHI al 25%, 50%, 100% y MM\_ *A. oris*. Se tomaron muestras de los cultivos cada 12 horas durante 120 horas y se estriaron en placas conteniendo BHI sólido al 25%, 50%, 100% y MM\_ *A. oris*. Se hicieron diluciones  $10^6$ ,  $10^4$  y  $10^2$ . La muestra proveniente de la zona MP se incubó en un ambiente libre de oxígeno. El cual fue creado de la siguiente forma: se puso una vela dentro de un frasco de vidrio que contenía las placas con la siembra de bacterias y se prendió la vela. Se tapó el frasco y se esperó a que el fuego se extinguiera todo esto a 37°C. Para aislar las bacterias, las colonias que crecieron en estas condiciones fueron re-sembradas en placas de medio BHI al 100%, de igual forma se incubaron durante 120 horas a 37°C. Una vez que se obtuvieron los cultivos con bacterias aisladas. Se les hizo una tinción de Gram. Las bacterias que resultaron positivas para la tinción fueron crecidas en medio BHI líquido durante 120 horas a 37°C. Posteriormente se extrajo el ADN por el método de fenol/cloroformo/alcohol isoamílico que se encuentra en la sección **7.4.4.2**. Una vez que se obtuvo el ADN se cuantificó utilizando el nanodrop a una longitud de onda de 260/280. A partir del ADN obtenido se amplificó el gen del 16s rRNA utilizando los oligos universales R1492 y F27 (Martin and Collen, 1998)(Tabla 5).

## 7. RESULTADOS Y DISCUSIÓN

### Capítulo 1

#### 7.1 RESOLUCIÓN FILOGENÓMICA DE ACTINOBACTERIAS BASALES

En esta primera sección de resultados y discusión se llevó a cabo un análisis filogenómico de Actinobacterias basales. Dichos análisis nos permitieron detectar familias de Actinobacterias con tamaño pequeño del genoma y un bajo contenido de (G+C), lo que sugiere que algunas de estas familias están bajo reducción del genoma debido a la pérdida de genes. Basados en estos resultados y en previos análisis, seleccionamos a la familia *Actinomycetaceae* como modelo de estudio para analizar el efecto de la pérdida de genes sobre la evolución de la especificidad por sustrato de una enzima bi-funcional involucrada en la biosíntesis de L-histidina y L-triptófano.

Basado en lo anterior, uno de nuestros primeros objetivos fue determinar si las Actinobacterias que se separan temprano del ancestro común poseen genomas pequeños debido a la pérdida de genes. Para analizar esto, primero decidimos hacer un análisis filogenético de aquellas Actinobacterias que se encuentran en la parte basal, según filogenias reportadas en la base de datos del 16S Ribosomal SILVA y por Barka *et al.* (2016). Escogimos genomas disponibles en bases de datos públicas como el NCBI, que fueran representativos de cada familia y cuyos genomas estuvieran en menos de 100 secuencias contiguas. Con estos criterios se seleccionaron 135 genomas (Tabla 3, Materiales y Métodos). Cada uno de los genomas fue re-annotado utilizando la tecnología de los subsistemas, RAST (Aziz *et al.* 2008), lo que nos permitió homogeneizar los datos para posteriormente llevar a cabo un BLASTp de todas las proteínas contenidas en un genoma contra todas las proteínas contenidas en los 135 genomas seleccionados, utilizamos sólo los mejores hits bidireccionales (BBH, Best Bidirectional Hits), para obtener el genoma núcleo de las Actinobacterias basales,

obtuvimos un total de 35 proteínas que se comparten entre los 135 organismos. La mayoría de estas proteínas ortólogas están relacionados con procesos celulares como los ribosomales y referentes a la maquinaria de traducción (Tabla 8). Interesantemente, encontramos que dentro de este genoma núcleo se encuentra una proteína de choque térmico de la familia GroES, llamada “Heat shock protein 60 family co-chaperone”. Dicha proteína se encuentra altamente conservada en organismos que presentan deterioro genómico ya que su función es asistir en el plegamiento y la agregación, particularmente importante cuando se acumulan mutaciones con efectos negativos (Bukau y Horwich.1998, Moran *et al.* 2006). Esto sugiere, que probablemente algunas Actinobacterias basales también estén experimentando una aceleración en la tasa de sustitución que puede afectar el plegamiento de sus proteínas y que por eso esta proteína está bajo selección purificadora.

**Función basada en RASTtk**

**ATP synthase alpha chain**

**ATP synthase beta chain**

**DNA-directed RNA polymerase beta subunit**

**Heat shock protein 60 family co-chaperone GroES**

**LSU ribosomal protein L11p**

**LSU ribosomal protein L13p**

**LSU ribosomal protein L16p**

**LSU ribosomal protein L18p**

**LSU ribosomal protein L1p**

**LSU ribosomal protein L23p**

**LSU ribosomal protein L27p**

**LSU ribosomal protein L29p**

**LSU ribosomal protein L2p**

**LSU ribosomal protein L6p**

**LSU ribosomal protein L7/L12**

**LSU ribosomal protein L9p**

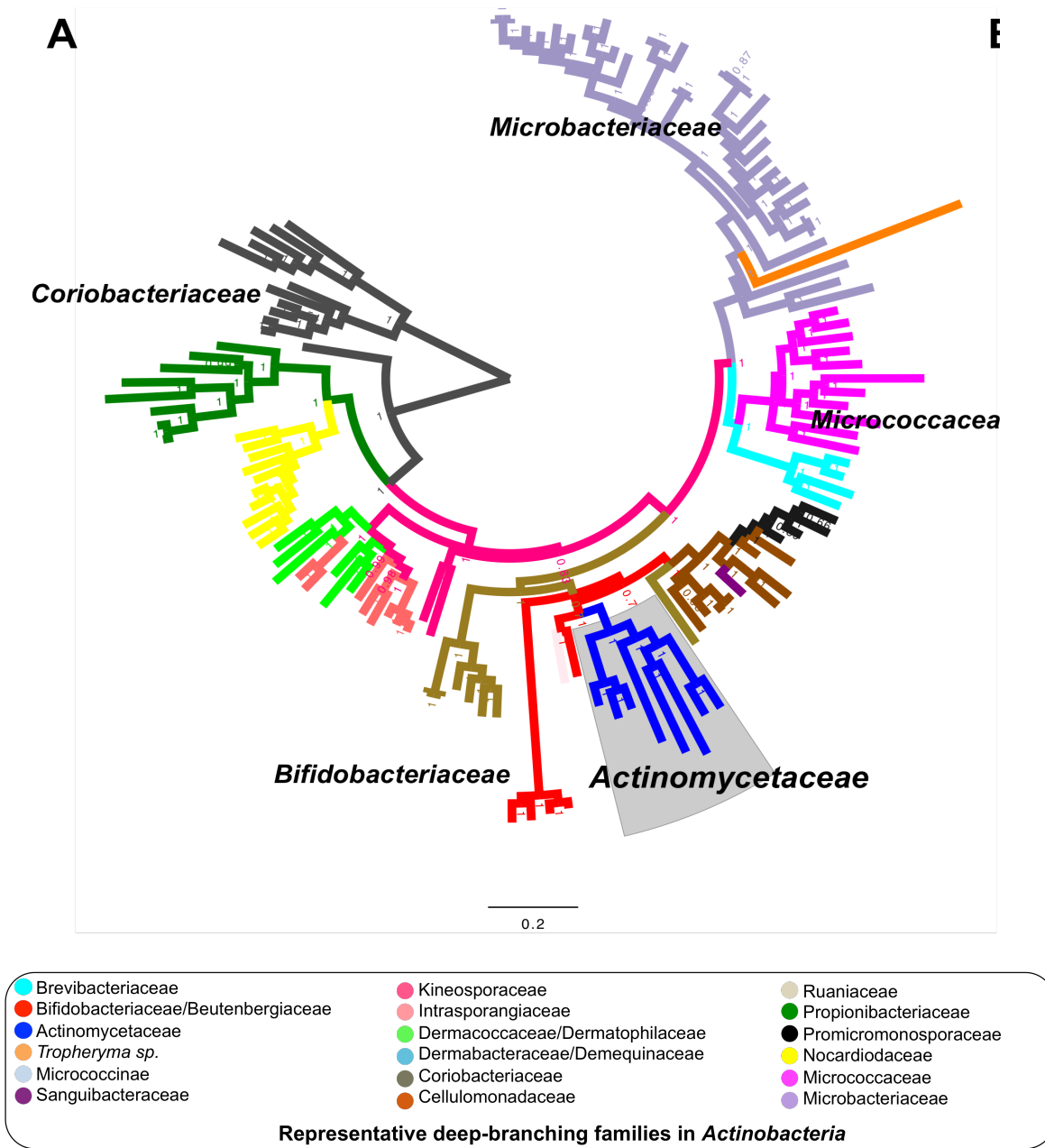
**Nucleoside 5-triphosphatase RdgB**

**Peptidyl-tRNA hydrolase**

**Preprotein translocase secY subunit**  
**Ribonuclease III**  
**Ribonucleotide reductase transcriptional regulator NrdR**  
**rRNA small subunit methyltransferase H**  
**SSU ribosomal protein S11p**  
**SSU ribosomal protein S13p**  
**SSU ribosomal protein S15p**  
**SSU ribosomal protein S18p @ SSU ribosomal protein S18p, zinc-independent**  
**SSU ribosomal protein S19p**  
**SSU ribosomal protein S7p**  
**SSU ribosomal protein S8p**  
**SSU ribosomal protein S9p**  
**Transcription termination protein NusA**  
**Translation elongation factor P**  
**Translation elongation factor Ts**  
**Translation elongation factor Tu**  
**Translation initiation factor 1**

**Tabla 8** Grupo de proteínas ortólogas altamente conservadas en el grupo de las actinobacterias basales.

Para resolver las relaciones taxonómicas de estos organismos y detectar familias bajo reducción del genoma, concatenamos el grupo de proteínas ortólogas y construimos una matriz de 135 organismos por 35 secuencias de proteínas. Con dicha matriz llevamos a cabo un análisis filogenómico utilizando un algoritmo Bayesiano. El árbol resultante se muestra en la Figura 15.

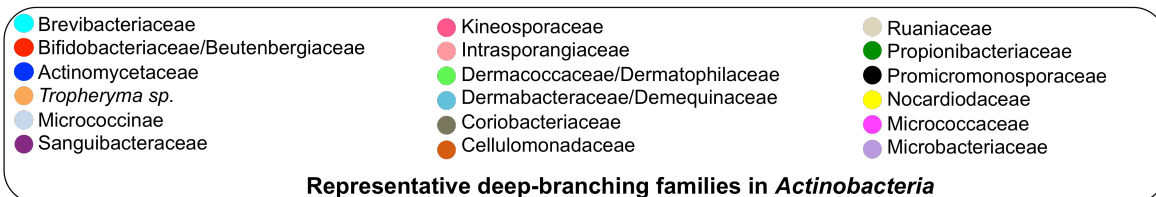
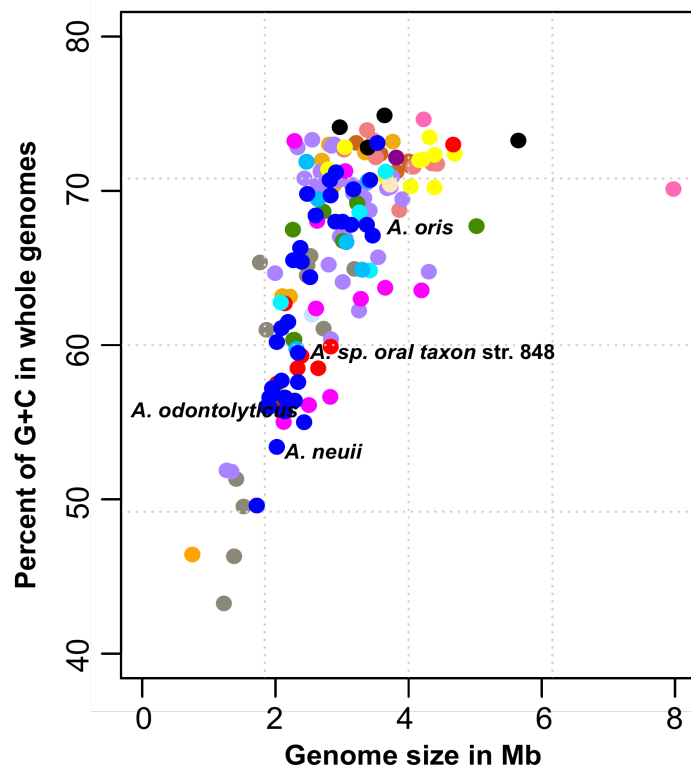


**Figura 15.** Árbol Filogenómico de familias representativas que conforman el grupo de las Actinobacterias basales. Cada familia esta marcada con un color diferente. En la caja se mesuestran las familias de organismos con su respectivo color de referencia. En los nodos, se muestran los valores de probabilidad posterior.

En dicha figura se muestra la distribución de los nodos internos que dan origen a las ramas, las cuales están altamente soportadas por valores de probabilidad posterior cercanos a 1. La filogenia



resultante coincide con filogenias previamente reportadas (Barka *et al.* 2016; Zhao *et al.* 2014), en donde las familias *Coriobacteriales*, *Propionibacteriaceae* y *Nocardiodaceae* se encuentran en la base de la filogenia (clados gris, verde y amarillo). Dos grupos, cuyas secuencias divergen rápidamente son *Troperyma* y *Bifidobacterium spp.*, los cuales se posicionan como taxones hermanos de *Micrococcales* y *Actinomycetales*, respectivamente. Cabe resaltar que, dentro del filo Actinobacteria las especies de *Troperyma* y *Bifidobacterium spp.* poseen los genomas con tamaños más pequeños y con más bajo contenido de G+C, lo que podría estar relacionado con una reducción del genoma y una tasa acelerada de mutaciones que se ve reflejada en la divergencia observada (ramas naranja y roja con longitudes grandes, Figura 15, Tabla 4). (Gupta *et al.* 2012; Barka *et al.* 2016). Además, los resultados también muestran que el género *Bifidobacterium* se encuentra como grupo externo a la familia *Actinomycetaceae*, la cual se sabe esta bajo un proceso de pérdida y ganancia de genes (Zhao *et al.* 2014). En general, la mayoría de los organismos seleccionados para éste análisis además de compartir ancestros comunes, constituyen un grupo con genomas pequeños (promedio 3.01 Mpb) si se compara con otros organismos que integran a las Actinobacterias como las del género *Streptomyces* y *Salinispora* que poseen genomas con un tamaño de más de 8 Mpb (Barka *et al.* 2016). Así mismo, para determinar si existía una relación entre el tamaño del genoma de los organismos seleccionados, contra el contenido-(G+C), se graficaron estas dos variables. Los resultados muestran que estos organismos poseen un contenido-(G+C) bajo (promedio 66.7 %) (Figura 16), si se compara con otras especies de Actinobacteria como *Streptomyces* (promedio >75%).



**Figura 16.** Contenido de (G+C) y tamaño del genoma en Actinobacterias basales. Cada punto representa un organismo perteneciente a una familia. Los colores representan a las familias, las cuales siguen el código de color propuesto en la Figura 15.

En la Figura 16 se observa que miembros de la familia *Actinomycetaceae* específicamente las clasificadas como parte del género *Actinomyces* poseen genomas de tamaño y contenido-(G+C) variables (puntos azul marino). Además, las ramas que dan origen a éstos taxa parecen evolucionar aceleradamente. Éste resultado se relaciona con observaciones previas sobre la dinámica genómica, relacionada con una pérdida y ganancia de genes dentro de ésta familia (Zhao et al. 2014).

Basados en éstos resultados, es decir, la ubicación taxonómica de la familia *Actinomycetaceae* como una familia de las más basales del filo Actinobacteria y la peculiaridad de éstas de poseer una enzima bi-funcional llamada PriA, la cual sólo se encuentra en Actinobacterias y en estudios previos se ha demostrado que en esta familia PriA parece estar divergiendo hacia la formación de nuevas subfamilias (Noda-García et al. 2015). Decidimos seleccionar a la familia *Actinomycetaceae* como modelo de estudio para tratar de determinar el efecto de la dinámica genómica, específicamente la pérdida de genes sobre la evolución de la función enzimática de PriA. Sin embargo, muchos miembros de esta familia son organismos oportunistas de mamíferos y en las bases de datos públicas se encuentran pocas secuencias genómicas de dichos organismos, por éstas razones decidimos tratar de aislar algunas cepas mediante un muestreo bucal (Thompson y Lovstedt, 1951; Belda Ferre *et al.* 2012), ya que se sabe que algunos de éstos organismos colonizan la cavidad oral humana y una vez obtenidas las cepas proseguir con la secuenciación genómica. Por lo que en las secciones subsiguientes describiremos los resultados obtenidos de dicho muestreo bucal.

## 7.2 MUESTREO BUCAL Y OBTENCIÓN DE NUEVAS ACTINOBACTERIAS

Con el objetivo de obtener un mayor número de secuencias genómicas de la familia *Actinomycetaceae* para llevar a cabo nuestro análisis evolutivo y funcional de PriA, se decidió llevar a cabo un muestreo de la cavidad oral de una persona adulta que presentaba problemas de caries. Las muestras fueron tomadas de tres zonas diferentes de un molar (Materiales y Métodos sección 6.4.4.3). Se logró aislar 21 morfotipos diferentes de bacterias (**Tabla 9**), 12 de ellas dieron positivo para la tinción de Gram (posibles Actinobacterias) y 9 negativo para la tinción de Gram. Todos los morfotipos, presentaron una respiración anaerobia facultativa, sólo el organismo etiquetado como 18 presentó una respiración microaerofílica.

<b>Identificador</b>	<b>Lugar de aislamiento</b>	<b>Tipo de respiración</b>	<b>Morfología según la tinción de Gram</b>
<b>FBG_1</b>	zona media profunda	Anaerobia facultativa	Bacilos Gram positivos
<b>FBG_2</b>	zona media profunda	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_3</b>	zona media profunda	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_4</b>	zona media profunda	Anaerobia facultativa	Cocos Gram positivos
<b>FBG_5</b>	zona media profunda	Anaerobia facultativa	Cocos Gram positivos
<b>FBG_6</b>	zona media profunda	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_7</b>	zona media profunda	Anaerobia facultativa	Bacilos Gram positivos
<b>FBG_8</b>	zona distal	Anaerobia facultativa	Bacilos Gram positivos
<b>FBG_9</b>	zona distal	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_10</b>	zona distal	Anaerobia facultativa	Cocos Gram positivos
<b>FBG_11</b>	zona distal	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_12a</b>	zona distal	Anaerobia facultativa	Bacilos Gram positivos
<b>FBG_12b</b>	zona distal	Anaerobia facultativa	Bacilos Gram positivos
<b>FBG_13</b>	zona distal	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_14</b>	zona intraocular	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_15</b>	zona intraocular	Anaerobia facultativa	Bacilos Gram negativos
<b>FBG_16</b>	zona intraocular	Anaerobia facultativa	Cocos Gram negativos
<b>FBG_17</b>	zona intraocular	Anaerobia facultativa	Bacilos Gram positivos
<b>FBG_18</b>	zona intraocular	Microaerofílico	Bacilos Gram positivos
<b>FBG_19</b>	zona intraocular	Anaerobia facultativa	Cocos Gram positivos
<b>FBG_20</b>	zona intraocular	Anaerobia facultativa	Bacilococos Gram positivos

**Tabla 9.** Cepas obtenidas del muestreo bucal, se muestra la etiqueta que se usó, el lugar de aislamiento, el tipo de respiración y el resultado de la tinción de Gram.

Con la finalidad de obtener la secuencia de la subunidad 16S del rRNA de éstos organismos para llevar a cabo análisis filogenéticos, se extrajo el ADN (Materiales y métodos sección 6.4.4.1) y se amplificó el 16s. Logramos amplificar el 16s ribosomal de 14 cepas de las 21 obtenidas (aprox. 66%). Cuatro provenientes de la Zona distal, cuatro de la Zona media profunda y seis de la Zona Intraocular. Para ubicar a nivel taxonómico los aislados, una vez que se obtuvieron las secuencias del 16s, se hizo una búsqueda por homología utilizando la base de datos del Microbioma Oral Humano HOMD, por sus siglas en inglés de Human Oral Microbiome Database. Esta base de datos contiene todos los 16s con un taxa asignado de todos los organismos que hasta Agosto 2016 han sido secuenciados y que fueron aislados de la cavidad bucal humana. En esta base de datos están contenidas especies pertenecientes al *filo* Actinobacteria de los géneros *Lactobacillus*, *Neisseria*, *Rothia* y *Actinomyces*, entre otros.

Como se observa en la tabla 10, todas las cepas aisladas pertenecen a los *filo* Firmicutes, Proteobacteria y Actinobacteria, encontramos que los organismos del *filo* Firmicutes del género *Lactobacillus* son los más frecuentes, puesto que se distribuyen en las tres zonas muestreadas, seguidos por especies del género *Neisseria* que sólo se encuentran en la zona intraocular y distal. Por último encontramos que el género *Rothia* del *filo* Actinobacteria, solo se encuentra presente en la Zona Media profunda (**Tablas 9 y 10**).

Secuencia semilla	Tamaño	Mejor Hit con la cepa de OHMD	Identities (%)
<b>FBG_1</b>	1491	Lactobacillus casei   HOT_568   Strain_JCM 1136   GB_D16552   Named   G   0	99.1
<b>FBG_12a</b>	1444	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99.9
<b>FBG_12b</b>	1374	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99
<b>FBG_15</b>	1417	Neisseria flava   HOT_609   Strain_U40   GB_AJ239301   Named   X   5	99.8
<b>FBG_16</b>	751	Neisseria flava   HOT_609   Strain_U40   GB_AJ239301   Named   X   5	100

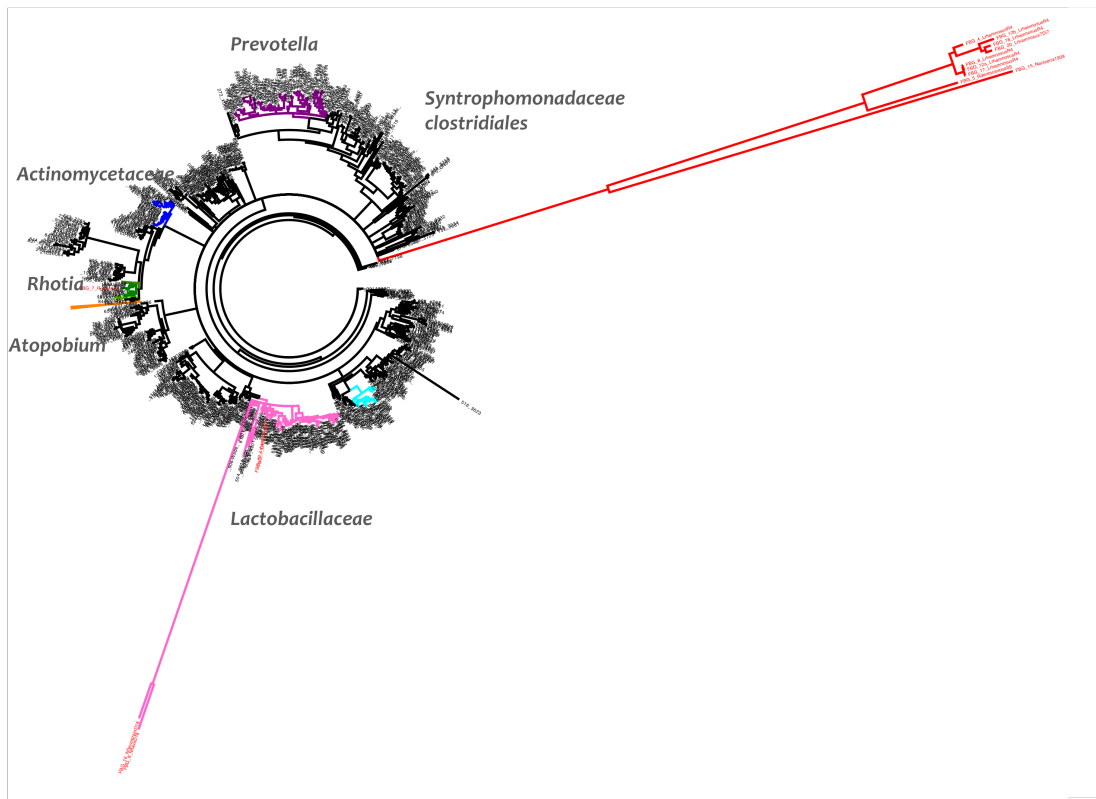
<b>FBG_17</b>	1430	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99.9
<b>FBG_18</b>	1394	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99.7
<b>FBG_19</b>	559	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99.5
<b>FBG_20</b>	1392	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	96.2
<b>FBG_4</b>	1475	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99.3
<b>FBG_5</b>	1438	Rothia dentocariosa   HOT_587   Strain_ATCC 17931   GB_M59055   Named   G   6	99.3
<b>FBG_7</b>	1448	Rothia aeria   HOT_188   Clone_ncd2429e02c1   GB_JF210891   Named   G   5	99.9
<b>FBG_8</b>	1490	Lactobacillus rhamnosus   HOT_749   Strain_F11   GB_AF243146   Named   G   2	99.4
<b>FBG_9</b>	777	Neisseria flava   HOT_609   Strain_U40   GB_AJ23930	99.5

**Tabla 10.** Identificación de organismos obtenidos del muestreo bucal.

Para determinar la ubicación filogenética de las especies obtenidas en el muestreo, llevamos a cabo un análisis filogenético, el cual incluye todas las secuencias del 16s rRNA obtenidas y 1 000 secuencias de 16s rRNA provenientes de la base de datos del HOMD. Nuestros resultados muestran que cepas pertenecientes al *filo* Actinobacteria como la cepa del género *Rothia* (ramas verdes, etiqueta roja, Figura 17 y FBG\_7 Tabla 12) se agrupan bien con otras especies de *Rothia* reportadas, estas especies tienen como taxones hermanos a la familia *Actinomycetaceae* y *Atopobium*. Las cepas marcadas FBG\_1, FBG\_16 y FBG\_19 asignadas como *Lactobacillus*, se agruparon junto con otras especies de la familia *Lactobacillaceae* (ramas rosas, Figura 17). De modo interesante, las secuencias marcadas como FBG\_4, FBG\_12a, FBG\_12b, FBG\_15, FBG\_17, FBG\_18 y FBG\_20 identificadas por homología como: *Lactobacillus*, *Rothia* y *Neisseria* (Tabla 10) se agruparon en un clado separado junto con especies de clostridiales (rama roja, Figura 17). Este resultado sugiere que a pesar del alto porcentaje de identidad que presentan éstas con secuencias reportadas (>99%)

parece que están evolucionando a tasas aceleradas. Además, el resultado podría estar enmascarado porque la mayoría de las secuencias en la base de datos HOMD poseen un tamaño de 600 pb. Tal vez, se necesitaría llevar a cabo un estudio más detallado tomando en cuenta hasta 1000 pb, lo que podría dar mayor soporte y resolución a nuestro análisis.

En resumen éste tipo de aproximaciones experimentales no nos permitieron obtener cepas de la familia Actinomycetaceae. Juntos nuestros resultados demuestran la necesidad de tratar de obtener un método para aislar cepas de dicha familia, el cual podría basarse en el uso de reconstrucciones metabólicas *in silico* para determinar el medio ideal para dichos aislados. Además, se necesita una base de datos con genomas de cavidades bucales provenientes de individuos mexicanos, dado que su historia evolutiva podría estar marcada por nuestros estilos de vida que son diferentes a los de la cultura Occidental de la cual provienen la mayoría de las secuencias genómicas reportadas y por ende algunas de las cepas obtenidas divergen mucho de las reportadas. Sería interesante secuenciar el genoma de la *Rothia* que encontramos para hacer estudios de genómica comparativa y determinar su adaptación a los estilos alimenticios como el alto consumo de chile (capsaicina) y bebidas conteniendo azúcar proveniente de caña.



**Figura 17.** Árbol filogenético inferido con Mr. Bayes, conteniendo las secuencias del 16s provenientes del muestreo bucal (FBG) y secuencias del 16s tomadas de Human Oral Microbiome Database (HOMD). En rojo se resaltan las secuencias del 16s provenientes de FBG.

Otra variable que pudo haber influido en la no obtención de cepas de *Actinomyces* durante nuestro muestreo, fue que la mayoría de las especies de *Actinomyces* crecen en condiciones de anaerobiosis. Aunque se trató de mantener condiciones de anaerobiosis, es probable que no se logró obtener la condición ideal para que crecieran dichos organismos. Dado que con éste muestreo no se pudieron aislar cepas de *Actinomyces* que es nuestro grupo de estudio en cuestión, decidimos explorar otras opciones de obtención de cepas. Una de ellas fue la búsqueda de investigadores que trabajaran con *Actinomyces*, contactamos al Dr. Hung That, del Instituto Nacional de Salud en Texas, NIH (por sus siglas en inglés de National Institute of Health) quién amablemente nos donó dos cepas de ésta familia, la primera fue la cepa *A. oris* MG-1 y la segunda una cepa marcada como UTB-18.



### 7.2.1 SECUENCIACIÓN GENÓMICA Y ANÁLISIS DE LAS CEPAS *A. ORIS* MG-1 Y UTB18 PROVENIENTES DEL GÉNERO *ACTINOMYCES*

La cepa de *A. oris* MG-1, es una cepa que se ha utilizada por décadas para el estudio de la formación de fimbrias en bacterias (Arunima *et al.* 2010). Fue aislada de la boca de un individuo con periodontitis, y hasta el 2008 no se contaba con una secuencia genómica de la misma. Por lo que, decidimos secuenciar el genoma al igual que el de la cepa desconocida, pero aparente *Actinomyces*, UTB18, la cual fue amablemente donada por Genne Tribble, también del NIH. Esta última cepa, fue aislada de una infección bucal en la que había 0.8 mm de profundidad. Los resultados de secuenciación fueron los siguientes: para la cepa *A. oris* MG-1, se obtuvo un total de 1 billón de pares de bases, todos los reads juntos poseen una cobertura de 222.25 del genoma, estos reads o lecturas del genoma fueron ensamblados utilizando velvet, utilizando un K-mer de 40, logramos obtener el genoma en 368 contigs. El tamaño calculado del genoma del fue de 2.87 Mpb, mientras que, el contenido de (G+C) fue estimado en un 68%. Para predecir los genes y encontrar los marcos de lecturas abiertos utilizamos Glimmer, contenido en la plataforma de RAST. Encontramos 53 ARNs y 2829 marcos de lectura.

Para la cepa UTB18 se obtuvieron 900 millones de pares de bases, todas las lecturas juntas nos dieron una cobertura del genoma de 181.28. También se usó Velvet para ensamblar el genoma, también con un k-mer de 40; obtuvimos un total de 484 contigs. El tamaño del genoma claculado para esta cepa fue de 2.43 Mpb. El contenido de (G+C) fue estimado para ser del 68%. De igual forma utilizamos RAST para predecir los marcos de lectura abiertos y encontramos 52 ARNs y 2995 marcos de lectura. Estos resultados de la secuenciación, no fueron tan buenos como se esperaba, debido a que aunque se intentó obtener el genoma en el menor número de contigs, los genomas siguen estando en pedazos. Una solución a éste problema sería secuenciar nuevamente ambas cepas o tratar de cerrarlos utilizando oligos específicos para amplificar los huecos, para ambos casos se necesitarían más de 250 pares de oligos a la par durante el desarrollo de éstos

experimentos se buscaron nuevos genomas en la base de datos del NCBI y hasta el 2013 se encontraron más de 25 genomas disponibles, dichos genomas fueron utilizados para los estudios siguientes.

### 7.3 RESOLUCIÓN FILOGENÓMICA DE LA FAMILIA *ACTINOMYCETACEAE*

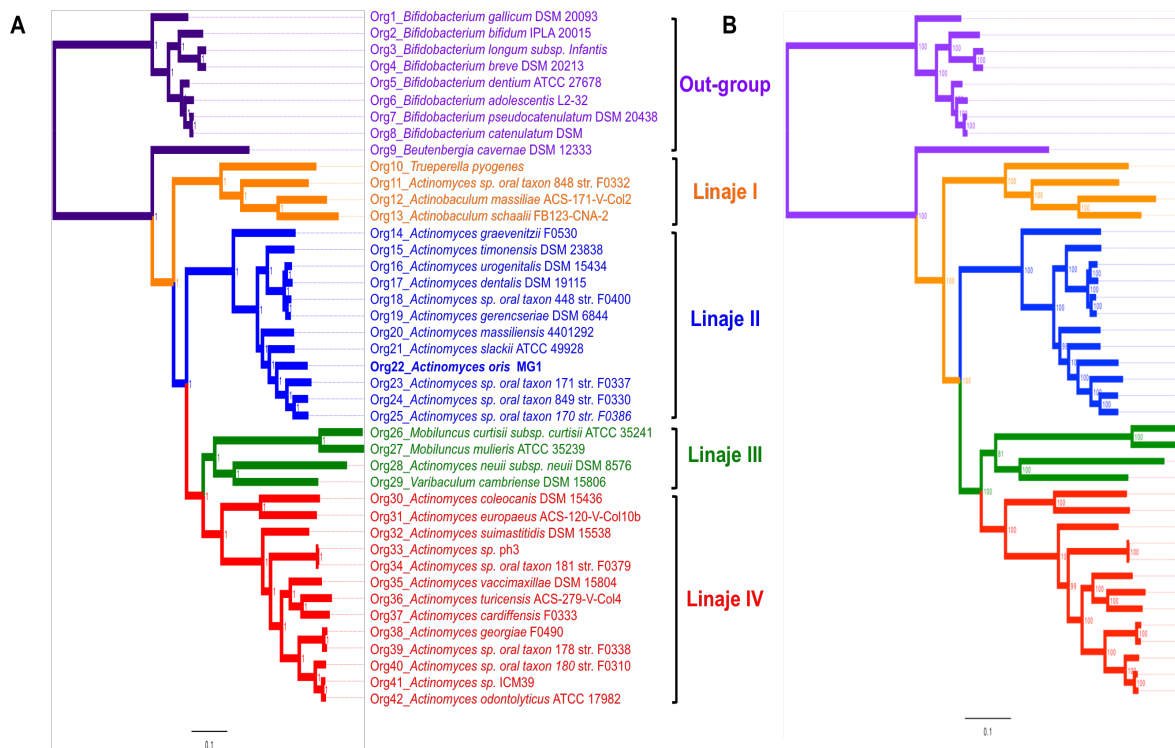
Como se mencionó en las secciones anteriores, decidimos utilizar a la familia *Actinomycetaceae* como modelo de estudio para elucidar el efecto de la pérdida de genes sobre la evolución de la función enzimática por lo que uno de los objetivos es tratar de resolver las relaciones taxonómicas dentro de la familia, hicimos una búsqueda de genomas completos en la base de datos del NCBI, basados en nuestros estudios filogenéticos de la sección 7.1, seleccionamos ocho genomas provenientes del género *Bifidobacterium spp* y a *Beutenbergia cavernae DSM 12333*; como grupo externo, ya que se comparten un ancestro común con la familia *Actinomycetaceae*. También seleccionamos 33 genomas pertenecientes a la familia *Actinomycetaceae*: 24 que han sido clasificadas como especies de *Actinomyces spp.*; 2 de *Mobilluncus spp.*; 1 *Trueperella*; 1 de *Varibaculum*, 2 de *Actinobaculum spp.*, y utilizamos el genoma de *A. oris* MG-1, que previamente secuenciamos (Tabla 4). En éste primer acercamiento para determinar las relaciones taxonómicas de la familia, decidimos dejar fuera el genoma de la cepa UTB-18, por que es un genoma nuevo que necesita ser analizado a más detalle, pero que retomaremos más adelante.

Para llevar a cabo un análisis filogenómico, de secuencias conservadas, decidimos como primer paso buscar los grupos de proteínas ortólogas que se conservan en todos los genomas seleccionados esto con la finalidad de poder llevar a cabo los análisis, reanotamos todos los genomas utilizando la tecnología de los subsistemas RASTtk (Aziz *et al.* 2008; 2015), mediante el uso de los mejores Hits bidireccionales, BBH, obtuvimos los grupos de proteínas ortólogas, que se encuentran presentes tanto en la familia *Actinomyceataceae* como en el grupo externo. A éste grupo de proteínas obtenidas lo denominamos como el genoma core o núcleo y esta compuesto por 205 grupos de

proteínas ortólogas. En la figura 18 se muestra la representación de las diferencias en el número de ortólogos presentes sólo en la familia *Actinomycetaceae* (435 grupos de ortólogos) y cuando integramos al outgroup vemos como hay una disminución en el número de grupos de ortólogos, obteniendo un total de 205 genes. Si tomamos en cuenta que el promedio de proteínas en éstos organismos es de 2,500, este resultado resalta que aproximadamente el 17% de las funciones que están codificadas en éstos organismos están conservadas, lo que sugiere que a pesar de provenir del mismo género existe una enorme diferencia, lo cual podría ser debida a una especialización reciente de algunos de éstos organismos que podrían estar formando nuevos linajes bacterianos.

Entre los 205 grupos de ortólogos que forman parte del genoma núcleo, se encuentran proteínas del metabolismo celular como algunas polimerasas, girasas entre otras, que participan en el proceso de replicación. También encontramos proteínas relacionadas con el metabolismo primario, algunas chaperonas como la proteína de choque térmico GroEL/GroES. Interesantemente, además de compartir un ancestro común, estos organismos comparten un sistema de secreción tipo II/IV ATP hidrolasa TadA/VirB11/CpaF, TadA, éste sistema de exportación de proteínas parece estar altamente conservado entre diferentes tipos de bacterias incluyendo las Gram positivas y parece jugar un papel importante durante la adherencia a algunos tejidos celulares (Planet *et al.* 2001).

Para llevar a cabo el análisis filogenómico, concatenamos éstos 205 grupos de ortólogos para obtener una matriz y hacer el análisis utilizando dos algoritmos; el de máxima verosimilitud y el Bayesiano. Los resultados de estos análisis muestran una consistencia entre ambas filogenias, además las ramas que dan origen a los clados detectados están soportadas con valores de probabilidad posterior y bootstrap cercanos a 1 y a 100 respectivamente, Figura 18A y 18B.



**Figura 18.** Árboles filogenéticos de la familia *Actinomycetaceae* construidos con la concatenación de 205 grupos de proteínas ortólogas. A) Filogenia construida con el algoritmo bayesiano, en los nodos se muestran los valores de probabilidad posterior. B) Filogenia construida con el algoritmo de máxima verosimilitud, RAxML, en los nodos se muestran los números de bootstrap.

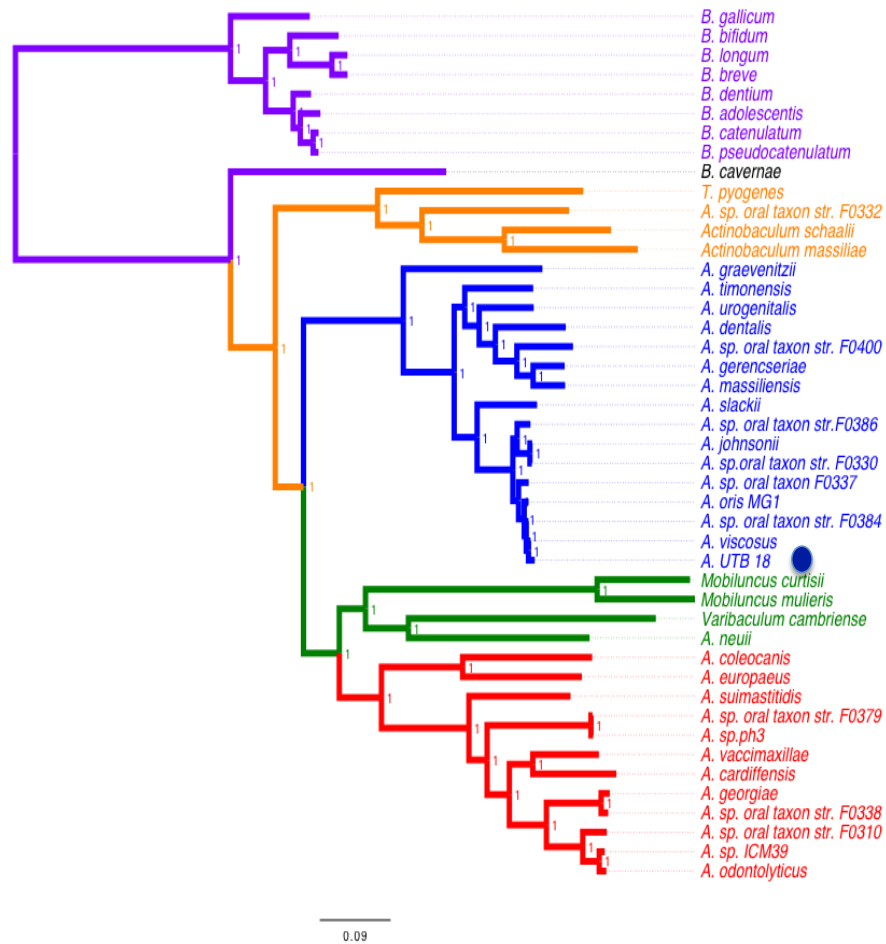
Ambas filogenias muestran que la familia *Actinomycetaceae* puede dividirse en tres grandes clados, los cuales hemos denominado como linaje I, II y III. Interesantemente, basados en clasificaciones previas (Gupta *et al.* 2012) proponemos subdividir el linaje III en uno más dando como resultado la existencia de cuatro linajes evolutivos: I, II, III y IV. El linaje I tiene un origen monofilético (clado naranja, Figura 19A-B), es el grupo más basal de la familia y en él se encuentran especies de los géneros *Trueperella*, *Varibaculum* y *Actinomyces* Org<sub>11</sub>. Lo que sugiere que este último organismo está más cercano a los géneros *Trueperella*, *Varibaculum* que al género *Actinomyces*. Este grupo parece haberse adaptado a diferentes nichos ecológicos, puesto que se ha descrito que estas especies colonizan diferentes cavidades de mamíferos. El linaje II (clado azul, Figura 18A-B),

esta compuesto sólo por especies del género *Actinomyces* sugiriendo un origen parafilético de éste género, ya que este clado no contiene a todas las especies que han sido clasificadas como *Actinomyces*. Éstas especies se han especializado en colonizar solo cavidades humanas (Könönen and Wade 2015). El linaje III es un grupo monofilético, (clado verde, Figura 19) conformado por organismos de los género *Mobiluncus*, *Varibaculum* y *Actinomyces* Org\_28, estas especies se han especializado en colonizar el tracto urogenital femenino. La inclusión del Org\_28 como parte de este linaje, sugiere que al igual que el Org\_11, éstos organismos han sido mal clasificados, y que si se toma en cuenta el uso de estos 205 grupos de ortólogos propuestos aquí es posible resolver las relaciones taxonómicas. El último linaje, IV (clado rojo, Figura 18), tiene un origen parafilético como ya se comentó puesto que comprende sólo especies del género *Actinomyces*.

Estos análisis sugieren que los organismos que han sido clasificados como especies del género *Actinomyces* están adaptándose para formar dos grandes linajes que aquí hemos denominado II y IV. Cómo ya se mencionó el II es el que se separa antes del ancestro común, además la rama que da origen a este linaje, parece que acumuló más mutaciones en comparación con la que da origen al linaje IV por lo que es posible que existan muchos cambios a nivel de nucleotídico en el linaje II, posteriormente parece que hubo un cese en la acumulación de mutaciones. Estudios de genómica de poblaciones del linaje II, utilizando modelos de coalescencia podrían explicar la acumulación de estas mutaciones, y podrían establecer si fue un evento neutral debido al tamaño de la población o hubo una selección hacia esta acumulación de mutaciones. Pero en primera instancia parece ser el linaje más diferente dentro de ésta familia, por lo que nos preguntamos si esta división taxonómica podría relacionarse con adaptaciones metabólicas y por supuesto de las enzimas.

Adicionalmente, las proteínas ortólogas fueron buscados en la cepa UTB-18, y posteriormente las utilizamos para resolver las relaciones taxonómicas de la cepa UTB-18, ésta cepa se agrupó con los organismos del linaje II, muy cercana a la cepa *A. oris* MG-1 (Figura 19). Este resultado sugiere que probablemente el subclado conformado por las especies de *A. oris* y la cepa UTB-18 sea el

clado que agrupa a los organismos que tienen la capacidad de establecer interacciones con otros organismos para formar la caries dental. Para poder afirmar esto se necesitan más estudios de genómica comparada de estos genomas, los cuales no son el objetivo principal de ésta tesis.

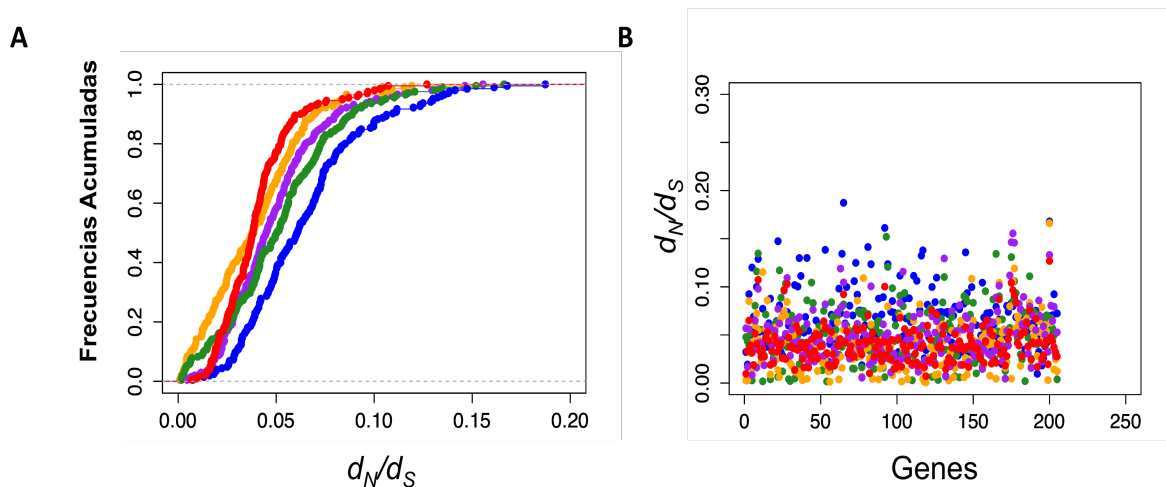


**Figura 19.** Ubicación taxonómica de la cepa UTB-18 (círculo azul), la cepa se ubica con los organismos del Linaje II (clado azul). La filogenia fue construida con el algoritmo bayesiano, en los nodos se muestran los valores de probabilidad posterior.

## 7.4 CÁLCULO DE LAS TASAS DE SUSTITUCIÓN NUCLEOTÍDICA ENTRE LOS LINAJES

La diversidad que observamos dentro de la familia *Actinomycetaceae* ha sido el resultado de las mutaciones, la selección natural y la deriva génica, cómo se observa en nuestros análisis filogenéticos, parece que a pesar del alto grado de conservación en los grupos de proteínas ortólogas, existe una diferencia en la acumulación de mutaciones que ocurrieron durante la formación de los nuevos linajes. Por lo que, esperamos que un análisis de las tasas de sustitución nucleotídica ( $d_N/d_S$ ), nos permitirá determinar el tipo de selección que actúa sobre las secuencias del grupo de ortólogos. Para dicho análisis incluimos al género *Bifidobacterium* y la determinación de las tasas de sustitución para cada rama seleccionada se muestra en materiales y métodos. Dado que nos interesa conocer si existen diferencias significativas en las tasas de sustitución dentro de cada linaje, después de aplicar el análisis de máxima verosimilitud (LRT, por sus siglas en inglés de Likelihood-ratio test) sólo tomamos los genes en cuyo caso particular la hipótesis alternativa fue aceptada *i.e* son genes que presentan diferentes tasas de sustitución entre los linajes propuestos y dados los grados de libertad ocupados ( $df=5$ ) con un valor de probabilidad del  $p=0.05\%$ , para el análisis estadístico de Chi2 se rechaza la hipótesis alterna cuando es menor a ( $p<11.07$ ). De estos 205 (100%), en sólo 174 (84 %) se aceptó la hipótesis alterna, es decir los grupos de genes que provienen de los diferentes linajes presentan diferentes tasas de sustitución (Ver materiales y métodos). Tal cual era de esperarse, el grupo de genes en dónde se rechaza la hipótesis alterna son algunos reguladores y proteínas ribosomales de subunidades pequeñas y grandes, estas proteínas están pasando por una presión de selección elevada, valores  $d_N/d_S < 1$  y se relaciona con su importancia celular (Información suplementaria, recuadros naranjas). De estos 174 genes, calculamos las diferencias que existían entre las tasas de sustitución en cada linaje. Utilizando frecuencias acumuladas obtuvimos los promedios para cada linaje. En la gráfica de frecuencias acumuladas Figura 20A, se observa que los organismos que provienen del Linaje II

presentan una mayor tasa de sustitución, en comparación con los otros linajes, es decir aunque los genes que se encuentran altamente conservados están sujetos a una presión de selección purificante, ésta presión es ligeramente más relajada ( $d_N/d_S > 0.05$ ) que la de los otros Linajes ( $d_N/d_S < 0.05$ ). A pesar de las diferencias en las tasas de sustitución entre los genes, ninguno de éstos genes está bajo selección positiva ( $d_N/d_S > 1$ ) (Figura 20B). Todo esto sugiere fuertemente que en etapas tempranas durante la separación del ancestro, los organismos que se agrupan en el Linaje II acumularon más mutaciones que los otros Linajes I, III, IV, lo que les permitió evolucionar por separado y conservar tamaños de genomas más grandes.



**Figura 20.** Tasas de sustitución calculadas para cada uno de los 205 genes provenientes de los diferentes organismos que conforman los linajes evolutivos en la familia *Actinomycetaceae*. A) Tabla de frecuencias acumuladas. B) Gráfica de dispersión de puntos. Cada punto representan la razón  $d_N/d_S$  calculada para cada gen. En morado el utgroup, naranja Linaje I, azul Linaje II, verde Linaje III y rojo Linaje IV.

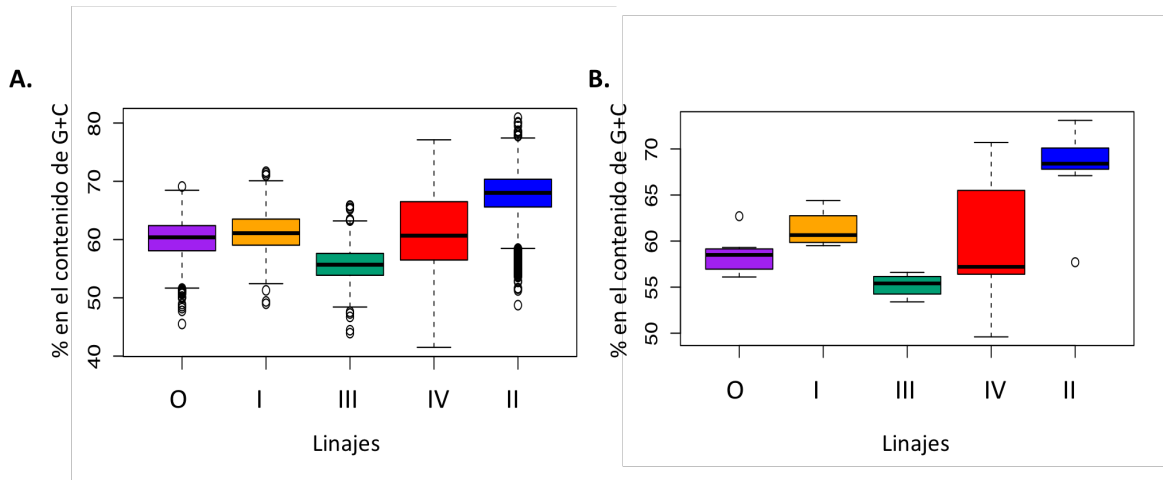
En resumen, nuestros resultados del análisis evolutivo, coinciden con la distribución taxonómica reportada anteriormente para la familia *Actinomycetaceae* (Zhao *et al.* 2014) en donde se ha postulado que el género *Actinomyces* puede subdividirse en más de dos linajes evolutivos como lo muestra la Figura 19. Así mismo, aunque se ha postulado que la evolución de estos linajes está dada por eventos de pérdidas y ganancias de genes, se desconoce la presión de selección que actúa sobre los genes conservados en dichos linajes. Lo que encontramos es que los organismos del Linaje II,



acumularon más mutaciones en etapas tempranas, en comparación con los organismos del Linaje IV, lo que les permitió separarse de los otros. Esta divergencia evolutiva podría relacionarse con otras características del genoma como diferencias en el tamaño, el contenido de (G+C) así como contenido de funciones metabólicas.

## 7.5 TAMAÑO DEL GENOMA Y CONTENIDO DE (G+C) DE LOS LINAJES EVOLUTIVOS

Para determinar las diferencias genómicas entre los cuatro linajes, calculamos el tamaño promedio de cada linaje así como el porcentaje en el contenido de (G+C). Nuestros resultados muestran que los organismos pertenecientes al linaje III son los que poseen los genomas de menor tamaño (2.18 Mpb) y más bajo contenido de (G+C) (55.1 %). Los organismos que forman parte del linaje IV, también presentan genomas pequeños, sus genomas miden en promedio 2.19 Mpb y poseen un contenido de (G+C) de 60.66%. Los organismos que se agrupan en el linaje III poseen genomas de tamaño promedio de 2.24 Mpb y de un contenido de (G+C) de 61.3%. En comparación con los otros los organismos del linaje II poseen los genomas más grandes de 3.03 Mpb en promedio y contenido de (G+C) más alto 68.3% (Figura 21A). Lo que se relaciona con su distribución taxonómica y las tasas de sustitución nucleotídicas. Dado que para calcular el contenido de (G+C) se utilizaron los genomas completos, era posible que el resultado obtenido podía estar enmascarado por eventos de THG debido a la dinámica de pérdida y ganancia de genes que se ha observado en éstos organismos, decidimos descartar ésta posibilidad calculando el contenido de (G+C) sólo para el grupo de proteínas ortólogas. Nuestros resultados mostraron que los porcentajes en el contenido de (G+C) de los grupos de ortólogos son muy similares a los obtenidos utilizando los genomas completos (Figura 21B) . Por lo que podemos concluir que estas diferencias en el contenido de (G+C) y tamaño del genoma son características específicas que distinguen a cada linaje.



**Figura 21.** Diferencias en contenido de (G+C) entre los Linajes de la familia *Actinomyceataceae*. A) Contenido de (G+C) en genes ortólogos, B) Contenido de (G+C) entre genomas completos. En el eje de las “x” se muestra la nomenclatura utilizada a cada Linajes, O=Out-group. La distribución de los datos se representa con un gráfico tipo caja, en donde se muestra la mediana con una línea negra y la distribución en cuartiles. Los *outliers* se representan en círculos, los cuales corresponden al contenido de (G+C) por gen o por genoma respectivamente.

Así mismo, nuestros resultados sugieren que los organismos que pertenecen a los linajes I, III y IV están en un proceso de reducción del genoma dirigido por la pérdida de genes como se ha observado en otros genomas bacterianos (Moran y Ochman 2006). De igual forma, se ha postulado que los organismos con genomas pequeños tienden a perder muchos de los genes involucrados en los mecanismos de reparación del ADN, lo que resulta en la deaminación de la citosina y la oxidación de la guanina, junto con los efectos de la muletera de Muller y la relajación en la selección purificante, estos cambios crean un sesgo hacia la fijación de mutaciones tipo A o T en la población (Sueoka 1962; Muto y Osawa 1987). De manera interesante, a pesar que se ha observado que una de las características del *filo* Actinobacteria es el alto porcentaje en el contenido de (G+C) mayor al 70% (Barka *et al.* 2016), nuestros resultados muestran que organismos los linajes propuestos poseen un contenido de (G+C) bajo y tamaños de genoma reducidos. Para determinar si éste cambio en el contenido de (G+C) se debe a la pérdida de los mecanismo de reparación entre

cada linaje, buscamos por homología cada una de las proteínas implicadas en estos mecanismos de reparación según Moran *et al.* 2006. Los resultados se muestran en la Tabla 11.

<i>Organismos/Mecanismos</i>	Reparación de bases y escisión	Ruta RecBCD	Ruta RecFOR	Ruta RecA	RecA y MutS	RecA y RecX	DinG	UvrABC	Uracil- ADN
<i>Actinobaculum massiliae</i>	1	1	1	1	0	0	1	1	1
<i>Actinobaculum schaalii</i>	1	1	1	1	0	0	1	1	1
<i>Actinomyces cardiffensis F0333</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces coleocanis</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces dentalis</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces europaeus</i>	1	1	1	1	0	0	1	1	1
<i>Actinomyces georgiae</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces gerencseriae</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces graevenitzii F0530</i>	1	1	1	1	0	0	1	1	1
<i>Actinomyces massiliensis</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces naeslundii</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces neuui</i>	1	1	1	1	0	0	1	1	1
<i>Actinomyces odontolyticus ATCC</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces oris MG1</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces slackii</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. ICM39</i>	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. oral taxon 170 str. F0386</i>	1	1	1	1	1	1	1	1	1

<i>Actinomyces sp. oral taxon 171 str. F0337</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. oral taxon 178 str. F0338</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. oral taxon 180 str. F0310</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. oral taxon 448 str. F0400 (888056.7)</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. oral taxon 848 str. F0332</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. oral taxon 849 str. F0330</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces sp. ph3</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces suimastitidis</i>	1	1	1	1	1	0	0	1	1	1
<i>Actinomyces timonensis</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces turicensis</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces urogenitalis</i>	1	1	1	1	1	1	1	1	1	1
<i>Actinomyces vaccimaxillae</i>	1	1	1	1	1	1	1	1	1	1
<i>Beutenbergia cavernae</i>	1	1	1	1	1	1	1	1	1	1
<i>Bifidobacterium adolescentis</i>	1	1	1	1	1	0	0	0	1	1
<i>Bifidobacterium bifidum</i>	1	1	0	1	1	0	0	0	1	1
<i>Bifidobacterium breve</i>	1	1	1	1	1	0	0	0	1	1
<i>Bifidobacterium catenulatum</i>	1	1	1	1	1	0	0	0	1	1
<i>Bifidobacterium</i>	1	1	1	1	1	0	0	0	1	1

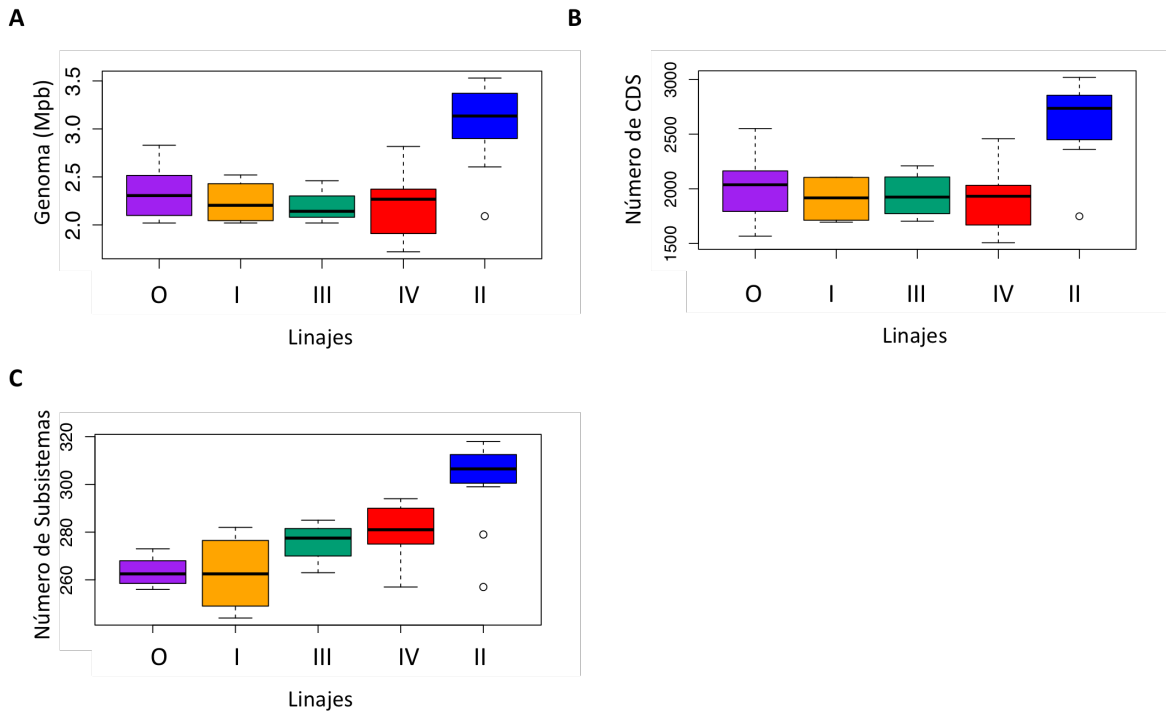
<i>dentium</i>										
<i>Bifidobacterium gallicum</i>	1	1	1	1	1	0	0	1	1	1
<i>Bifidobacterium longum</i>	1	1	1	1	1	1	1	0	1	1
<i>Bifidobacterium pseudocatenulatum</i>	1	1	0	1	1	0	0	0	1	1
<i>Mobiluncus curtisii subsp. curtisii ATCC</i>	1	1	1	1	1	0	0	1	1	1
<i>Mobiluncus mulieris</i>	1	1	1	1	1	0	0	1	1	1
<i>Trueparella pyrogenes</i>	1	1	1	1	1	0	0	1	1	1
<i>Varibaculum cambriense</i>	1	1	1	1	1	1	1	1	1	1

**Tabla 11.** Mecanismos de reparación presentes/ausentes en la familia *Actinomycetaceae*

Encontramos que algunos organismos pertenecientes al linaje IV han perdido los mecanismos relacionados con el sistema RecA, MutS y RecX (Tabla 11). Por un lado la proteína RecA y RecX se encuentran en un operon y regulan la activación de genes que participan en el sistema de reparación SOS, mientras que MutS reconoce las bases mal apareadas y se une a ellas. Es probable que al no estar presente RecA, genes que se regulan por ésta proteína estén siempre expresados, sin embargo se ha visto que la mutación en RecX es letal dado que no suprime los efectos tóxicos de RecA (Vierling *et al.* 2000). Con base en estos resultados podemos decir que los organismos del Linaje IV están acumulando daño celular relacionado con los mecanismos de reparación del ADN y explica el sesgo hacia el contenido de (A+T).

## 7.6 DIFERENCIAS EN EL NÚMERO DE SECUENCIAS CODIFICANTES ENTRE LOS LINAJES

Para tener una aproximación de las capacidades metabólicas de cada linaje decidimos cuantificar el contenido de genes que codifican para proteínas, calculamos los promedios de CDS, por sus siglas en inglés de Coding Sequences y el número de subsistemas, roles funcionales asociados con un proceso celular o la síntesis de algún metabolito, presentes en cada en cada linaje. Obtuvimos que los organismos del linaje II, tienen más genes que codifican para proteínas y a su vez éstas se agrupan en más subsistemas y que los organismos del linaje III son los que presentan menos CDS y menos funciones asociadas a un subsistema. Mientras que los organismos del Linaje IV, presentan un bajo número de CDS y se agrupan en más subsistemas (Figura 22). Esto sugiere que la red metabólica de los organismos pertenecientes a este linaje es más compleja y pueden ser capaces de llevar a cabo más funciones metabólicas en comparación con los organismos de los Linajes I y III y no así con los del Linaje II. Además, a nivel funcional, parece ser que los organismos pertenecientes a los linajes II y IV son más parecidos, lo que se relaciona con su origen evolutivo. Estos resultados sugieren que, los organismos que provienen de los Linajes I, III y IV podrían estar perdiendo genes, incluidos los de las rutas de biosíntesis de aminoácidos como se ha observado ocurre durante el deterioro genómico ya que los aminoácidos son aportados por el medio en el cual habitan los organismos (Moran y Ochman 2006).



**Figura 22.** Diferencias en el tamaño del genoma, número de secuencias codificantes y subsistemas entre miembros de la familia *Actinomycetaceae*. A) Diferencias en el tamaño del genoma, expresado en mega pares de bases. B) Diferencias en el número de CDS (Secuencias codificantes) y C) Diferencias en el número de funciones asignadas a un subsistema basado en la anotación de RAST. En el eje de las “x” se muestra la nomenclatura utilizada a cada Linajes, O=Out-group. La distribución de los datos se representa con un gráfico tipo caja, se muestra la mediana (línea negra) y la distribución en cuartiles. Los *outliers* se representan en círculos.

En resumen, juntos todos estos resultados, sugieren que los organismos que se encuentran en la parte basal de la filogenia de las Actinobacterias, son organismos que poseen genomas pequeños, un bajo contenido de (G+C), y menos funciones metabólicas, en comparación con Actinobacterias que se separan más tarde del ancestro común. En la familia *Actinomycetaceae* la formación de los Linajes evolutivos podría relacionarse con una especialización del nicho, en cuyo caso los organismos que forman parte del linaje II tuvieron que adaptarse rápidamente a éstos cambios, acumulando mutaciones como se observó en el análisis de dn/ds. Por lo que, la selección actuó a diferentes niveles en los diferentes linajes. Los organismos del Linaje II, se caracterizan por poseer

tamaños de genomas, contenido de (G+C) y número de CDS mayores en comparación con los otros linajes. Un tamaño de genoma “grande” les permite codificar para más funciones metabólicas las cuales se agrupan en un mayor número de subsistemas y tal vez esto les de alguna ventaja adaptativa, relacionada con la patogénesis reportada para especies de éste linaje (That *et al.* 2010). La pérdida de genes implicados en la reparación del ADN explica la disminución en el contenido de (G+C) observado en los Linajes I, III y IV. A pesar de esto, su evolución ha estado sujeta a una fuerte presión de selección. Tal vez en estos casos la presión de selección aumenta para mantener genes que son esenciales. A un nivel general, el estudio de esta familia basal del *filo* Actinobacteria, podría servir como modelo para entender cómo es que en las Actinobacterias se seleccionó el alto porcentaje en el contenido de (G+C) que observamos en organismos que se separan tarde, se pudo haber seleccionado para dar una mayor regulación a la transcripción de genes, ya que regiones con alto contenido de (G+C) son regiones que se transcriben lentamente. También, esta selección para el uso de codones "raros" como se les ha llamado a estos codones ricos en G+C pudo haberse seleccionado como un mecanismo de defensa, así cualquier otro organismo que se beneficiara de la adquisición de algún gene proveniente de Actinobacteria tendría que invertir en mantenerlo y transcribirlo. Por el contrario, cualquier gene nuevo con un bajo contenido de (G+C) que se integrara al genoma Actinobacteriano, sería transcrito más rápidamente por lo que sería fácil mantenerlo en genoma.



## CAPÍTULO 2

### 7.7 EVOLUCIÓN DEL METABOLISMO DE LA FAMILIA *ACTINOMYCEATAACEAE*

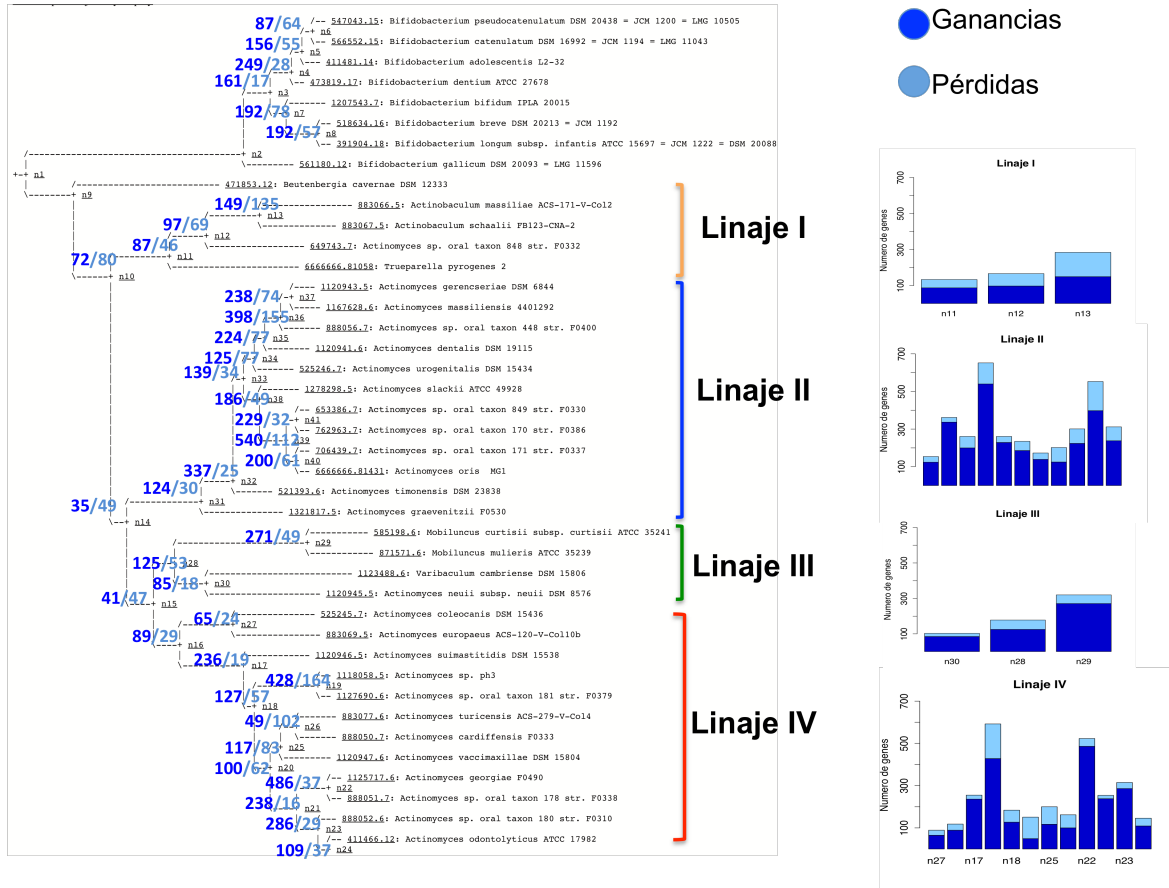
Como demostramos en el capítulo anterior, durante la evolución de la familia *Actinomycetaceae* surgieron cuatro linajes evolutivos. Tres de ellos, el I, III y IV, están en proceso de reducción del genoma, por lo que se espera encontrar pérdidas de genes incluidas las rutas de biosíntesis de L-histidina y L-triptófano. Proponemos que estas pérdidas de genes podrían estar asociadas con la ubicación de la secuencia de PriA de *A. odontolyticus* como posible intermediario evolutivo (Noda García *et al.* 2015). En el siguiente capítulo, nos enfocaremos en detectar pérdidas de genes asociadas a estas rutas de biosíntesis y su relación con la evolución de la especificidad por sustrato en PriA.

#### 7.7.1 ANÁLISIS DE PÉRDIDAS Y GANANCIAS DE FAMILIAS DE PROTEÍNAS EN *ACTINOMYCETACEAE*

Para entender cómo ha sido el proceso de evolución de estos linajes por pérdida de genes, decidimos usar la herramienta que está disponible en RAST, llamada *What change* (Ross *et al.* Sin publicar). Utilizando como plantilla el árbol filogenómico de la familia *Actinomycetaceae* (Figura 19), mapeamos las pérdidas y ganancias de genes que ocurrieron en cada uno de los nodos que dieron origen a los cuatro linajes evolutivos. Nuestros resultados muestran que, a pesar de la diferencia en el número de taxones que forman parte de cada linaje, es evidente que el linaje II adquirió el mayor número de genes desde su divergencia (nodo 31 y 32, Figura 23). En contraste, los linajes I, III y IV fueron los que a partir de su divergencia (nodos 11, 15 y 28, respectivamente) perdieron más genes de los que ganaron. Dichos resultados sugieren que mientras la evolución del Linaje II está dirigida por ganancias masivas de genes, la evolución de los linajes I, III y IV está

dirigida por pérdidas masivas de genes, lo que se relaciona con los tamaños de genoma pequeños y el bajo número de funciones metabólicas detectadas en dichos linajes. Por otra parte, es altamente probable que los linajes I, III y IV estén evolucionando hacia un estilo de vida simbiótico o intracelular, cómo se ha observado en otras bacterias en proceso de reducción de su genoma, principalmente en rutas de biosíntesis de aminoácidos (Moran *et al.* 2006 & 2012).

Al menos a primera vista pareciera que no existe una tendencia de pérdida genes, por ejemplo de ciertas rutas metabólicas, por lo que es probable que estos procesos de pérdidas sean aleatorios y por ello no tienen un efecto en la adecuación de dichos organismos. Es decir, la mayoría de estos genes, dadas las condiciones ambientales, son dispensables más que esenciales. Sin embargo, para poder determinar esto, se necesitaría llevar a cabo experimentos en medios adecuados para determinar el efecto de ciertas pérdidas en la adecuación. Así mismo, cabe resaltar que el linaje II es el más divergente, por lo que en la siguiente subsección nos enfocaremos en el estudio de la dinámica génica, tanto pérdidas como ganancias, que han ocurrido en este linaje.



**Figura 23 .** Pérdidas y ganancias de familias de proteínas por nodo. Los números en los nodos de la filogenia representan el número de familias de proteínas ganadas (azul marino) y las familias de proteínas que se pierden (azul cielo). A la derecha se muestra la distribución de las pérdidas y ganancias.

### 7.7.2 FAMILIAS DE PROTEÍNAS ESPECÍFICAS AL LINAJE II

Basados en el análisis de pérdidas y ganancias de genes de la sección 7.7.1 decidimos buscar familias de proteínas específicas para cada linaje. Para este fin utilizamos como punto de partida al Linaje II, porque es el que más genes ha ganado, y por ende poder determinar si a nivel metabólico es el más divergente. Para este análisis también se utilizó la búsqueda de familias de proteínas mediante *What change*, y tomamos como punto de partida el nodo 31 y 32 que es en donde ocurre

la divergencia. Los resultados obtenidos se muestran en la Tabla 15, en donde un *score* o puntaje de 1 sugiere que la familia de proteínas es única al Linaje (nodo 14 n14, Figura 23). A medida que el *score* va disminuyendo, esto sugiere que las familias de proteínas no sólo se encuentran en el Linaje II, sino que también pueden ser detectadas en los otros linajes I, III y IV. De este análisis observamos que desde su divergencia, el Linaje II ha mantenido 93 familias de proteínas que lo distinguen de los otros linajes, de éstas 96 sólo seis son únicas para este Linaje. Entre estas proteínas se encuentra ParA, la cual participa durante la replicación del ADN y de plásmidos (Bignell y Thomas 2001); 1-*sn*-glicerol-3-fosfato acil-transferasa, que participa en el metabolismo de lípidos (Lands 1960); la succinato deshidrogenasa, que es una proteína que ha sido asociada con un metabolismo anaeróbico (Iwasaki *et al.* 2002); y tres proteínas hipotéticas (Tabla 13). Así mismo, durante la divergencia, el Linaje perdió una chaperona de choque térmico CspC, la cual se encuentra altamente conservada en los linajes I, III y IV.

Score	Tamaño promedio de la proteína	Función	Linaje I	Linaje II	Linaje III	Linaje IV	Adquirida desde	Pérdida en
1	256.75	Succinate dehydrogenase cytochrome b subunit	presente			n14		
1	271	1-acyl- <i>sn</i> -glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	presente			n14		
1	85.833	hypothetical protein	presente			n14		
1	126.4	Cold shock protein CspC	presente		presente	presente		n14
0.966	422	3-oxoacyl-[acyl-carrier-protein] synthase, KASII (EC 2.3.1.179)	presente			n14,11		
0.966	574.923	ABC transporter, ATP-binding subunit	presente			n14,n9		
0.966	108.154	Thioredoxin		presente			n14,29	
0.966	1148.769	conserved hypothetical hypothetical protein	presente		presente	presente		n14,9
0.966	465.154	D-alanyl-D-alanine carboxypeptidase	presente		presente	presente		n14,9
0.929	293.143	hypothetical protein	presente			n14,11		
0.929	312.286	Hydrogen peroxide-inducible genes activator	presente			n14,12		

<b>0.929</b>	192.067	putative two-component system response regulator	presente	n14,9,11
<b>0.929</b>	240.357	DNA-binding response regulator, LuxR family	presente	n14,27,29
<b>0.909</b>	394.909	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	presente	n14,29,9
<b>0.909</b>	243.818	hypothetical protein	presente	n14,28
<b>0.909</b>	212.667	Two-component response regulator	presente	n31
<b>0.909</b>	290	hypothetical protein	presente	n31
<b>0.909</b>	286.727	Methionine ABC transporter substrate-binding protein	presente	n31
<b>0.909</b>	568.545	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3)	presente	n31
<b>0.909</b>	301.818	Chromosome (plasmid) partitioning protein ParA	presente	n14
<b>0.909</b>	260	putative azaleucine resistance protein AzlC	presente	n31
<b>0.909</b>	389.846	RNA methyltransferase, TrmA family	presente	n31
<b>0.909</b>	268.636	hypothetical protein	presente	n31
<b>0.909</b>	226.818	hypothetical protein	presente	n31
<b>0.909</b>	640.182	kinase domain protein	presente	n31
<b>0.909</b>	488.545	Dihydroxyacetone kinase family protein	presente	n31
<b>0.889</b>	165.333	Thiol peroxidase, Tpx-type (EC 1.11.1.15)	presente	n31
<b>0.875</b>	423.583	Beta-hexosaminidase (EC 3.2.1.52)	presente	n14 n36
<b>0.875</b>	225.833	two component transcriptional regulator, LuxR family	presente	n12,28,14
<b>0.875</b>	104.167	hypothetical protein	presente	n31 n9
<b>0.875</b>	114.25	hypothetical protein	presente	n31,n17
<b>0.875</b>	369.083	FIG019733: possible DNA-binding proteína	presente	n9, 31
<b>0.875</b>	304.583	1-phosphofructokinase (EC 2.7.1.56)	presente	n9,31
<b>0.875</b>	758.667	ATP-dependent DNA helicase SCO5183	presente	n9,n14
<b>0.875</b>	127	hypothetical protein	presente	n12,31
<b>0.875</b>	91.5	hypothetical protein	presente	n9,n31
<b>0.875</b>	429.333	Possible membrane protein	presente	n9,31
<b>0.875</b>	603.25	Cell division protein FtsI [Peptidoglycan synthetase] (EC 2.4.1.129)	presente	n12,n31
<b>0.875</b>	569.167	DNA repair helicase	presente	n11,n14
<b>0.847</b>	271.562	Twin-arginine translocation protein	presente	n14,n30

		TatB					
<b>0.847</b>	100.625	hypothetical protein	presente			n9, 31	
<b>0.847</b>	278.333	Carbamate kinase (EC 2.7.2.2)	presente			n9,12,14,16	
<b>0.847</b>	369.188	NADH dehydrogenase (EC 1.6.99.3)	presente	presente	presente		n9,11,14, 30
<b>0.839</b>	748	Isocitrate dehydrogenase [NADP] (EC 1.1.1.42); Monomeric isocitrate dehydrogenase [NADP]	presente	presente	presente	presente	n13,17,25,14,9
<b>0.839</b>	505	Lipid A export ATP-binding/permease protein MsbA	presente	presente		n10,	n15
<b>0.839</b>	180.846	hypothetical protein	presente			n21,31	
<b>0.839</b>	261.857	Cytoplasmic copper homeostasis protein cutC(1 family)	presente			n21,32	
<b>0.839</b>	434.667	Lipid A export ATP-binding/permease protein MsbA	presente			31,11,30	
<b>0.839</b>	400.538	possible peptidase	presente			n21,31	
<b>0.839</b>	234.231	hypothetical protein	presente			n21,32	
<b>0.839</b>	139.154	hypothetical protein	presente			12,26,31	
<b>0.801</b>	245.909	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	presente			9,31	
<b>0.801</b>	297.3	Phosphoserine phosphatase (EC 3.1.3.3)	presente			31	36
<b>0.801</b>	165.3	hypothetical protein	presente			32	
<b>0.801</b>	99.6	hypothetical protein	presente			32	
<b>0.801</b>	313.588	hypothetical protein	presente			32	36
<b>0.801</b>	248.6	hypothetical protein	presente			32	
<b>0.801</b>	389	sensor histidine kinase	presente			9,10,14,30	
<b>0.801</b>	533.2	UDP-N-acetylmuramoylalanyl-D-glutamate--L-ornithine ligase	presente			31	33
<b>0.801</b>	76.5	protein-N(pi)-phosphohistidine--sugar phosphotransferase( EC:2.7.1.69 )	presente			14	36,38
<b>0.801</b>	639.733	Ferrous iron transport protein B	presente			31	37
<b>0.801</b>	232.765	DUF1794		presente			14 36,38
<b>0.801</b>	217.7	hypothetical protein	presente			31	
<b>0.801</b>	272.5	hypothetical protein	presente			30	11,14
<b>0.801</b>	260.526	hypothetical protein	presente			32	
<b>0.801</b>	130.2	hypothetical protein	presente			31	36
<b>0.801</b>	91.24	hypothetical protein	presente			9,10,14,29,30	
<b>0.801</b>	273.9	Protein serine/threonine phosphatase PrpC, regulation of stationary phase	presente			31	36
<b>0.801</b>	362.636	hypothetical protein	presente			31	34
<b>0.801</b>	297.278	hypothetical protein	presente			32	34

0.801	271.6	hypothetical protein	presente	9,10,14,29,30	
0.801	235.9	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	ausente		31
0.801	250.8	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)		31	
0.8	566.357	Formate--tetrahydrofolate ligase (EC 6.3.4.3)(5 families)	presente		31
0.8	469.714	FIG00549624: hypothetical protein	presente		31
0.8	402.8	Cystathionine gamma-synthase (EC 2.5.1.48)	presente	13,14,18	35
0.8	444.286	N-acyl-L-amino acid amidohydrolase (EC 3.5.1.14)	presente	16,14	18,38
0.8	319.333	ABC transporter, ATP-binding protein	presente	8,12,31,9	
0.768	546.364	HtrA protease/chaperone protein	presente	12,18,31	
0.768	714.182	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	presente	25,14	
0.768	391.909	Aspartate aminotransferase (EC 2.6.1.1)	presente	32,9	
0.768	186.909	Putative MutT-family protein	presente	27,14	35
0.768	205.545	Transcriptional regulator, TetR family	presente	7,31	34
0.768	404.333	hypothetical protein	presente	9,14	32,40
0.768	481.462	6-phospho-beta-glucosidase (EC 3.2.1.86)	presente	31,9	34
0.768	430.545	Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE	presente	9,31	34
0.768	82.182	Acyl carrier protein(1 family)	presente	7,31	34
0.768	75.636	COG1264: Phosphotransferase system IIB components	presente	9,14	39
0.768	172.909	Alkylhydroperoxidase protein D	presente	9,32	
0.759	339.8	ABC-type Fe <sup>3+</sup> -siderophore transport system, permease component	presente	5,31	35
0.759	86	Copper chaperone	presente	12,17,14,9	41
0.759	336.85	ABC transporter, ATP-binding protein	presente	9,25	12,14,35
0.759	420.2	hypothetical protein	presente	31,20	
0.753	548.833	ABC transporter ATP-binding protein	presente	10,14	15
0.753	402.333	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	presente	12,18,23,14	
0.753	352.778	Threonine dehydrogenase and related Zn-dependent dehydrogenases	presente	12,18,23,15	

**Tabla 12.** Familias de proteínas que diferencian al linaje II de los otros linajes

Además, en la Figura 23 se observa que si excluimos al organismo *A. graevenetzi* del nodo 14, tenemos un grupo de organismos más homogéneo, es decir, si observamos los genes únicos a este linaje específicamente en el nodo 31, existen 20 genes únicos al Linaje II. Entre ellos se encuentran seis proteínas hipotéticas, además de las enzimas glicerol-fosforil di-éster fosfodiesterasa (EC 3.1.4.46), la cual participa en la utilización del grupo glicerol proveniente de algunos fosfolípidos y triglicéridos; formato-tetra-hidrofolato ligasa, involucrada en el metabolismo del carbono como única fuente de carbono durante condiciones aeróbicas (Marx *et al.* 2003); el determinante de resistencia a 4-azaleucina AzlC (Ward *et al.* 1973); dos metil-transferasas (RNA y TrmA) involucradas en la resistencia a antibióticos en bacterias (Vester & Long 2009); di-hidroxicetona cinasa (FakA), la cual pertenece a una familia de proteínas involucrada en el metabolismo de lípidos e implicada en *Staphylococcus aureus* como regulador del factor de virulencia (Parsons *et al.* 2014); thiol peroxidase, del tipo Tpx (EC 1.11.1.15), asociada con la adecuación al estrés oxidativo de *M. tuberculosis* en macrófagos.

Todas estas proteínas diferenciales presentes únicamente en el Linaje II, sugieren que dichos organismos han evolucionado para adaptarse y beneficiarse de los recursos del huésped, especialmente para sobrevivir bajo estrés de tipo oxidativo y para la utilización de lípidos exógenos. Esta adaptación a la utilización de lípidos exógenos es una estrategia que se ha observado en algunas bacterias Gram positivas como *Streptococcus* y *Staphylococcus*, la cual se integra a la síntesis de la membrana lipídica (Yao & Rock 2016). Dado que los organismos del Linaje II poseen un nicho parecido a las especies antes mencionadas, es probable que estén bajo una dinámica parecida y por eso los genes diferenciales se centran en la utilización de lípidos (Tabla 12). Esta observación concuerda con la hipótesis de que los organismos patogénicos de esta familia son los que se encuentran formando parte del Linaje II, incluida la cepa UTB18 cuyo genoma fue secuenciado en el capítulo anterior.

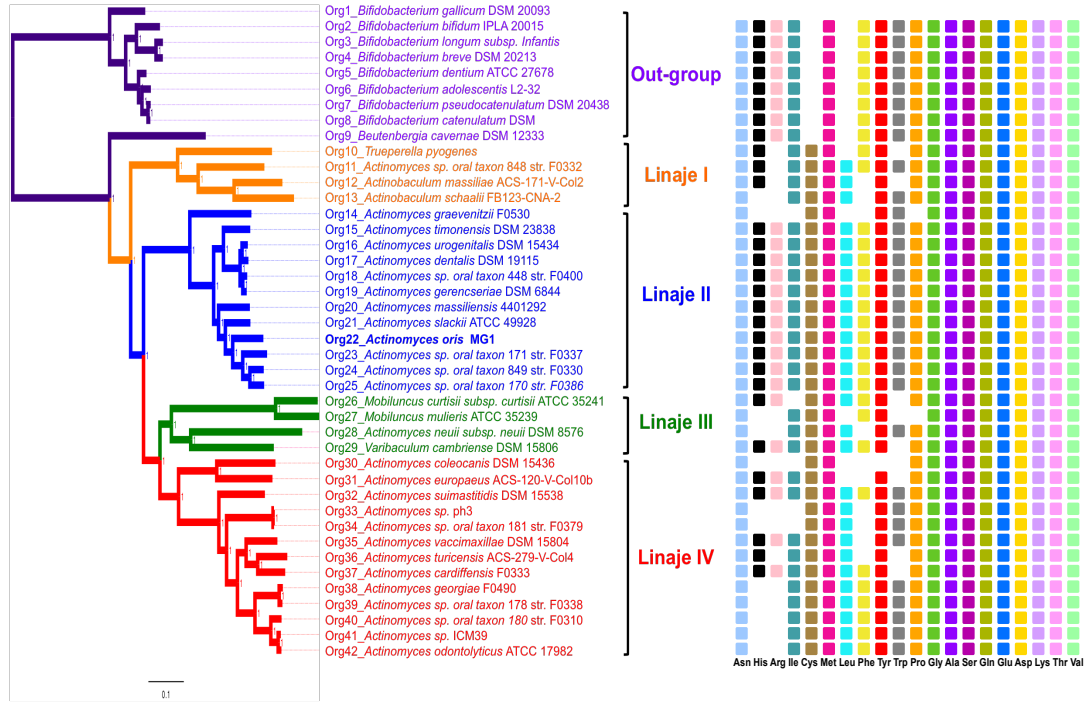


La importancia de estos resultados desde el punto de vista odontológico y médico es que algunos de estos genes podrían usarse como marcadores moleculares para el diagnóstico de infecciones asociadas con éstos organismos, o bien como posibles blancos farmacéuticos. Así mismo, sería interesante llevar a cabo un análisis de selección sobre estos genes para determinar si se encuentran bajo selección positiva por la ventaja adaptativa que podrían conferir a los organismos del linaje II. Como punto de partida en este sentido decidimos explorar a detalle como se encontraban los genes asociados a las rutas de biosíntesis de L-histidina y L-triptófano, por lo que se llevó a cabo un análisis de ocurrencia (presencia y ausencia) de (i) rutas biosintéticas, con énfasis en estos dos aminoácidos y (ii) específicamente de la enzima PriA.

### 7.7. 3 DIFERENCIAS EN LAS RUTAS DE BIOSÍNTESIS DE AMINOÁCIDOS ENTRE LOS DISTINTOS LINAJES

Para determinar diferencias en las rutas de biosíntesis de aminoácidos entre los distintos Linajes, usamos como guía el árbol filogenético del capítulo anterior. Todos los análisis de presencia / ausencia de las rutas de biosíntesis de aminoácidos fueron llevados a cabo en RAST. Para ello, mediante similitud de secuencias buscamos en cada uno de los 33 genomas de la familia *Actinomycetaceae*, y del grupo externo, las proteínas ortólogas que se sabe catalizan alguna reacción en las 20 rutas de biosíntesis de aminoácidos, según KEGG (Kanehisa *et al.* 2004). Se determinó que una ruta no era funcional, es decir, no estaba presente, cuando alguna de las enzimas reportadas para la conversión de algún sustrato relacionado con la biosíntesis de cada uno de los 20 aminoácidos, estaba ausente. Bajo este criterio, detectamos una clara distribución entre aquellas rutas que se mantienen y aquellas que se pierden de manera específica para cada linaje. Nuestro análisis arrojó que en las especies del género *Bifidobacterium* los genes *cysKDCH*, necesarios para la biosíntesis de L-cisteína, no están presentes en ninguno de los genomas analizados, por lo que esta ruta fue tomada como ausente. Esto sugiere que durante la evolución de este género la ruta

biosintética de la cisteína se pierde, adquiriéndose un transportador de cisteína que permite la incorporación de dicho aminoácido. Analizando a detalle las rutas de biosíntesis en *Bididobacterium*, se encontró que la biosíntesis de L-glicina, se lleva a cabo a partir de la de L-treonina por una treonina aldolasa, relacionado a esto, la biosíntesis de L-treonina está completa y ocurre vía *thrA-asd-hom-thrC-thrB*. En la biosíntesis de L-arginina, ArgJ cataliza el primer y quinto paso en la misma ruta metabólica, por lo que se trata de una enzima bi-funcional con actividad ArgA y ArgE (EC 2.3.1.1), la biosíntesis de L-metionina ocurre mediante un transportador de metionina, puesto que no tiene genes par la biosíntesis de novo. La biosíntesis de L-lisina ocurre de novo a partir de Aspartato. Las biosíntesis de L-Leucina, L-triptófano, L-Tirosina y L-Fenilalanina ocurren de novo a partir de corismato, la L-prolina la sintetizan a partir de L-Glutamato y sintetizan L-alanina mediante los genes canónicos utilizando una alanina racemasa. Estos organismos pueden sintetizar L-serina a partir de 3-fosfoglicerato con una serie de tres pasos catalizados por SerA, SerC y SerB (Figura 24).



**Figura 24.** Retención y pérdida en rutas de biosíntesis de aminoácidos en los diferentes organismos que conforman cada Linaje evolutivo dentro de la familia *Actinomycetaceae*, cada ruta de biosíntesis presente, se marca con un cuadro de color diferente, las ausencias se marcan en blanco.

Todos los organismos pertenecientes a la familia *Actinomyceataceae* parecen ser capaces de sintetizar L-glicina a partir de la L-serina y L-treonina: al igual que las *Bifidobacterium*, la biosíntesis de L-treonina está completa y ocurre vía *thrA-asd-hom-thrC-thrB*, además de que poseen a la enzima aspartato aminotransferasa. En todos los Linajes, la biosíntesis de L-lisina ocurre de novo a partir de Aspartato, y la biosíntesis de L-leucina también está completa. En algunos organismos del Linaje III y IV hay genes que se encuentran ausentes, por lo que podrían ser *missing genes* definidos como en Overmars & Overbeek (2003). Por ejemplo, en la biosíntesis de L-fenilalanina y L-tirosina están ausentes las proteínas aminotransferasa AroJI $\beta$  y AroJI $\gamma$ . Los organismos del Linaje II y IV sintetizan prolina a partir de L-glutamato. En algunos organismos de los linajes I y III, la síntesis de L-prolina no está clara, puesto que al menos una forma de Pirrolina-5-carboxilato reductasa podría estar codificada por ProC, ProG o ProI (presentes), mientras que

ProA y ProB están ausentes. Sin embargo, no se sabe si estos organismos puedan sintetizar el aminoácido. La biosíntesis de L-alanina en todos los linajes involucra a una alanina racemasa, y no es necesaria la presencia de IlvE. Los organismos del Linaje I pueden sintetizar L-serina a partir de L-glicina. Mientras que los Linajes II y IV sintetizan L-serina a partir de 3-fosfoglicerato con una serie de tres pasos catalizados por SerA, SerC y SerB. O bien pueden sintetizar L-serina a partir de L-glicina con la primera enzima de la biosíntesis de L-glicina GlyA, o los organismos del Linaje II y I pueden sintetizar L-metionina vía O-Succinil-L-Homoserina. Específicamente para el desarrollo de esta tesis, estos análisis predicen que algunos organismos del Linaje IV y III no pueden sintetizar L-histidina, debido a que no cuentan con los genes *hisBCDEFHG*. En cambio, dichos organismos parecen poseer un transportador que les permite utilizar histidina del medio. Por otro lado, la biosíntesis de L-Triptófano esta completa en los organismos del Linaje II; mientras que organismos de los linajes I, III y IV pierden los genes *trpABCDE* de la ruta de biosíntesis de L-triptófano de manera diferencial. Sin embargo, estos organismos conservan a la enzima bi-funcional PriA, la cual cataliza los pasos llevados a cabo por las enzimas HisA (biosíntesis de L-Histidina) y TrpF (biosíntesis de L-triptófano), al igual que el gen *trpE*. En resumen, estos análisis demuestran que los organismos que provienen del Linaje II conservan completas las rutas de biosíntesis de aminoácidos, mientras que los organismos provenientes de los Linajes I, III y IV pierden de forma diferencial algunas rutas de biosíntesis, entre ellas las de biosíntesis de L-histidina y L-triptófano, lo que podría relacionarse con un proceso de reducción del genoma tal cual se ha propuesto (Moran *et al.* 2006).

### 7.7.3.1 RELACIÓN ENTRE LA PÉRDIDA DE BIOSÍNTESIS DE AMINOÁCIDOS Y LA PROPORCIÓN DE AMINOÁCIDOS PRESENTES EN LAS SECUENCIAS DE PROTEÍNAS

Basados en nuestro árbol filogenético de especies, y en la propuesta de los cuatro linajes evolutivos en la familia *Actinomycetaceae*, decidimos determinar si los organismos de los linajes I, III y IV que pierden las rutas de biosíntesis de aminoácidos antes mencionadas (Figura 25) eran capaces de adaptar su proteoma a las limitaciones en estos aminoácidos. Esta posibilidad se ha observado en otros organismos patógenos, cuyo genoma está en detrimento, como *Chlamydia* (Lo CC *et al.* 2012). Una estrategia para determinar esto, es contar el número de aminoácidos presentes a lo largo de proteínas que están conservadas, como por ejemplo las proteínas ribosomales. Dado que los 20 aminoácidos se distribuyen de forma regular a lo largo de una proteína, una medida que permite determinar la proporción de cada aminoácido es el cálculo del coeficiente de determinación  $R^2$ , en donde se comparan los tamaños de las proteínas con la proporción de residuos. Para este análisis se usó el grupo de proteínas ortólogas que obtuvimos anteriormente, y calculamos el coeficiente de determinación para cada grupo de ortólogos.

Los resultados muestran diferencias en las proporciones de estos aminoácidos (Tabla 13), se observa que los aminoácidos como la histidina, el triptófano y la cisteína son los que presentan proporciones más bajas en comparación con los otros aminoácidos. Este resultado sugiere que la adaptación no es linaje específica, ya que todos los organismos presentaron proporciones bajas para estos aminoácidos, con valores de  $R^2 < 0.7$ . Por lo que, podemos concluir que no existe una relación entre la pérdida de la biosíntesis de aminoácidos con la adaptación del proteoma en estos organismos. Cabe resaltar que dadas las propiedades fisicoquímicas del residuo Cys, se espera que se encuentre en una baja proporción en las secuencias de proteínas.

Aminoácido/Linaje	Outgroup	Linaje I	Linaje II	Linaje III	Linaje IV
N	0.845	0.793	0.726	0.848	0.816
H	0.656	0.682	0.665	0.66	0.663
R	0.774	0.807	0.799	0.815	0.821
I	0.877	0.84	0.822	0.856	0.838
M	0.75	0.732	0.746	0.761	0.776
C	0.33	0.297	0.333	0.374	0.318
L	0.91	0.91	0.903	0.896	0.9
F	0.806	0.781	0.709	0.793	0.792
Y	0.794	0.772	0.765	0.767	0.776
W	0.447	0.443	0.444	0.459	0.449
P	0.884	0.844	0.847	0.864	0.848
G	0.918	0.918	0.916	0.92	0.911
A	0.846	0.846	0.855	0.853	0.866
S	0.823	0.875	0.821	0.861	0.856
Q	0.794	0.793	0.764	0.805	0.783
E	0.84	0.851	0.835	0.824	0.843
D	0.912	0.901	0.913	0.898	0.901
K	0.778	0.655	0.553	0.723	0.653
T	0.877	0.876	0.88	0.889	0.888
V	0.9	0.909	0.904	0.9	0.899

**Tabla 13.** Proporción de aminoácidos presentes en las secuencias de proteínas.

Se remarcen los aminoácidos con la proporción más baja en las secuencias de proteínas.

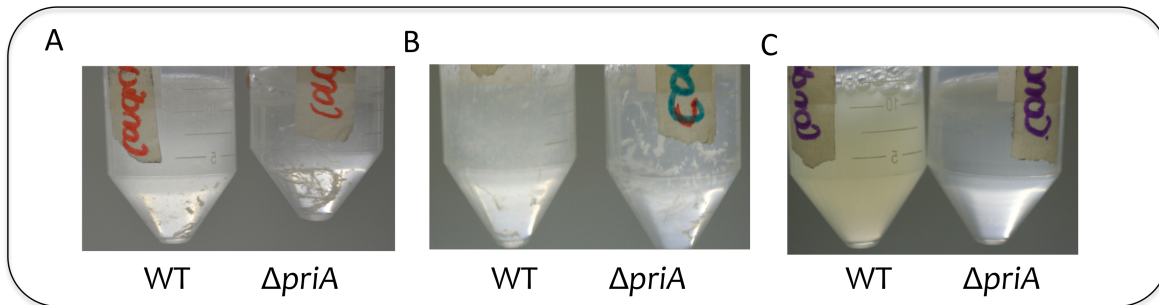
#### 7.7.4 ESENCIALIDAD DE LAS RUTAS DE BIOSÍNTESIS DE L-HISTIDINA Y L-TRIPTÓFANO EN ACTINOMYCES

La historia evolutiva de la familia *Actinomycetaceae* muestra que durante la evolución de los organismos pertenecientes a los Linajes I, III y IV, éstos perdieron diferencialmente genes de la

biosíntesis de L-histidina y L-triptófano (Figura 25). Este análisis sugiere la posibilidad de que los organismos que pierden estas rutas presenten auxotrofia para tales aminoácidos; ó bien que existan rutas alternas para la producción de éstos aminoácidos en dichos organismos, lo cual es poco probable. Para poder explorar estas posibilidades, decidimos construir una mutante en *priA* (actividad E.C 5.3.1.24, TrpF y E.C 5.3.1.16, HisA) utilizando como organismo modelo la cepa *A. oris* MG-1, cuyo genoma fue secuenciado en el capítulo anterior. Seleccionamos a esta cepa dado que (i) evolutivamente está muy relacionada con organismos del linaje I, III y IV, (ii) comparte el mismo nicho ecológico que muchos miembros de estos linajes, incluido *A. odontolyticus*, organismo que perdió los genes de biosíntesis de L-histidina, (iii) presenta todos los genes necesarios para la biosíntesis de L-histidina y L-triptófano, y (iv) existen métodos para su manipulación genética (That *et al.* 2011).

La cepa *A. oris* MG-1 se creció en un medio líquido sintético químicamente definido, con únicamente glucosa como fuente de carbono y de nitrógeno sulfato de amonio. Además, para determinar prototrofias se adicionaron los 20 aminoácidos esenciales y vitaminas como niacina y tiamina; dicho medio fue nombrado MM\_AA. Primeramente, se determinó que la cepa *A. oris* MG-1 WT y la mutante obtenida *A. oris* MG-1  $\Delta priA$  fuesen capaces de crecer en el medio MM\_AA, demostrándose que ambas son capaces de utilizar los componentes del medio (Figura 25A). El siguiente experimento consistió en crecer ambas cepas en el medio MM\_AA, pero sin histidina ni triptófano, medio MM\_AA\_DoHW. En este medio la cepa *A. oris* MG-1 WT fue capaz de crecer, mientras que la cepa *A. oris* MG-1  $\Delta priA$  fue incapaz de crecer. Esto sugiere que a nivel fisiológico en estas condiciones la biosíntesis de L-histidina y L-triptófano son relevantes para el organismo. Para determinar la relevancia de cada una de estas rutas, hicimos un tercer experimento, el cual consistió en crecer ambas cepas en medio MM\_AADoH y en MM\_AA\_DoW. En ambas condiciones la cepa *A. oris* MG-1 WT presentó crecimiento, mientras que, en ausencia de

triptófano, la cepa *A. oris* MG-1  $\Delta priA$  fue incapaz de crecer, mientras que en ausencia de histidina mostró crecimiento (Figura 26B-C).



**Figura 25.** Caracterización fenotípica de la cepa *A. oris* MG-1 en diferentes medios. A) Medio con todos los 20 aminoácidos, B) Medio con 19 aminoácidos sin histidina y C) Medio con 19 aminoácidos sin triptófano.

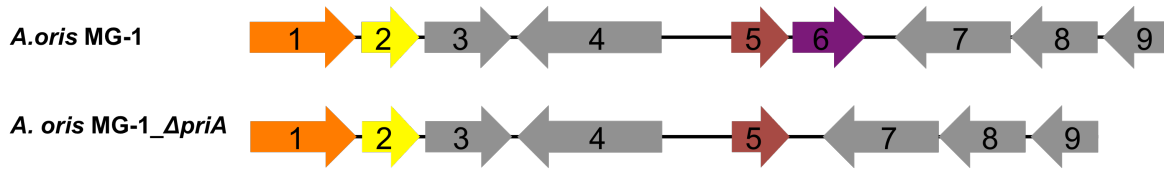
Para soportar estos resultados, y eliminar la posibilidad de que el crecimiento de la mutante en ausencia de histidina se debiese a un problema técnico durante la mutación del gene *priA*, decidimos secuenciar la mutante *A. oris* MG-1  $\Delta priA$ . Los resultados de la secuenciación de la cepa se muestran en la Tabla 14.



<b>Estadísticas</b>	<b>A. oris MG-1</b>	<b>A. oris MG-1 <math>\Delta priA</math></b>
<b>Tamaño (bp)</b>	2 930 837	3 078 390
<b>Numero de contigs*</b>	501	182
<b>Contenido G+C (%)</b>	68.2	68.4
<b>Tamaño del contig más corto</b>	69	128
<b>Median sequence size</b>	3 188	440
<b>Mean sequence size</b>	6 043	16 914
<b>Longest contig size</b>	65 533	380 414
<b>Número de secuencias codificantes</b>	2 742	2 800
<b>Número de RNAs</b>	53	53

**Tabla 14.** Estadísticas de secuenciación de la cepa *A. oris* MG-1  $\Delta priA$

Si comparamos los resultados obtenidos con los de la secuenciación de la cepa *A. oris* WT, obtuvimos una mayor cobertura y mejor ensamblado para la cepa mutante (número de contigs y tamaño del genoma). Más aún, estos resultados permitieron hacer una búsqueda por Blastp utilizando como semilla la secuencia de PriA de *A. oris* MG-1. El hit que se obtuvo, con un valor de expectancia (e-value) significativo fue con la proteína HisF, presente en ambos genomas. Para comprobar que la región en donde se hizo la mutación no afectó otros genes, buscamos los contigs en donde se ubica la mutación y los comparamos con la región de la cepa WT (Figura 26). Los resultados obtenidos muestran que el único gen eliminado corresponde a *priA*, y que la mutación no presenta ningún efecto polar en el genoma, al menos en la sección analizada.



Strain	Size (nt)	Strand	Function	ID
<i>A. oris</i> MG1	1143	+	EC 2.6.1.9	1
<i>A. oris</i> MG1	600	+	EC 4.2.1.19	2
<i>A. oris</i> MG1	915	+	MshA	3
<i>A. oris</i> MG1	1584	-	hypothetical protein	4
<i>A. oris</i> MG1	639	+	EC 2.4.2.-	5
<i>A. oris</i> MG1	753	+	<b>PriA</b>	6
<i>A. oris</i> MG1	1293	-	EC 2.3.1.51	7
<i>A. oris</i> MG1	936	-	EC 2.7.7.41	8
<i>A. oris</i> MG1	699	-	EC 2.7.8.5	9
<i>A. oris</i> MG1_Δ <i>priA</i>	1143	+	EC 2.6.1.9	1
<i>A. oris</i> MG1_Δ <i>priA</i>	600	+	EC 4.2.1.19	2
<i>A. oris</i> MG1_Δ <i>priA</i>	915	+	MshA	3
<i>A. oris</i> MG1_Δ <i>priA</i>	1584	-	hypothetical protein	4
<i>A. oris</i> MG1_Δ <i>priA</i>	639	+	EC 2.4.2.-	5
<i>A. oris</i> MG1_Δ <i>priA</i>	1293	-	EC 2.3.1.51	7
<i>A. oris</i> MG1_Δ <i>priA</i>	936	-	EC 2.7.7.41	8
<i>A. oris</i> MG1_Δ <i>priA</i>	699	-	EC 2.7.8.5	9

**Figura 26.** Contexto genómico de la mutación en *priA* de la cepa *A. oris* MG-1. Las flechas indican el sentido de la transcripción de cada gen. Los colores hacen alusión a la ruta en la cual esta involucrada PriA y los genes en gris son proteínas hipotéticas o no relacionadas con la biosíntesis de histidina.

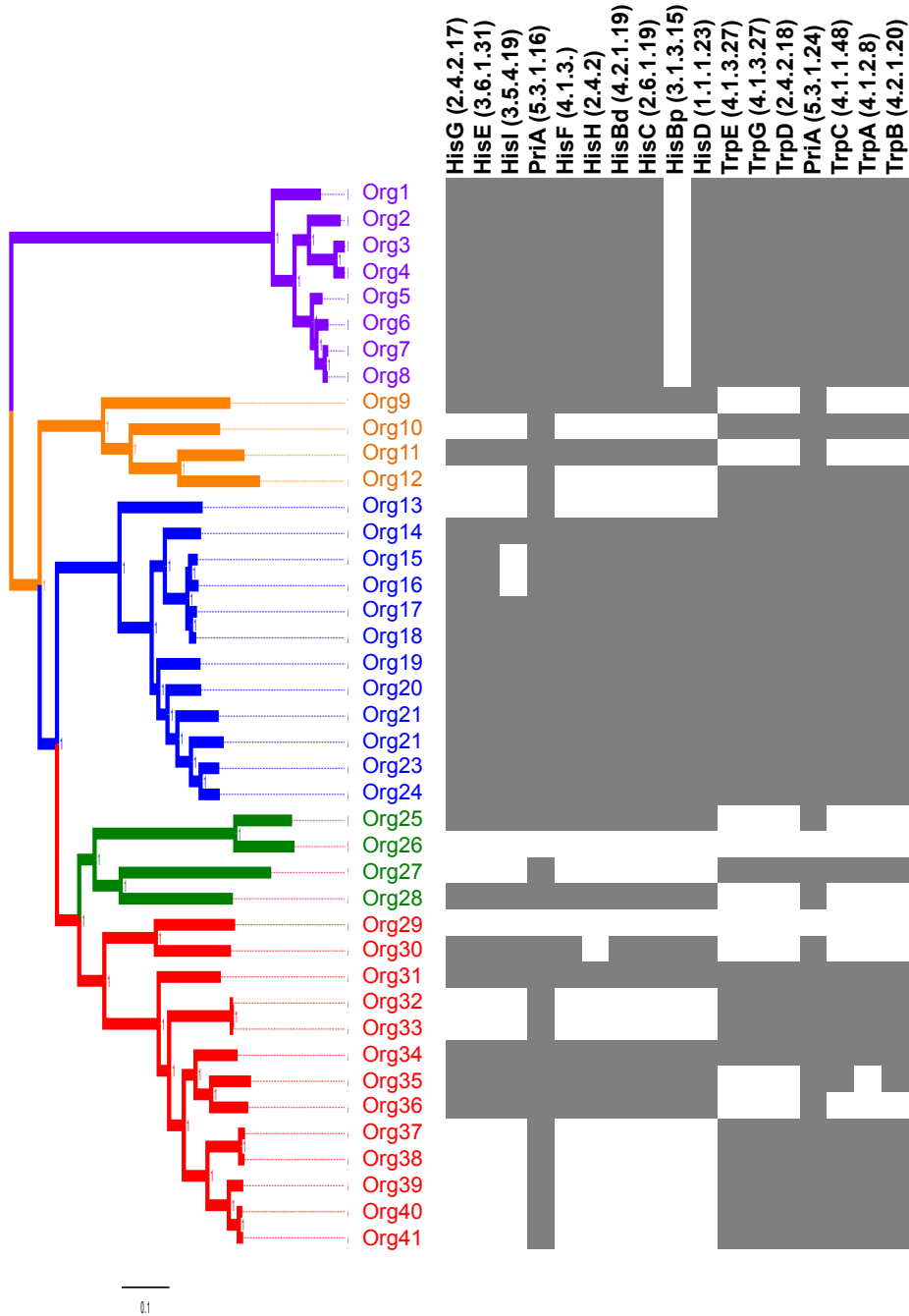
Estos resultados, además de confirmar la relevancia fisiológica de PriA cuando menos para un aminoácido como se discute en el siguiente párrafo, podrían ser explicados biológicamente por alguna de las siguientes posibilidades: (i) la existencia de una enzima con función ProFAR isomerasa E.C 5.3.1.16 de reciente evolución, cómo se demuestra en *Chlamydia* para la enzima FoIE E.C EC 3.5.4.16 (ii) que el producto de PriA, phosphoribulosylformimino-AICAR-P (siglas de su nombre en inglés) se encuentra presente en el medio celular como resultado de una ruta alternativa y (iii) que ante la escasez de histidina en el medio, estos organismos son capaces de minimizar su uso, es decir, existe una adaptación previa que se vería reflejada en la incorporación de estos aminoácidos a su proteoma. Todas estas posibilidades biológicas merecen por sí mismas de mayor investigación.

Todos estos resultados juntos muestran que en ausencia de triptófano en el medio la actividad PRA isomerasa codificada por el gen *priA* es esencial para el crecimiento de la cepa. Por lo que los

organismos han mantenido este gen a lo largo de su evolución. Además, se sugiere que si la histidina y el triptófano están presentes en el medio, los organismos maximizan su metabolismo tomando los aminoácidos del medio mediante transportadores. Es probable que lo que observamos sea el paso de ser un gen esencial a convertirse en uno dispensable. Por lo que como habíamos comentado en la introducción PriA parece ser un buen modelo para estudiar la evolución de la especificidad por sustrato durante la pérdida de genes. Para ello decidimos llevar a cabo un análisis más detallado de la co-ocurrencia (presencia ausencia) de los genes *his* y *trp* a lo largo de la familia *Actinomycetaceae*.

#### 7.7.5 ANÁLISIS DE CO-OCURRENCIA DE LOS GENES HIS Y TRP

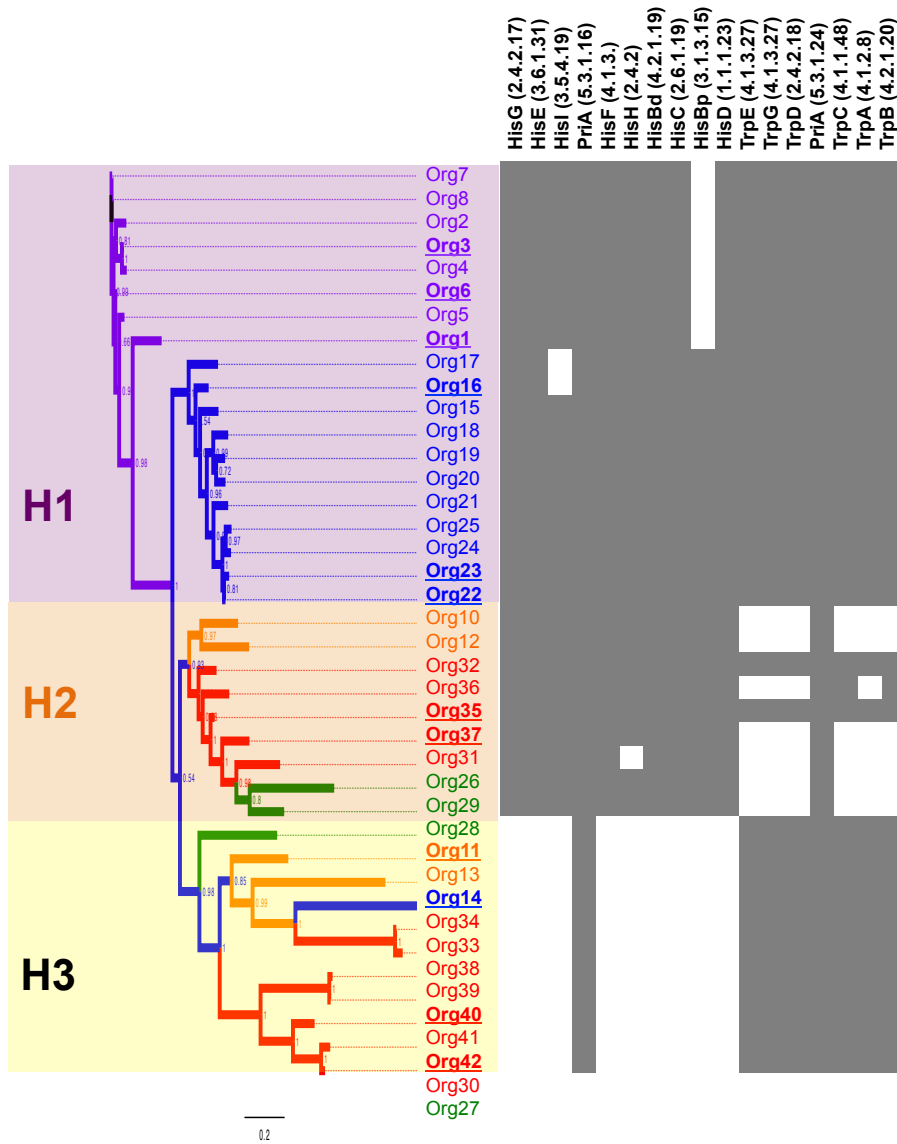
Generalmente, la conservación evolutiva de un gen es un buen indicador de su relevancia funcional (Korbel *et al.* 2004). En bacterias, la codirección génica y la posición de los genes vecinos se pueden conservar formando operones, lo que sugiere que están relacionados funcionalmente. Para deducir a nivel sistemático la significancia de la conservación de PriA en la familia *Actinomycetaceae*, tomamos la filogenia obtenida en la sección 7.3 para evaluar la presencia y ausencia de los genes *his* y *trp*. Posteriormente buscamos cada uno de los genes ortólogos de la biosíntesis de L-Histidina y L-Triptófano en cada genoma, y utilizando como mapa la filogenia propuesta, construimos una matriz de ocurrencias (Figura 27).



**Figura 27.** Análisis de ocurrencia de los genes *his/trp* utilizando como mapa la filogenia de especies. La presencia de un gen se denota en cuadros grises, mientras que la ausencia en cuadros blancos. En la parte superior se muestra el producto de cada gen con su respectivo número E.C. Dado que PriA está involucrada en ambas rutas, la enzima se marca en ambas rutas, sin embargo, sólo existe una copia génica por genoma.

En la Figura 27 se observa que los organismos que constituyen el grupo raíz (out-group) pierden al gen *hisBp*, que codifica para una Histidol fosfatasa. Siendo éste un paso alterno en la ruta de biosíntesis de L-histidina, a pesar de su ausencia, predecimos que estos organismos pueden sintetizar histidina. Interesantemente, en los linajes III (Org\_27) y IV (Org\_30) no se encuentran los genes *his/trp*, incluido el gen *priA*, sugiriendo que estos organismos se han adaptado a tomar dichos aminoácidos del hospedero. Los organismos del Linaje II poseen completos los genes *his/trp*, incluido el gen *priA*, por lo que asumimos que ambas rutas siguen siendo relevantes, como se mostró experimentalmente en la sección 7.7.4. Los organismos de los Linajes I, III y IV pierden de forma diferencial, ya sea los genes *his* o bien los genes *trp*, aunque conservan *priA*. Esto nos sugiere una asociación funcional ligada a la presencia de los genes *his/trp*.

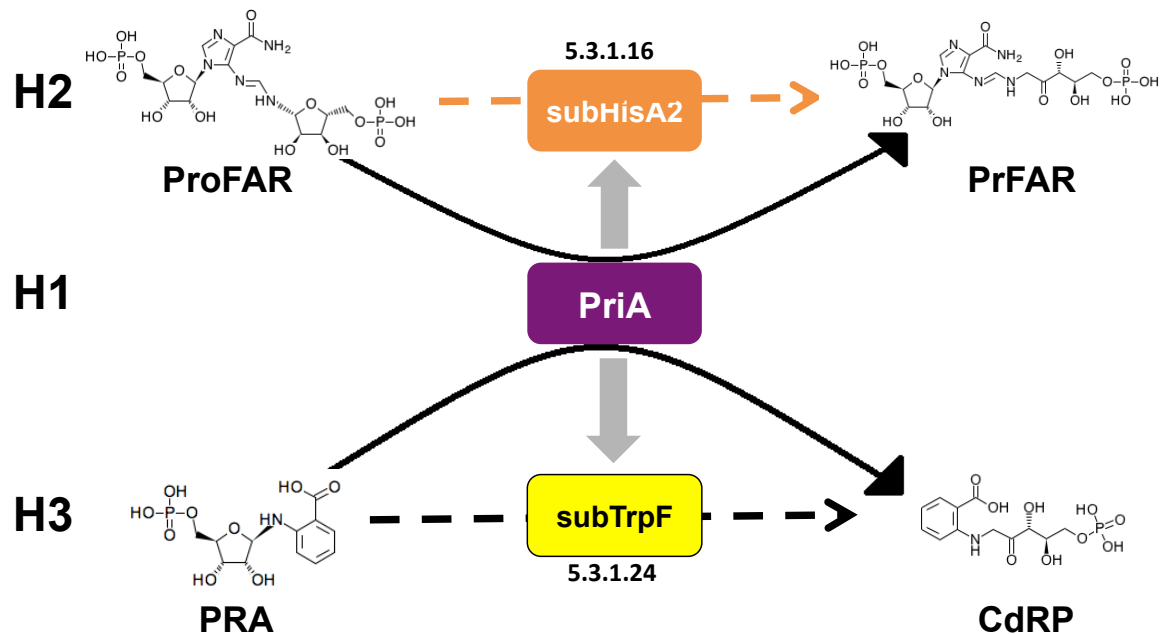
Basados en los resultados obtenidos decidimos utilizar a la proteína PriA como un marcador filogenético de la asociación de esta co-ocurrencia de genes. Así, construimos una filogenia utilizando a los ortólogos de PriA presentes en los 40 genomas analizados. El árbol resultante fue tomado como guía para mapear nuevamente la presencia y ausencia de genes (Figura 28). Nuestros resultados muestran una asociación funcional entre PriA y la co-ocurrencia de dichos genes, de manera inversa a las relaciones taxonómicas construidas con la filogenia de especies. Con base en la secuencia de PriA, se observan cuatro clados: en el primero se agrupan las PriAs que co-existen con todos los genes *his/trp* (Recuadro morado, Figura 28). En el Segundo clado se agrupan aquellas PriAs asociadas con la presencia únicamente de los genes *his* (Recuadro naranja, Figura 28). En el tercer clado se agrupan todas las PriAs asociadas únicamente a la presencia de los genes *trp* (Recuadro amarillo, Figura 28). Basados en estos resultados proponemos tres hipótesis evolutivas o asociaciones funcionales de PriA relacionadas con la co-ocurrencia de genes, llamadas H1, H2, H3.



**Figura 28.** Análisis de ocurrencia de los genes *his/trp* utilizando como marcador filogenómico a la enzima PriA. La presencia de un gen se denota en cuadros grises, mientras que la ausencia en cuadros blancos. En la parte superior se muestra el producto de cada gen con su respectivo número E.C. Dado que PriA está involucrada en ambas rutas, la enzima se marca en ambas rutas. Sin embargo, sólo existe una copia génica por genoma. Los recuadros denotan los clados con una asociación funcional.

### 7.7.6 ASOCIACIONES FUNCIONALES DE LOS ORTÓLOGOS DE PriA BASADOS EN LA CO-OCURRENCIA

Basados en nuestros resultados anteriores propusimos tres caminos evolutivos por los cuales PriA pudo haber evolucionado su especificidad por sustrato; por una parte, tenemos que los ortólogos de PriA que evolucionan en un contexto en donde se conservan ambas rutas biosintéticas (Linaje II, ocurrencia de genes *his/trp*), mantenga su bi-funcionalidad, es decir, PriA sea capaz de convertir los sustratos PRA y ProFAR (H1). Los ortólogos de PriA que evolucionan en un contexto en donde únicamente se pierden los genes *trp* (Linajes I, III y IV, caja naranja Figura 29, H2), han acumulado mutaciones que impiden aceptar PRA como sustrato, es decir, tienen actividad únicamente de ProFAR isomerasa. Mientras que aquellos ortólogos que evolucionan en un escenario en donde se han perdido únicamente los genes *his* (Linajes I, III y IV, caja amarilla, Figura 29, H3), se han sub-funcionalizado para aceptar al sustrato PRA, es decir, tienen únicamente actividad de PRA isomerasa. Con estas hipótesis proponemos que los ortólogos de PriA provenientes de la familia *Actinomycetaceae* han evolucionando su especificidad por sustrato debido a la pérdida en los genes *his* o *trp*, y proponemos que han dado origen a dos nuevas subfamilias las cuales proponemos nombrarlas como SubHisA2 (ProFAR isomerasa) y SubTrpF (PRA isomerasa) (Figura 29). La subfamilia SubHisA2 fue nombrada así para diferenciarla de la subfamilia SubHisA previamente reportada en el género *Corynebacterium* (Noda-García *et al.* 2013).



**Figura 29.** Hipótesis propuestas sobre la evolución de los ortólogos de PriA que dan origen a diferentes subfamilias. Las cajas representan a las subfamilias de PriA y su proceso evolutivo. Por encima de las cajas se muestran los números E.C. de cada subfamilia. En el caso de PriA, ésta posee dos números E.C.

## 7.8 PRESIONES DE SELECCIÓN DETECTADAS EN LOS ORTÓLOGOS DE PRIA

Nuestros análisis filogenéticos hechos con PriA, muestran que los ortólogos de PriA pueden agruparse en tres clados, por lo que se decidió determinar si las ramas que dan origen a cada uno de los clados pasó por diferentes tipos de presión de selección. Para ello utilizamos el modelo de ramas contenido en el programa codeml, el cual calcula el número de sustituciones no sinónimas sobre las sinónimas  $d_N/d_S$ , el cual se define como el valor de omega ( $w$ ). Para este fin, decidimos probar dos hipótesis, como se define en materiales y métodos (sección 6.1.4.2). La primera es la hipótesis nula, en donde se asume que las tasas de sustitución que dan origen a los clados tienen el mismo valor de  $w$ , y la contrastamos con la hipótesis alterna, la cual supone que las presiones de selección que actuaron en las diferentes ramas fueron diferentes. Esto se traduce en que las ramas que dan origen a las distintas subfamilias de PriA estarían bajo diferentes presiones de selección. Los



resultados fueron complementados con un análisis de máxima verosimilitud (Tabla 15), y se aceptó la hipótesis alterna relacionada con que diferentes presiones de selección actuaron sobre la evolución de los ortólogos de PriA.

Hipótesis	Omega w	np	InL
<b>H0=</b> <b>wPriA=wSubTrpF=wSubHisA2</b>	w=0.09459	77	-15172.40253
<b>H1=</b> <b>wPriA≠wSubTrpF≠wSubHisA2</b>	wPriA= 0.06066 wSubTrpF= 0.14540 wSubHisA2=0.09277	81	-15155.85071

Tabla 15. Presiones de selección para las diferentes subfamilias de PriA

Por tal motivo, se sugiere que estas presiones de selección se relacionan con la co-ocurrencia de los genes *his/trp*. En las hipótesis propuestas postulamos que los ortólogos se mantienen funcionales pero, que pueden subfuncionalizarse o especializarse debido a una relajación en la presión de selección. En la H1 por ejemplo, existe una selección purificadora con un valor de  $w$  igual a 0.06066, que elimina las mutaciones que afectan la actividad, por lo que se espera que éstos ortólogos mantengan su bifuncionalidad. Para las hipótesis H2 y H3 se observó que las presiones de selección disminuyeron, siendo los valores de  $\omega$  más grandes si se comparan con los resultados obtenidos para las enzimas PriA bifuncionales, 0.14540 y 0.09277 para subTrpF y subHisA2, respectivamente. Estos resultados juntos muestran que una relajación en la presión de selección puede permitir la sub-funcionalización o especialización enzimática, lo que se relaciona con la co-ocurrencia en las pérdidas de genes *his/trp*. Hasta este punto tenemos hipótesis con asociación funcional que son dignas de ser caracterizadas a nivel bioquímico y estructural, como se describe en el siguiente capítulo.

## CAPÍTULO 3

### 7.9 CARACTERIZACIÓN FUNCIONAL DE ENZIMAS RETENIDAS: NUEVAS SUBFAMILIAS EN LA BIOSÍNTESIS DE L-HISTIDINA Y L-TRIPTÓFANO EN *ACTINOMYCETACEAE*

En este capítulo abordaremos las implicaciones funcionales y estructurales de la pérdida de los genes *his/trp* sobre los ortólogos de PriA que se han retenido en los diferentes linajes evolutivos, dando lugar a dos subfamilia de enzimas de PriA previamente desconocidas. Para este fin, se reportan experimentos mediante los cuales se caracterizan in vivo e in vitro la actividad bioquímica y las características estructurales de enzimas selectas.

#### 7.9.1 CARACTERIZACIÓN IN VIVO DE ORTÓLOGOS DE PRIA PROVENIENTES DE LOS TRES ESCENARIOS EN CEPAS DE *E. COLI* HISA- Y TRPF-

Basados en los resultados del capítulo, decidimos caracterizar a nivel funcional ortólogos selectos provenientes de los tres escenarios propuestos: tres provenientes de *Bifidobacterium* (outgroup), Org1, Org3, Org6. Cuatro posibles enzimas bi-funcionales provenientes del Linaje II, Org16, Org23, Org22. Dos provenientes de la hipótesis H2, en donde postulamos que estas enzimas presentan únicamente actividad de ProFAR isomerasa, tales como Org35 y Org37. Y cuatro provenientes de la hipótesis H3, la que postula únicamente la ocurrencia de la actividad PRA isomerasa, provenientes de los organismos Org11, Org14, Org40 y Org42.

Como un primer acercamiento se decidió determinar las actividades biológicas mediante ensayos de complementación *in vivo*, utilizando cepas de *E. coli* mutadas en los genes *hisA* (HfrG6) y *trpF* (FBG-Wf)(Wright *et al.* 2008), las cuales fueron transformadas con el plásmido pASK (Noda-García *et al.* 2015). Este vector incluye el promotor inducible por tetraciclina (tetO), el cual está altamente controlado y no responde a ningún otro mecanismo genético, por lo que cambios en el

fondo genético no afectan su expresión (Noda-García *et al.* 2015). Los ensayos de complementación se llevaron a cabo utilizando medio mínimo M9, el cual fue suplementado con todos los aminoácidos menos histidina o triptófano. La expresión del promotor se llevó a cabo utilizando 100 ng/ml de anhidrotetraciclina, lo que fue suficiente para observar los resultados. Como control negativo se utilizó el vector pASK vacío. Las enzimas fueron capaces de complementar el crecimiento en aproximadamente 48 horas. Los resultados se muestran en la Tabla 16.

<b>Enzima<sup>§</sup></b>	<b>Linaje</b>	<b>Grupo</b>	<b>Actividad ProFAR</b>	<b>Actividad PRA</b>
<b>pASK-vacío</b>	-	control	-	-
<b>PriA_Org3_B. longum</b>	-	outgroup	+	+
<b>PriA_Org1_B. gallicum</b>	-	outgroup	+	+
<b>PriA_Org6_B. adolescentis</b>	-	outgroup	+	+
<b>PriA_Org16_A. urogenotalis*</b>	I	H1	+	+
<b>PriA_Org23_A. sp. oral taxon 171</b>	I	H1	+	+
<b>PriA_Org22_A. oris MG-1</b>	I	H1	+	+
<b>SubHisA2_Org35_A.vaccimaxiliae</b>	IV	H2	+	+
<b>SubHisA2_Org37_A. cardiffensis</b>	IV	H2	+	+
<b>SubTrpF_Org11_A. sp. oral taxon 848</b>	IV	H3	-	+
<b>SubTrpF_Org14_A. graevenitzii</b>	II	H3	-	+
<b>SubTrpF_Org40_A. sp. oral taxon 180</b>	I	H3	-	+
<b>SubTrpF_Org42_A. odontolyticus*</b>	IV	H3	-	+

**Tabla 16.** Ensayos de complementación de ortólogos de PriA

<sup>§</sup>Las enzimas están acomodadas conforme a su ubicación taxonómica

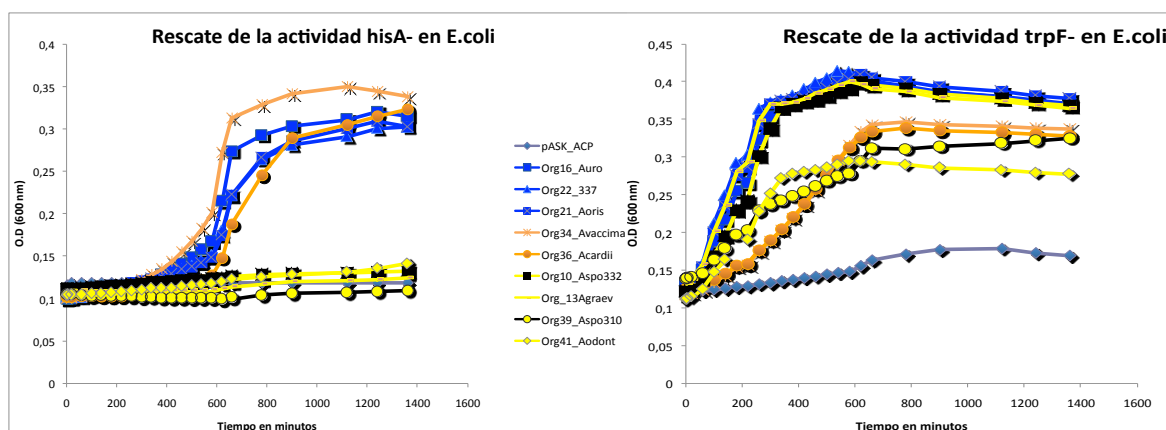
\*Enzimas previamente reportadas en la tesis de maestría (Juárez-Vázquez 2011).

Estos resultados nos ofrecen un panorama general de las actividades de estas enzimas, demostrándose que los homólogos asociados a la hipótesis H1 (Figura 28, cuadro morado y figura 29), son capaces de rescatar ambas actividades de manera eficiente; mientras que las enzimas de los organismos correspondientes a la hipótesis H2 (Figura 28, cuadro naranja y figura 29), complementan bien la actividad ProFAR isomerasa (*hisA*) pero parecen tener muy poca actividad PRA isomerasa (*trpF*). En cambio, los homólogos asociados a la hipótesis H3 (Figura 28, cuadro amarillo y figura 29), complementan únicamente la actividad codificada por *trpF*, es decir, PRA isomerasa. Todos estos resultados concuerdan con las hipótesis planteadas. En el caso de los homólogos provenientes de la H2, es altamente probable que estos estén en un proceso de estasis evolutiva, en donde al acumular mutaciones irán perdiendo la actividad de PRA isomerasa.

Cómo se mencionó en las subsecciones anteriores, observamos que las subfamilias SubHisA2 y SubTrpF presentan mutaciones, que en algunos casos las hacen insolubles *in vitro*, lo que nos hace pensar en que este tipo de proteínas podrían tener una desventaja adaptativa para los organismos que las han retenido. Por lo que decidimos repetir los experimentos de complementación de ésta sección en medio líquido y establecer si existía una relación fisiológica de las bajas actividades con la tasa de crecimiento de las cepas de *E. coli* mutantes en los genes *hisA*- y *trpF*-. Estos ensayos de complementación se llevaron a cabo tal cual se describe en la sección de materiales y métodos y los resultados de las cinéticas de crecimiento complementadas con los genes ortólogos de PriA provenientes de los diferentes escenarios que se reportan provienen de tres experimentos diferentes.

Observamos que bajo el fondo genético *hisA*-, solo los genes provenientes de los organismos Org16, Org21, Org22, Org34 y Org36 (Figura 35, lado izquierdo, etiquetas azules y naranjas) rescataron la actividad de ProFAR isomerasa (*hisA*). Por el otro lado en el fondo genético *trpF*-, todos los genes fueron capaces de complementar la actividad PRA isomerasa, mientras que las enzimas provenientes de los organismos Org34 y Org36 fueron las que tardaron más en

complementar dicha actividad, lo que concuerda con nuestros resultados sobre que estas enzimas están en el proceso de perder por completo la actividad de PRA isomerasa.



**Figura 30.** Cinéticas de crecimiento de las cepas de *E.coli hisA-* y *trpF-*. Cada punto representa una lectura tomada a 600 nm

Dado que en el fondo genético de *E. coli* se encuentran los genes que codifican para las actividades de PRA y ProFAR isomerasa por separado, genes *hisA* y *trpF*, se procedió a llevar a cabo experimentos análogos de complementación en el fondo genético de la mutante en *priA* de *A. oris* MG-1, generada en el capítulo anterior.

## 7.9. 2 CARACTERIZACIÓN IN VIVO DE LOS HOMÓLOGOS DE PRIA EN A. ORIS MG-1

Los resultados descritos en la sección anterior presentan evidencia sólida de las actividades de estos homólogos en cepas de *E. coli*. Sin embargo, éste no es el fondo genético en donde han evolucionado estas enzimas. Por ejemplo, en Actinobacterias se pierde el la secuencia promotora

hisP1, mientras que en *E. coli* la regulación se da mediante dicho promotor (Alifano *et al.* 1996). Debido a esto, consideramos realizar los ensayos en un fondo genético carente de *priA*, mediante el uso de la mutante correspondiente de *A. oris MG-1* (ver sección 7.6.7). Para ello utilizamos el plásmido pJRD215, el cual en realidad es un cósmido que se transfiere por conjugación, y se ha utilizado para complementar actividades en especies del género *Actinomyces* (Davison *et al.* 1987; Yeung y Kozalsky 1994). Los doce genes (Tabla 16) fueron clonados utilizando los sitios de corte *NdeI* y *HindIII*. La cepa mutante fue transformada con las construcciones, y se puso a crecer en un medio rico con infusión de cerebro y corazón (BHI por sus siglas en inglés, Brain Heart Infusion).

Las cepas con el cósmido integrado fueron identificadas debido a que presentaron resistencia a kanamicina y estreptomicina, conferidas por el cósmido. Sin embargo, cuando intentamos crecer estas transformantes en un medio mínimo definido, provisto por nuestros colaboradores de la Universidad e Texas, (Ton-That y Wu, datos sin publicar) carente de histidina y triptófano, y con los respectivos antibióticos, no se observó crecimiento alguno en ningún caso. Dado que en la sección 7.5 comprobamos la integridad genética de la mutante, sugerimos que la replicación del cósmido con los diferentes homólogos es demasiado costosa a nivel metabólico, y afecta el crecimiento de la bacteria. Un medio rico en aminoácidos como el BHI tampoco fue suficiente para promover el crecimiento de la cepa. Por otra parte, para poder determinar diferencias a nivel de las actividades específicas de cada enzima decidimos llevar a cabo ensayos enzimáticos sobre los ortólogos selectos. En los casos en donde fue posible, se determinaron los parámetros cinéticos  $k_{cat}$  y  $K_M$  para ambos sustratos, ProFAR y PRA, los cuales se describirán en las secciones subsiguientes.

### 7.9.3 CARACTERIZACIÓN *IN VITRO* DE LOS HOMÓLOGOS DE PRIA

De manera inicial se llevaron a cabo los ensayos de sobre-expresión de los ortólogos de PriA utilizando los vectores pET15b, pET22a, pET28b y pMCSG68 que contienen el promotor T7, así como una cola de histidinas (6x) ya sea en el extremo amino ó carboxilo-terminal, según sea el caso. La cepa de *E. coli* BL21\* star (DE3) y BL21 Magic fueron transformadas con estos vectores y la expresión se llevó a cabo utilizando hasta 1mM de IPTG. La solubilidad de todas las enzimas analizadas fue verificada por electroforesis en geles de poliacrilamida (Tabla 17).

	Plásmido/cepa	Uso de codones	Expresión	Solubilidad
<b>PriA_Org3_ <i>B. longum</i></b>	pET28a/BL21	Silvestre	+	+
<b>PriA_Org1_ <i>B. gallicum</i></b>	pET22b/BL21	Silvestre	+	+
<b>PriA_Org6_ <i>B. adolescentis</i></b>	pET22b/BL21	Silvestre	+	+
<b>PriA_Org16_ <i>A. urogenotalis</i>*</b>	pET28a/BL21	Optimizado	+	+
<b>PriA_Org23_ <i>A. sp. oral taxon 171</i></b>	pMCSG68/BL21	Optimizado	+	+
<b>PriA_Org22_ <i>A. oris</i> MG-1</b>	pET22b/BL21	Silvestre	+	+
<b>SubHisA2_Org35_ <i>A. vaccimaxiliae</i></b>	pET22b/BL21	Optimizado	-	-
<b>SubHisA2_Org37_ <i>A. cardiffensis</i></b>	pET22b/BL21	Optimizado	+	+
<b>SubTrpF_Org11_ <i>A. sp. oral taxon 848</i></b>	pMCSG68/BL21 Magic cells	Optimizado	+	+
<b>SubTrpF_Org14_ <i>A. graevenitzii</i></b>	pET22b/BL21	Optimizado	-	-



<b>SubTrpF_Org40_A. sp. oral taxon 180</b>	pET22b/BL21	Optimizado	-	-
<b>SubTrpF_Org42_A. odontolyticus*</b>	pET15b/BL21	Optimizado	+	+

**Tabla 17. Condiciones de expresión de homólogos selectos de PriA**

La secuencia de codones optimizado se indica en materiales y métodos. La solubilidad de la proteína usando vectores pET, fue observada en cepas DE3, mientras que si se trató del vector pMCSG68, fue en BL21 Magic. La solubilidad se observó utilizando el amortiguador Tris-HCl pH 7.5 50 mM, Glicerol 10% v/v y NaCl 50 mM. \*Reportadas previamente en tesis de maestría (Juárez-Vázquez 2011).

Las concentraciones de enzima óptimas para llevar a cabo los ensayos enzimáticos varían entre 20 y 100 nM. Los datos obtenidos se muestran en la Tabla 18. Estos resultados al igual que los ensayos de complementación y los de selección indican que existe una relación entre la actividad y el escenario del cual provienen. Los detalles de esta relación y sus implicaciones evolutivas son analizadas a continuación.

	Actividad enzimática <sup>a</sup>					
	ProFAR isomerasa (HisA)			PRA isomerasa (TrpF)		
	$K_M$ $\mu M$	$k_{cat}$ $s^{-1}$	$k_{cat}/K_M$ $s^{-1} \mu M^{-1}$	$K_M$ $\mu M$	$k_{cat}$ $s^{-1}$	$k_{cat}/K_M$ $s^{-1} \mu M^{-1}$
PriA_Org3_B. longum	2.7 ± 0.5	0.4 ± 0.1	0.1	6.1 ± 0.1	2.1 ± 0.5	0.3
PriA_Org1_B. gallicum	1.7 ± 0.3	0.3 ± 0.1	0.2	40 ± 9	3.5 ± 0.1	0.09
PriA_Org6_B. adolescentis	17 ± 4.3	2.3 ± 0.01	0.1	21 ± 5	0.9 ± 0.2	0.04
PriA_Org16_A. urogenitalis	4.0 ± 0.9	0.2 ± 0.03	0.04	23 ± 6.5	0.5 ± 0.05	0.02
PriA_Org23_A. sp. oral taxon 171	3 ± 0.3	0.3 ± 0.09	0.1	8 ± 2	0.4 ± 0.1	0.04
PriA_Org22_A. oris MG-1	10 ± 2	0.2 ± 0.09	0.02	30 ± 7	0.3 ± 0.03	0.01
SubHisA2_Org35_A. vaccimaxillae						
SubHisA2_Org37_A. cardiffensis	56 ± 17	0.14 ± 0.05	0.002	n.d.	n.d.	n.d.

SubTrpF_Org11_A. sp. oral taxon 848	n.d.	n.d.	n.d.	n.d.	n.d.	0.0001
SubTrpF_Org14_A. graevenitzii						
SubTrpF_Org40_A. sp. oral taxon 180						
SubTrpF_Org42_A. odontolyticus	n.d.	n.d.	n.d.	85 ±0.9	0.15 ± 0.06	0.02

Tabla 18. Caracterización enzimática de los ortólogos de PriA

<sup>a</sup> Cada punto proviene de al menos tres experimentos, utilizando purificaciones de enzimas independientes. n.d., actividad no detectada, utilizando condiciones saturantes de enzima. Los recuadros vacíos representan nuestra incapacidad de poder obtener la proteína soluble para llevar a cabo los experimentos. Los límites de detección para las actividades de PRA y ProFAR isomerasas usadas en el presente estudio son  $0.0001 \mu\text{M}^{-1}\text{s}^{-1}$  y  $0.001 \mu\text{M}^{-1}\text{s}^{-1}$ , respectivamente.

#### 7.9.4 EVOLUCIÓN DE LA FUNCIÓN ENZIMÁTICA DE LA SUB-FAMILIA PRIA (HIPÓTESIS 1)

Las enzimas PriA provenientes de los organismos del Linaje II (Org16, 22 y 23) son bi-funcionales, Sin embargo, si se comparan los valores de eficiencia catalítica de los organismos que conforman el grupo externo del género *Bifidobacterium* (Org1, 3 y 6), con los de estos organismos, sus valores son un orden de magnitud más bajo:  $k_{cat}/K_M^{\text{ProFAR}}$  de 0.01 a  $0.1 \mu\text{M}^{-1}\text{s}^{-1}$  y  $k_{cat}/K_M^{\text{PRA}}$  alrededor de  $0.01 \mu\text{M}^{-1}\text{s}^{-1}$  (Tabla 18). Esto podría ser el resultado de la acumulación de mutaciones debido al proceso por el cual están pasando estos organismos, es decir, el deterioro genómico podría afectar la actividad. Alternativamente, dado que estas enzimas provienen de organismos que se separan temprano del ancestro común (Capítulo 1), sus bajas actividades podrían ser el resultado de la compensación que ocurrió durante el proceso de pasar de ser una enzima HisA a una tipo PriA (Noda García *et al.* 2015; Planch *et al.* 2016).

Así mismo, detectamos una afinidad  $K_M^{\text{PRA}}$  más alta en comparación con la  $K_M^{\text{ProFAR}}$ , lo que también se observa en las enzimas provenientes de *Bifidobacterium*. Este hecho soporta que a pesar de su bi-funcionalidad, estas enzimas conservan mayor afinidad por su sustrato ancestral, ProFAR. Interesantemente, a pesar de que estos ortólogos de PriA provienen de especies del mismo Linaje evolutivo (Capítulo 2), poseen en promedio menos del 70% de identidad entre ellos. Esto es importante porque como ya se demostró, aunque estén sujetos a una “fuerte” selección purificadora, la variación en estas secuencias es notable. Esto significa que la selección purificante actúa manteniendo los residuos que llevan a cabo la catálisis, y algunos otros que son importantes para la conformación del sitio activo, pero al mismo tiempo permite la acumulación de mutaciones “neutrales”, que en un principio no habrían de afectar la adecuación del organismo pero sí la actividad de la enzima.

Esto último es congruente con el hecho de que estos organismos son oportunistas de humanos y tienen a su disposición fuentes de histidina y triptófano en el medio. Dado que mantienen los genes de estas rutas completas (sección 7.6.8), su contribución para el mantenimiento fisiológico de este tipo de organismos es esencial, y por eso se mantienen como enzimas bi-funcionales. Esto podría estar relacionado a un mecanismo de respuesta a niveles bajos de histidina y triptófano disponibles en el medio, ya que como se mencionó, la biosíntesis de histidina está regulada por las concentraciones intracelulares de histidina.

### 7.9.5 EVOLUCIÓN DE LA FUNCIÓN ENZIMÁTICA DE LA SUB-FAMILIA SUBHISA2 (HIPÓTESIS H2)

Los ortólogos de PriA asociados a la hipótesis H2, provienen de organismos de los Linajes I, III y IV, los cuales incluyen organismos que pierden los genes *trp*. La única enzima que logramos obtener como soluble para su caracterización, fue la SubHisA2\_Org37. A pesar de que esta enzima muestra actividad PRA isomerasa *in vivo*, aun en condiciones saturantes, no detectamos actividad

de PRA isomerasa *in vitro*. Interesantemente, esta enzima tiene muy reducida la actividad de ProFAR isomerasa,  $k_{cat}/K_M^{\text{ProFAR}}$  0.002 unidades, un orden de magnitud por debajo de las enzimas bi-funcionales. Adicionalmente, presenta una afinidad por ProFAR demasiado alta,  $K_M^{\text{ProFAR}}$  de 56  $\mu\text{M}$  (Tabla 21). Esto podría explicarse por una leve relajación en la presión de selección purificante, como lo mostramos en el capítulo 2. La acumulación de mutaciones, si bien mantiene la actividad, afecta la afinidad y por ende su eficiencia catalítica. Esto podría ser un efecto que pudiera causar la pérdida de la actividad PRA isomerasa.

Otra interpretación de la pérdida de función podría ser que la biosíntesis de histidina no es esencial para algunos organismos de este tipo. Como lo demostramos en la sección 7.6.5, la cepa *A. oris* MG-1 mutante pudo crecer en un medio sin histidina, sugiriendo que esta ruta podría estar evolucionando de ser esencial a ser dispensable dadas ciertas condiciones. Por el contrario, la baja actividad también podría reflejar una regulación en la ruta en la cual participa esta enzima, es decir, se está adaptando a las eficiencias que presentan las enzimas río arriba y río abajo en la ruta. Resulta por ende interesante el estudiar a fondo a las enzimas involucradas en la ruta de biosíntesis de histidina en éstos organismos y ver si éstas también tienen una  $k_{cat}/K_M$  baja para la actividad que catalizan.

### 7.9.6 EVOLUCIÓN DE LA FUNCIÓN ENZIMÁTICA DE LA SUBFAMILIA SUBTRPF (HIPÓTESIS H3)

Los ortólogos de PriA asociados a la hipótesis H3, provienen de organismos de los Linajes I, III y IV, los cuales incluyen organismos que pierden los genes *his*. Estos ortólogos de PriA fueron capaces de rescatar la actividad de TrpF en *E. coli*. Se logró obtener proteínas solubles para llevar a cabo ensayos de cinética enzimática de dos enzimas representativas, Org42 y Org11, con  $k_{cat}/K_M^{\text{PRA}}$  de 0.01 y 0.0001  $\mu\text{M}^{-1}\text{s}^{-1}$ , respectivamente. En el caso de la enzima proveniente del Org11, detectamos que ésta tiene una inserción de 50 aminoácidos extra, la que podría desestabilizar a la proteína e interferir con sus actividades catalíticas. Consideramos que esta proteína es el resultado

de la acumulación de mutaciones, resultado de una relajación en la presión de selección como se describió anteriormente.

En resumen con la caracterización *in vitro* logramos detectar diferencias en las afinidades por sustrato de los ortólogos de PriA provenientes de las tres hipótesis propuestas. Interesantemente, estos escenarios se relacionan con la co-ocurrencia de los genes *his/trp* y las presiones de selección. En este punto, nos pareció relevante tratar de establecer la relación estructura-función con el surgimiento de las nuevas subfamilias, es decir, qué tipo de cambios a nivel estructural afectan los valores de  $k_{cat}/K_M$  observados.

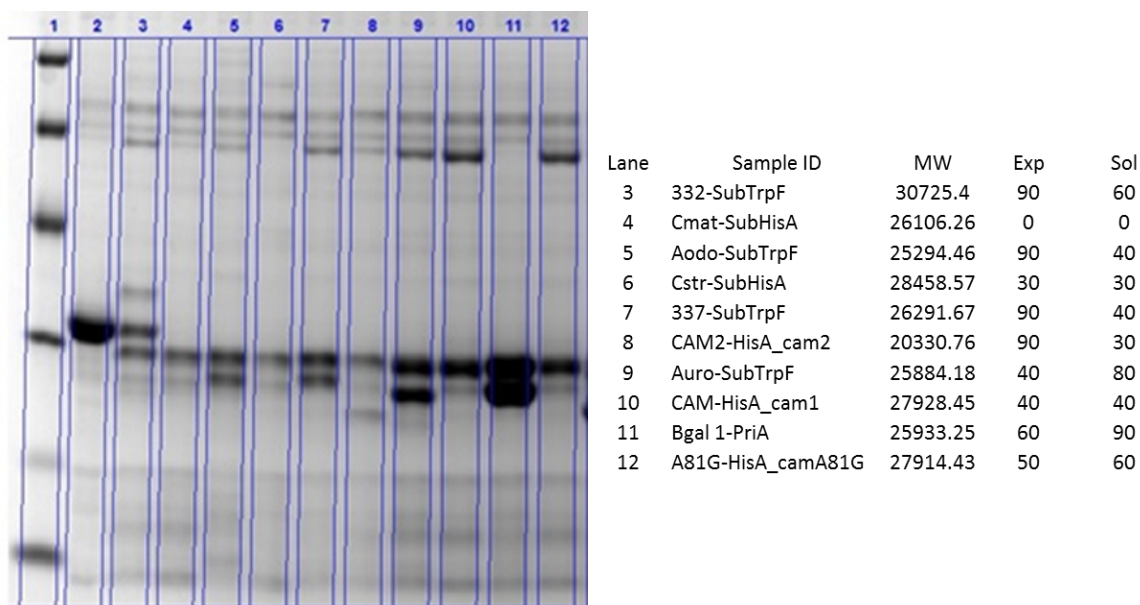
## 7.10 ANÁLISIS ESTRUCTURAL DE LA EVOLUCIÓN DE LA ESPECIFICIDAD POR SUSTRATO EN LOS ORTÓLOGOS DE PRIA.

La caracterización de la relación estructura función de enzimas tipo PriA y SubHisA ya ha sido reportada (Due *et al.* 2011; Noda-García *et al.* 2013). En esta subsección nos enfocaremos en tratar de entender los mecanismos generales a nivel estructural que permitieron el surgimiento de las Subfamilias SubHisA2 y SubTrpF en *Actinomycetaceae*.

### 7.10.1 DETERMINACIÓN DE ESTRUCTURAS CRISTALOGÁFICAS DE PRIAS DE ACTINOMYCES

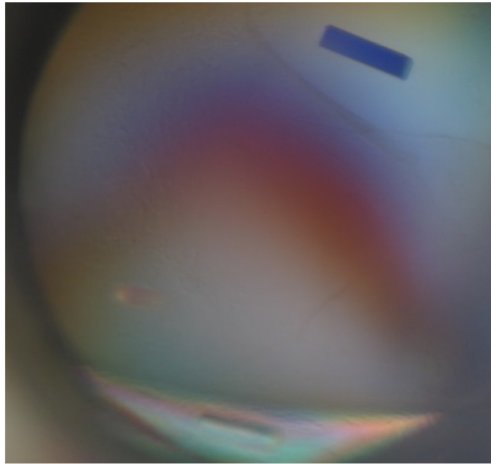
Para llevar a cabo los experimentos de cristalografía por rayos X, las proteínas que se obtuvieron solubles (Tabla 17), fueron llevadas a la unidad de Cristalografía de Rayos X en el Laboratorio Nacional del Departamento de Energía Argonne en Chicago. Las proteínas fueron purificadas por afinidad a Níquel y por cromatografía de exclusión molecular. Posteriormente, se les cortó la cola de histidinas. Las proteínas fueron concentradas hasta 37 mg/ml. Para verificar el corte en la cola de histidinas, la purificación, solubilidad y expresión de las proteínas se visualizaron en un gel SDS-page. Los resultados obtenidos se muestran en la Figura 31. Aquí se puede observar la pureza de las proteínas y su porcentaje de solubilidad. Con estas proteínas se llevaron a cabo los ensayos de

crystalización utilizando el método de gota sentada y el protocolo descrito en materiales y métodos. A pesar de que se ocuparon muchas condiciones diferentes para poder obtener cristales de las diferentes enzimas con los sustratos, sólo se pudieron obtener cristales de la proteína PriA\_Org16\_ *A. urogenitalis*.



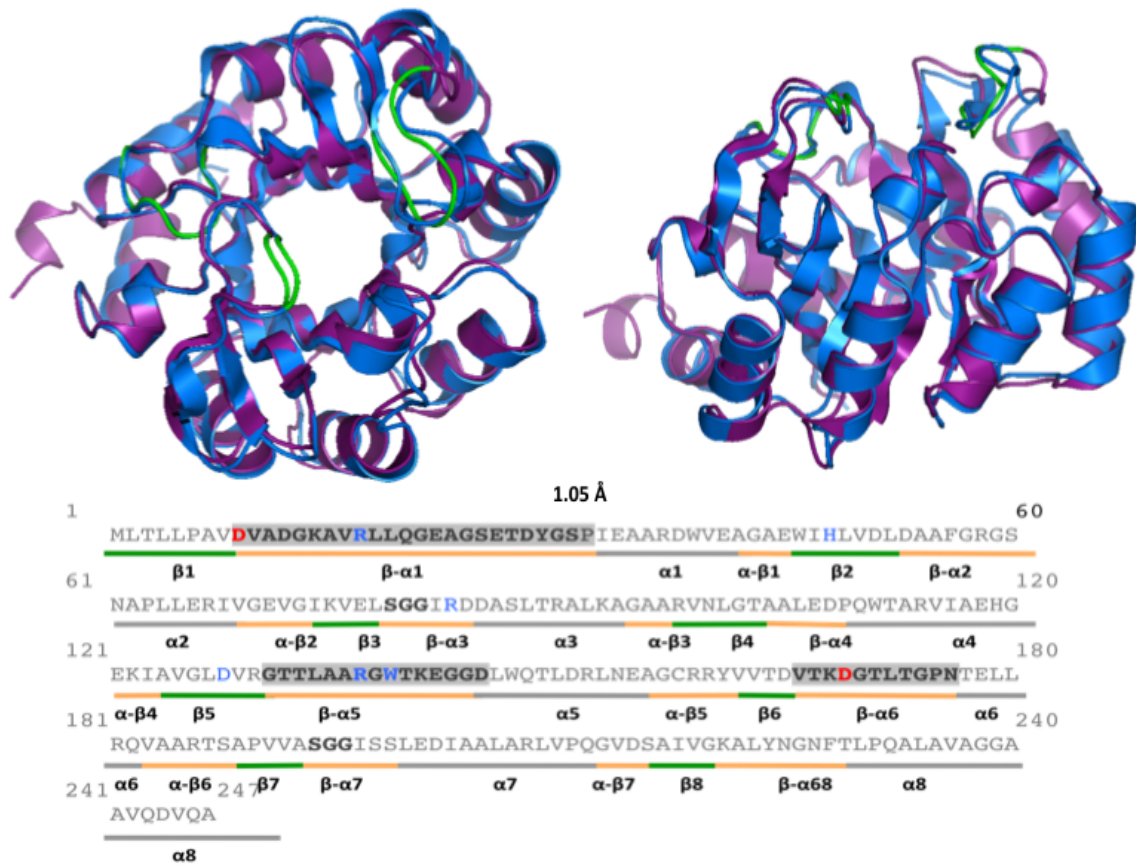
**Figura 31.** Purificación de los ortólogos de PriA, y corte de cola de histidinas. Al lado derecho se muestra la referencia de cada carril. Las líneas de interés corresponden a los carriles 3\_SubTrpF\_Org11, 5\_SubTrpF\_Org42, 7\_PriA\_Org23, 9\_PriA\_Org16.

Los cristales de la PriA\_Org16 aparecieron en condiciones de 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 0.1 M CAPS:NaOH pH 10.5, y 1.2 M NaH<sub>2</sub>PO<sub>4</sub>/0.8 M K<sub>2</sub>HPO<sub>4</sub>. Antes de ser colectados, los cristales fueron criopreservados en 2.4 M K<sub>2</sub>HPO<sub>4</sub>. (Figura 32). La difracción y afinamiento de datos fue llevada a cabo en el Centro de Biología Estructural en Argonne por Karolina Milkasha.



**Figura 32.** Cristales de PriA\_Org16\_A. *urogenitales*. Los cristales fueron obtenidos en la condición 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 0.1 M CAPS:NaOH pH 10.5, y 1.2 M NaH<sub>2</sub>PO<sub>4</sub>/0.8 M K<sub>2</sub>HPO<sub>4</sub>.

Una vez que los datos fueron colectados, se utilizó a la PriA de *M. tuberculosis* (PDB:1VZW) (Due *et al.* 2011) para hacer reemplazo molecular. Los datos con el afinamiento estructural fueron depositados en la basa de datos de proteínas con el número de acceso PDB:4X2R. La resolución atómica se obtuvo a 1.05 Å, la proteína presenta un plegamiento tipo  $\beta\alpha_8$  como se ha observado en otras proteínas. La estructura tridimensional esta compuesta por un solo monómero al igual que otras PriAs. Interesantemente no se logró resolver la estructura del loop  $\beta\alpha_1$  y 5, el cual se ha postulado que es altamente flexible (Figura 33). Dado que este primer acercamiento no nos permitió identificar residuos asociados con una disminución en la eficiencia catalítica de ésta PriA. Decidimos llevar a cabo un análisis de secuencia con otras PriAs previamente reportadas.



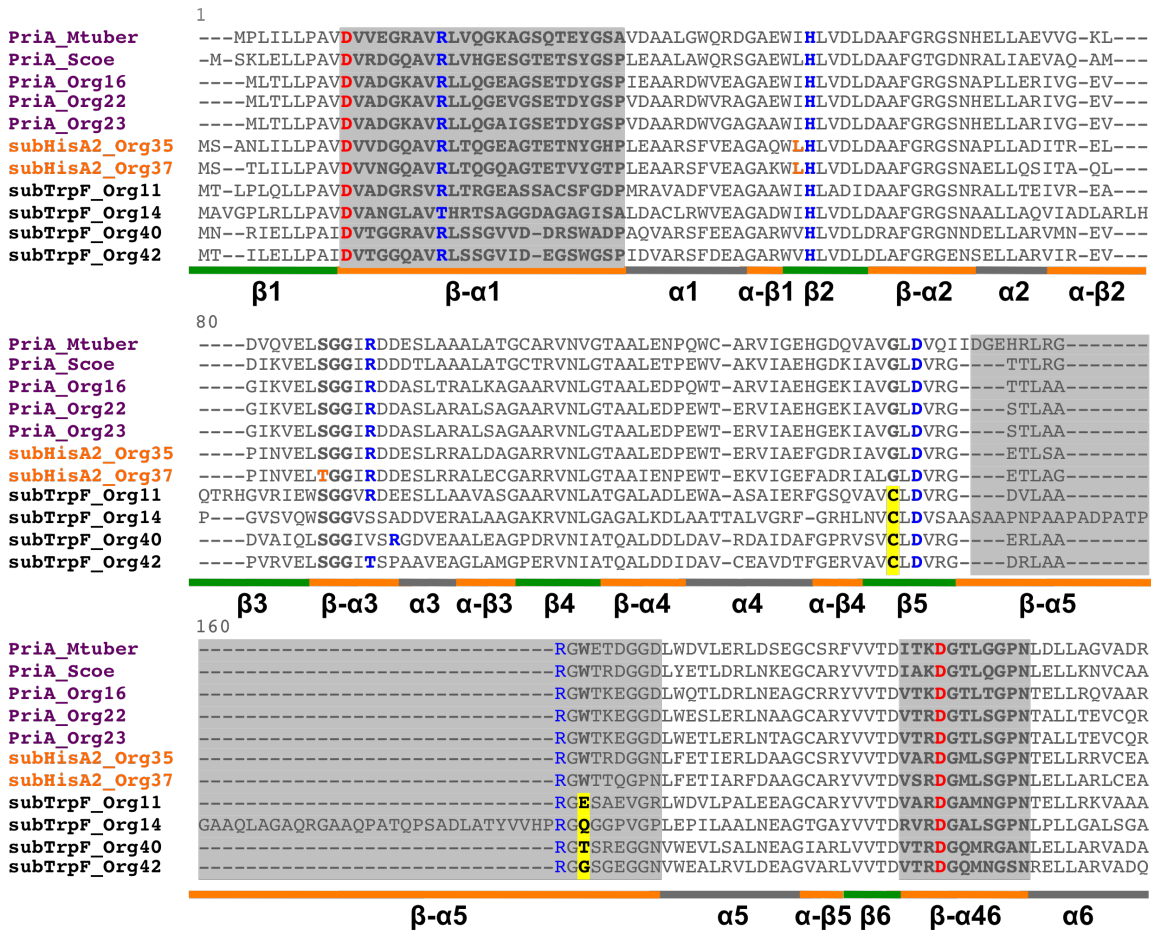
**Figura 33.** Comparación estructural de PriA\_*A. urogenitales* y la PriA\_*M. tuberculosis*. Vista frontal y lateral del monómero 4x2R (estructura morada). Se comparó con la estructura de PriA de *M. tuberculosis* PDB:1VZW (estructura azul). Los cambios estructurales más relevantes se resaltan en verde. En la parte de abajo se muestra la secuencia de la proteína, los residuos catalíticos se resaltan en rojo, residuos involucrados en la catálisis se resaltan en azul y en gris se marcan las asas.

Basados en estos resultados, decidimos hacer modelos estructurales por homología de una proteína proveniente de la hipótesis H2 y H3. Específicamente modelos las proteínas SubHisA2 Org37\_*A. cardiffensis* y SubTrpF Org42\_*A. odontolyticus* que fueron proteínas de las cuales obtuvimos parámetros cinéticos para después compararlos con la estructura de PriA:4X2R y determinar diferencias relacionadas con la especialización de las subfamilias.



### 7.10.2 DIFERENCIAS A NIVEL DE SECUENCIA Y ESTRUCTURA ENTRE PRIA Y SUBHISA2

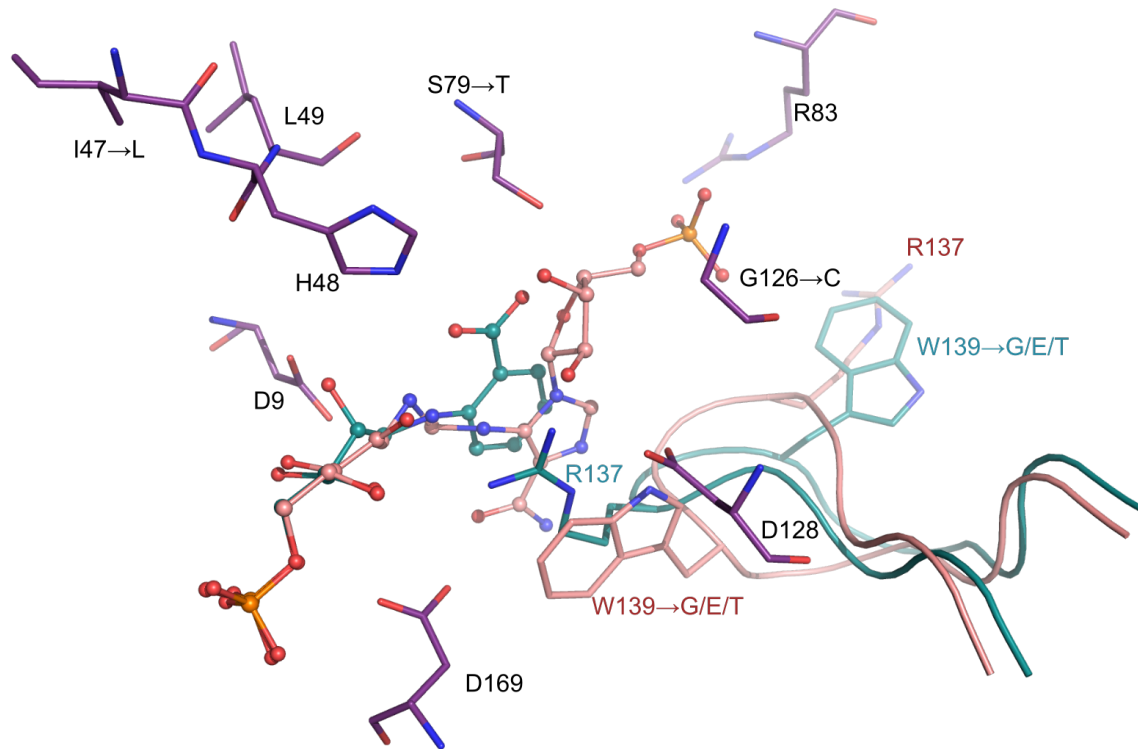
Para llevar a cabo el análisis de secuencia se alinearon todas las 40 secuencias de PriA reportadas en este estudio junto con las secuencias de PriAs previamente reportadas. Sin embargo, dado que nuestro interés recae en tener datos sólidos decidimos delimitar nuestro análisis a solo las secuencias caracterizadas (Figura 34, secuencias resaltadas en naranja). Detectamos cambios en residuos conservados de una PriA (H1) a una SubHisA2 (H2), como el cambio en la Ile47Leu y una Ser79Thr. Estudios previos de la mutación en estos dos residuos, matan totalmente la actividad de PRA isomerasa en las enzimas tipo SubHisA de *Corynebacterium* (Noda-García et al. 2013). Parecido a lo que se observa en SubHisAs previamente reportadas, la subHisA2\_Org37 también presenta la mutación Ser79 a Thr79 (Figura 34). En esta mutación el grupo metil de la treonina podría afectar el contacto entre PRA y el grupo hidroxilo común en estos residuos, por eso se ve afectada la actividad de PRA isomerasa como se muestra en la superposición de la PriA\_ *M. tuberculosis* con el modelo de la subHisA2\_Org37 (Figura 35).



**Figura 34.** Alinamiento de secuencias provenientes de la familia PriA y subfamilias SubHisA2 y SubTrpF. En morado se marcan las secuencias de la familia PriA, en naranja las secuencias de la subfamilia SubHisA y en negro las secuencias provenientes de la subfamilia SubTrpF. En rojo se resaltan los residuos catalíticos y en azul residuos conformacionales. En gris se resaltan las asas. En amarillo y naranja se resaltan los cambios relevantes detectados en cada familia.

La ocurrencia de la familia SubHisA2 en *Actinomycetaceae* junto con la presencia de las SubHisA en *Corynebacterium*, demuestran que la sub-funcionalización de PriA en enzimas tipo SubHisA con actividad de ProFAR isomerasa, ha ocurrido al menos dos veces en Actinobacteria. Sin embargo, los escenarios que dan origen a estas dos subfamilias son diferentes. Por un lado se sabe que la subfamilia SubHisA se origina por selección positiva debido a la adquisición de un operón de Triptófano, lo que hace que la PriA en estos organismos se especialice hacia una actividad tipo SubHisA (Noda-García *et al.* 2013). Por otra parte, las enzimas tipo SubHisA2 evolucionaron por

relajación de la selección en un contexto en donde se pierden los genes *trp*. Interesantemente, parece ser que existe una selección de la mutación Ser80 a Thr80. Tal vez la evolución de éstas SubHisA2 pase por el cambio antes mencionado. Si éste es el caso, este es un ejemplo de convergencia evolutiva en bacterias originado por dinámicas genómicas distintas, o tal vez éste sea el camino más parsimonioso para pasar de ser una enzima tipo PriA a una tipo subHisA.



**Figura 35.** Análisis estructural entre PriAs, SubTrpFs y SubHisA2s. Se muestra una superposición de los PDBs:4x2R (morado), SubTrpF\_Org42 y SubHisA2\_Org37. La figura muestra la correcta posición de los residuos catalíticos durante la isomerización de PRA (verde claro) y la isomerización con ProFAR (rosa). Se muestran los cambios detectados en las secuencias SubTrpF y SubHisA2.

### 7.10.3 DIFERENCIAS A NIVEL DE SECUENCIA Y ESTRUCTURA ENTRE PRIA Y SUBTRPF

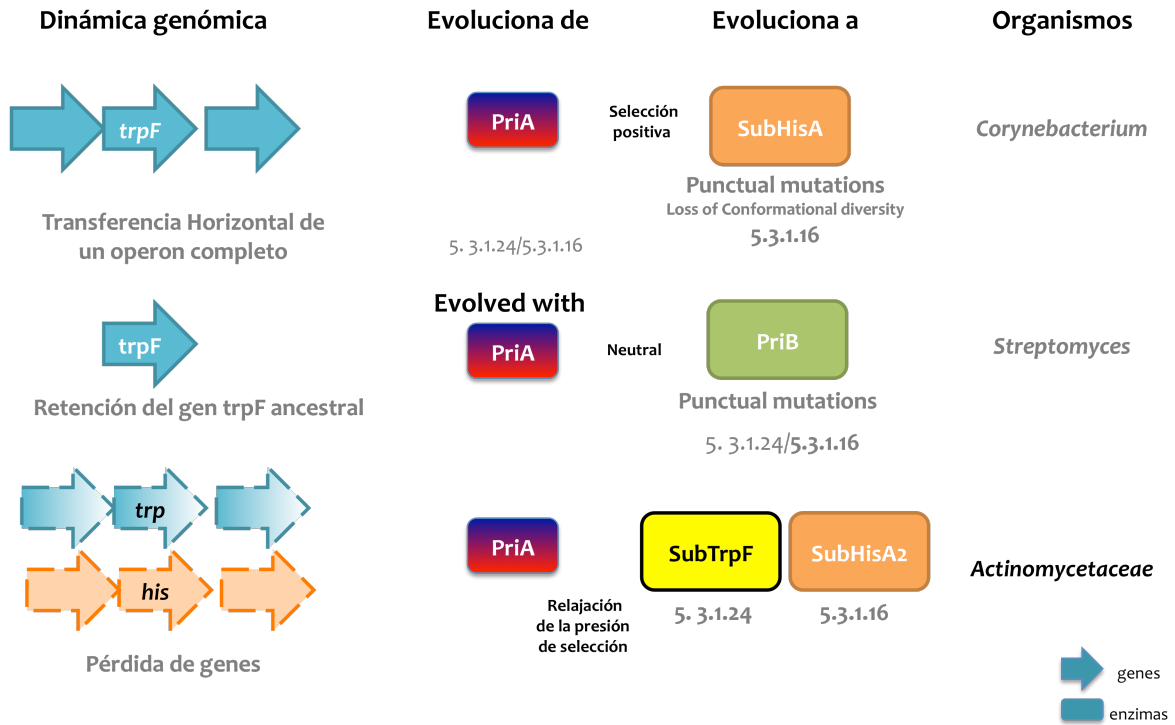
Basados en el alineamiento de la sección anterior (Figura 34), detectamos cambios relevantes entre las SubTrpFs (H3) y las PriAs (H1). Estos cambios explican la actividad de ProFAR isomerasa observada en la sección 7.9. Las mutaciones que observamos están en el cambio de la Gly126Cys y el Trp145Gly. En PriA la Gly126 esta cerca del residuo catalítico Asp128. La mutación hacia una cisteína en las SubTrpFs podría influir en el acomodo del la Arg137 con respecto al Asp128, obstruyendo el correcto acomodo de ProFAR (Figura 35). La unión de este sustrato se vería afectada porque este sustrato es largo tiene un motivo fósforo-azúcar que tal vez no interaccionaría en un una forma correcta para su catálisis. Así mismo observamos un cambio del Trp145 por una Gly. Este Trp 145 es importante para la correcta posición del sitio activo durante la interacción con ProFAR. El grupo Indol del Trp145 en PriAs puede formar un Puente de hisrógeno con el Ap128 estabilizando la conformación del sitio activo para la catálisis de ProFAR. La Arg83 esta perdida diferencialmente en algunas SubTrpFs y ocasionalmente se encuentra desplazada. La Arg83 interactúa con el segundo grupo fosfato de ProFAR, siguiendo a la correcta conformación del sitio activo y la catálisis de ProFAR (Figura 34). Muchos de estos cambios están conservados, por eso a pesar que observamos que la presión de selección que da origen a estas enzimas esta relajada. Es probable que exista una presión de selección positiva para el cambio en los residuos antes mencionados. Este evento de sub-funcionalización de PriA en una SubTrpF es único en Actinobacterias, y es es la primera vez que se describe. Interesantemente, en un estudio reciente, se observó que la secuencia subTrpF\_Org42 podría ser un intermediario evolutivo entre HisA y PriA (Noda García *et al.* 2015), lo que sustenta que estudios de éste tipo son capaces de detectar nuevas sub-familias de enzimas. En general podemos decir que los cambios que observamos dieron origen a la perdida de la actividad ancestral ProFAR fueron más drásticos si se comparan con los cambios sutiles para perder la actividad secundaria, es decir la actividad de PRA isomerasa. Por lo que

surgen algunas preguntas sobre la habilidad que tienen las enzimas de perder actividades ancestrales, ¿será posible que en las enzimas PriA sea más fácil perder la actividad de PRA isomerasa que la de ProFAR isomerasa?. ¿El andamiaje que presenta PriA favorece más las interacciones con ProFAR?

## 9. CONCLUSIONES

Este trabajo provee un ejemplo del uso de la filogenómica y la genómica comparada para detectar eventos de pérdidas y ganancias de genes dentro de Linajes selectos, éstos análisis de pérdidas de genes sirven para detectar enzimas retenidas y establecer asociaciones funcionales dependiendo del contexto en donde se mantengan. Encontramos que dentro del filo Actinobacteria existen familias que presentan genomas pequeños menos de 6 Mpb y presentan contenido-(G+C) menor al 70%, lo que sugiere que están pasando por un proceso de reducción del genoma. Algunos de éstos organismos pertenecen a la familia *Actinomycetaceae* en donde encontramos que sus estilos de vida los han llevado a evolucionar en cuatro Linajes evolutivos, con diferencias metabólicas. Por ejemplo el Linaje II, acumuló más mutaciones lo que les permitió formar el grupo más divergente de la familia, este linaje adquirió más funciones metabólicas, la mayoría están relacionadas con el uso de ácidos grasos exógenos como se ha observado en otras especies patógenas como *Streptococcus spp.* y *Staphylococcus spp.* Interesantemente éste Linaje conservó completos los genes involucrados en la biosíntesis de los 20 aminoácidos, mientras que los Linajes I, III y IV, pierden diferencialmente los genes para la biosíntesis de L-Histidina, L-Triptófano, L-Arginina, L-Fenilalanina, L-Tirosina y L-Prolina. Sin embargo, cuentan con transportadores para cada uno de estos aminoácidos. Aunque pareciera que una de las estrategias a las limitaciones en éstos aminoácidos podría ser a disminución en el número de los mismos en la secuencia de sus proteínas, no detectamos una co-relación entre la pérdida de las rutas de biosíntesis de aminoácidos en los Linajes I, III y IV y la presencia de éstos aminoácidos en la secuencia de proteínas. Interesantemente, todas las secuencias de proteínas en todos los Linajes propuestos tienen una proporción baja de Histinas, Cisteínas y Triptófanos, lo que podría explicarse porque la biosíntesis de estos aminoácidos es costosa, además algunos de éstos residuos como la Cys y el Trp tienen propiedades estabilizantes, lo que puede interferir con la catálisis de las enzimas.

Detectamos que en la cepa *A. oris* MG-1 *ΔpriA* la biosíntesis de L-Triptófano es esencial, mientras que es altamente probable que tengan un mecanismo aún no conocido para la obtención de histidina, por ejemplo aún queda la pregunta de si ¿éstos organismos han reclutado otra enzima capaz de llevar a cabo dicha actividad para compensar las actividades tan bajas que detectamos? cómo se ha detectado en otros patógenos como especies de *Chlamydia* (Adams *et al.* 2014). Éste ejemplo destaca la versatilidad que tienen los organismos en proceso de reducción del genoma para compensar la pérdida de genes, y la adaptación que presentan sus enzimas ante la pérdida de alguna ruta metabólica cómo el caso de PriA, la cual evolucionó en dos nuevas subfamilias SubHisA2 y SubTrpF. Este ejemplo de sub-funcionalización enzimática pone una vez más en evidencia los modelos actuales que explican la evolución de una función enzimática partiendo de un evento de duplicación (Nasvall *et al.* 2012), los cuales deberían de tomar en cuenta o proponer rutas alternas para aquellos organismos en donde la THG o la duplicación no son frecuentes. A su vez resalta la evolvabilidad de las enzimas provenientes de organismos bajo reducción del genoma para evolucionar hacia nuevas funciones. Esto último abre algunas preguntas sobre ¿el umbral que se ha propuesto (Tokuriki *et al.* 2010), para explicar el efecto de las mutaciones sobre la evolución de la función y la adecuación de los organismos, es el mismo para las proteínas que están bajo una tasa acelerada de acumulación de mutaciones?, ¿cuál será el efecto real de éstas mutaciones sobre la adecuación de los organismos? ¿las enzimas provenientes de organismos con deterioro genómico serán más plásticas? Finalmente, podemos concluir que existen diferentes escenarios por los cuales puede evolucionar una función, dichos escenarios son dependientes completamente del contexto en el cual participan éstas enzimas (Figura 36).



**Figura 36.** La evolución de la función enzimática está fuertemente relacionada con la dinámica genómica de los organismos. Diferentes presiones de selección actúan para modificar las actividades enzimáticas dando origen a nuevas sub-familias de proteínas.



## 10. PERSPECTIVAS

A pesar de que no tenemos evidencia sólida sobre otros episodios en dónde se pueda detectar la evolución de PriA en SubTrpF, una de las perspectivas debido a la cantidad de genomas disponibles es determinar si los genomas analizados en el capítulo 1 podrían estar bajo una dinámica de pérdida de los genes *his/trp* para determinar si la sub-funcionalización de PriA en SubHisA2 y SubTrpF son casos únicos en la familia *Actinomycetaceae* o podrían haber ocurrido en otras familias de Actinobacteria. Para determinar si la evolución de la especificidad por sustrato de PriA no es un caso único, proponemos buscar enzimas bi-funcionales que participen en dos o más rutas diferentes y detectar mediante análisis filogenómicos si están bajo una dinámica de pérdida de genes para observar si la pérdida puede dirigir la evolución de una función enzimática y determinar si éstas enzimas están bajo una presión de selección relajada. Para detectar la enzima que podría haber sido reclutada para llevar a cabo la actividad de HisA se pretende hacer una librería con todo el genoma de *A. oris* MG-1, utilizando como control la cepa *A. oris* MG-1  $\Delta priA$ . Otra estrategia sería llevar a cabo análisis de transcriptoma de la cepa *A. oris* MG-1  $\Delta priA$  en condiciones de medio mínimo carente de histidina y ver que genes se inducen. Entre ellos buscar un posible candidato para rescatar la actividad de ProFAR isomerasa.

Por último bastaría con probar si las familias de proteínas únicas al Linaje II pueden ser utilizadas como marcadores moleculares para el diagnóstico de infecciones causadas por estos organismos.

# 11. RESUMEN DE PUBLICACIONES RELACIONADAS Y CONTRIBUCIONES

## **Insights into the evolution of enzyme substrate promiscuity after the discovery of ( $\beta\alpha$ )<sub>8</sub> isomerase evolutionary intermediates from a diverse metagenome**

Lianet Noda-García\*, Ana L. Juárez-Vázquez\*, María C. Ávila-Arcos, Ernesto A. Verduzco-Castro, Gabriela Montero-Morán, Paul Gaytán, Mauricio Carrillo-Tripp and Francisco Barona-Gómez.

BMC Evolutionary Biology. 2015.

\* Primera autoría compartida.

En este análisis utilizamos Modelos Ocultos de Markov (HMM, por sus siglas en inglés de Hidden Markov Models) para clasificar familias de proteínas altamente relacionadas pero, con diferentes especificidades por sustrato. Se construyó un perfil utilizando estos HMM y con ello minando una base de datos de secuencias provenientes de metagenomas marinos logramos encontrar intermediarios evolutivos, es decir enzimas tipo HisA promiscuas para la actividad de PRA isomerasa (TrpF). Mis contribuciones en éste proyecto fueron reclasificar las secuencias obtenidas en éstos perfiles, para ello me base en los valores de corte propuestos para una enzima HisA y una PriA. Llevar a cabo análisis filogenéticos utilizando secuencias provenientes de los diferentes escenarios, es decir seleccioné enzimas tipo HisA, PriA y aquellas que se encontraban en la zona intermedia para llevar a cabo análisis filogenéticos y de secuencias, en donde propuse un camino evolutivo a nivel de codones para evolucionar de una actividad HisA hacia una TrpF sin perder la actividad ancestral. También encontramos que la distribución obtenida en nuestros análisis se relacionó con la clasificación de los perfiles. Caractericé *in vivo* e *in vitro* junto con la Dra. Gabriela Montero algunas enzimas tipo HisA y PriA selectas y las mutantes propuestas para

evolucionar una actividad HisA a una PriA. La caracterización funcional de estos intermediarios evolutivos nos permitió entender la historia evolutiva de PriA, una isomerasa bifuncional a partir de una enzima HisA promiscua. Este trabajo además sugiere que las enzimas promiscuas, tienen mayor robustez mutacional debido a que como se mostró en este trabajo las mutaciones en estos intermediarios no eliminaron las actividades de ancestrales de la enzima, sólo afectaron un poco sus actividades catalíticas para ambos sustratos.

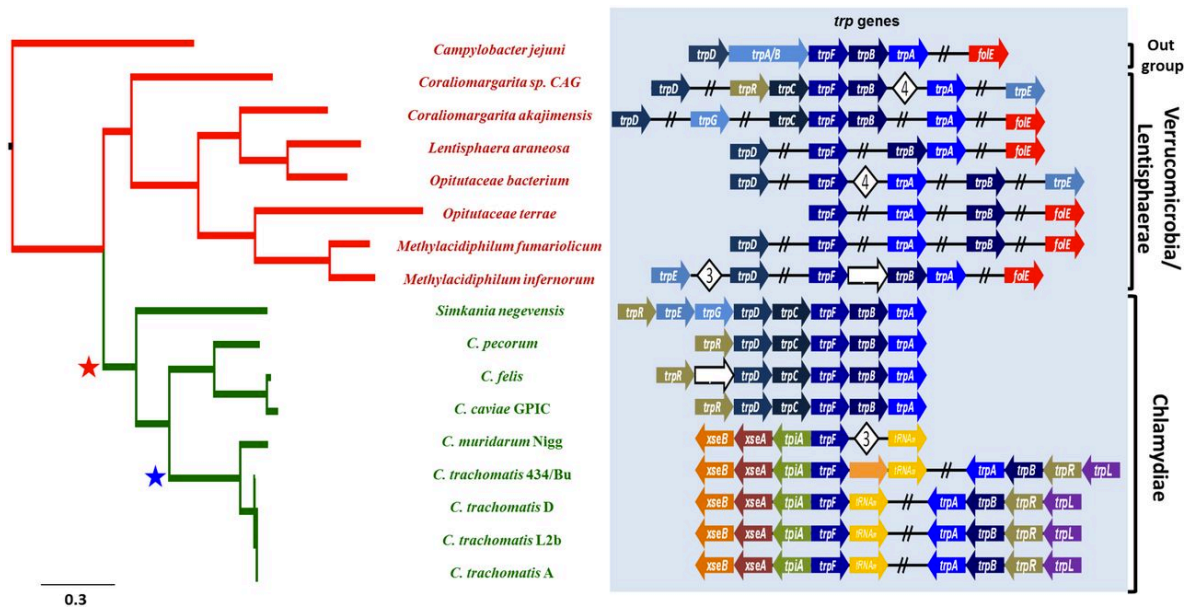
**Promiscuous and adaptable enzymes fill "holes" in the tetrahydrofolate pathway in  
*Chlamydia* species**

Adams N, Thiaville J, Proestos J, **Juárez-Vázquez AL**, McCoy AJ, Barona-Gómez  
F, Iwata-Reuyl D, de Crécy-Lagard V, Maurelli AT.

mBio. 2014.

Otra contribución de mi trabajo de doctorado fue en el artículo titulado “Promiscuous and adaptable enzymes fill "holes" in the tetrahydrofolate pathway in *Chlamydia* species”, el cual fue publicado en revista mBio. En este artículo se plantea la elucidación de la ruta de biosíntesis de Tetrahydrofolatos en especies de *Chlamydia*. A pesar que muchas especies de *Chlamydia* han perdido muchos genes de biosíntesis, han mantenido la biosíntesis de Folatos, especialmente de tetrahydrofolatos (THF), los cuales son cofactores esenciales requeridos por todos los organismos para la síntesis de ADN y aminoácidos. La vía de biosíntesis de THF involucra diez pasos catalizados por 10 enzimas: los genes *folEQBK* codifican para enzimas requeridas para la formación de pterina, los genes *pabABC* codifican para enzimas requeridas para la formación de PABA y los genes *folPCA* codifican para enzimas requeridas para los pasos de ligación de los precursores de pterina y PABA y la glutamilación y reducción. Se ha encontrado que muchas especies de *Chlamydia* no son auxotrofas para folato, lo que sugiere que pueden sintetizar folato de novo, poseen actividad de FoliA y FoliP. También muchas especies contienen homólogos de los genes *folBKPA*. Sin embargo, los genes que codifican para los dos primeros pasos de la biosíntesis FoliE y FoliQ, se encuentran ausentes. En *E.coli* el gen *folE* codifica para una GTP ciclohidrolasa I (EC 3.5.4.16), la cual cataliza la reacción de GTP a 7,8-dihidroneopterina trifosfato. Como ya lo mencionamos éste gen parece estar ausente en *Chlamydia* y lo más probable es que obtenga al precursor de la biosíntesis de THF de la célula huésped o que haya reclutado una enzima para llevar a cabo esta actividad. El objetivo de este

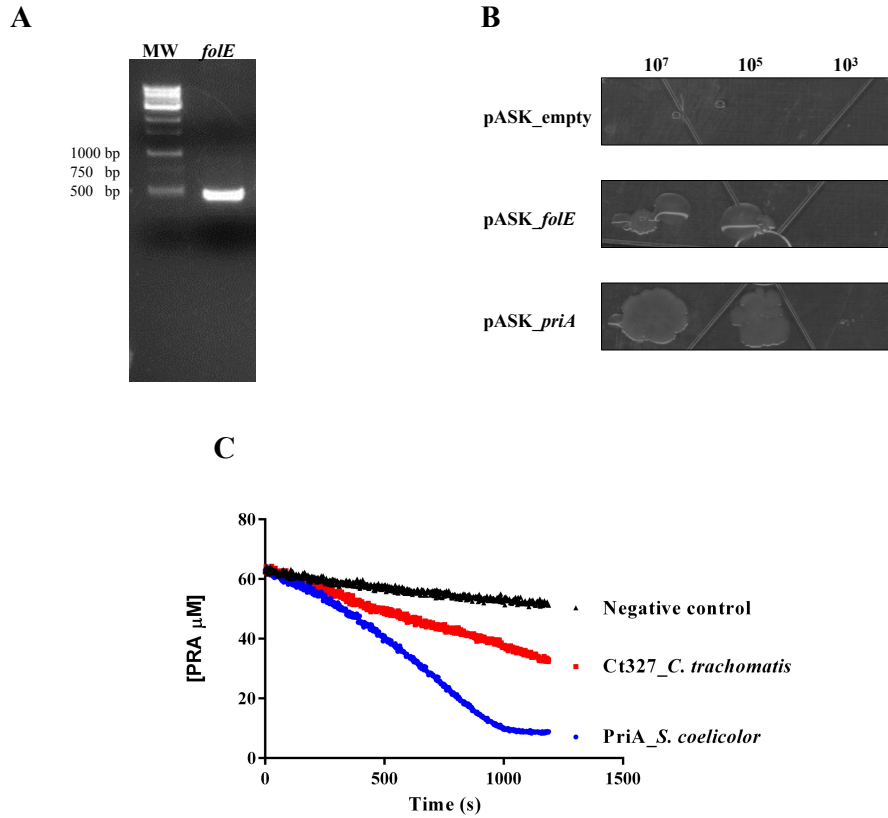
trabajo fue examinar la platicidad existente en el metabolismo de *Chlamydia*, por medio de un screening para encontrar una enzima capaz de llevar a cabo la actividad de *folE*. Y se encontró que el gen *trpF* que codifica para una fosforibosil Antranilato isomerasa puede llevar a cabo esta actividad. Mi contribución a éste trabajo fue llevar a cabo análisis filogenéticos utilizando a la enzima TrpF y analizar el contexto genómico del gen *trpF* en especies de *Chlamydia* y comprobar si ésta enzima además de rescatar la actividad de *folE*, conservaba la actividad Trp. Una de las hipótesis que planteamos fue que debido a la incorporación de éste gen en una nueva ruta era probable que éste gen se encontrara en un contexto genómico diferente al reportado, por ejemplo los genes *trp* casi siempre se encuentran formando un operon (Figura 37).



**Figura 37.** Análisis filogenético y contexto genómico del gen *trpF* en especies de *Chlamydia*. Se distinguen dos clados que se relacionan con la pérdida de los genes *trp* y la retención de *trpF* en un contexto diferente (estrella azul).

Por medio de éste análisis filogenético y del contexto genómico de *trpF*, encontramos que la familia *Verrucomicrobia* y *Lentisphaerae*, familias hermanas a *Chlamydiae* poseen el gen *folE*, y que los genes para la biosíntesis de triptófano se encuentran formando un operón salvo una excepción. De manera relevante, se observa que en la familia *Chlamydiae*, el gen TrpF se sigue conservando, pero

en el clado en donde cae el organismo *C. trachomatis*, el gen *trpF* se encuentra en un contexto diferente, notablemente ninguno de los genes adyacentes a *trpF* en este nuevo contexto participan en la biosíntesis de THF. Estos resultados sugerían que las TrpFs provenientes de éste clado podrían haber perdido la capacidad de catalizar a PRA como sustrato. Por lo que decidimos probar si aún conservaban la actividad de PRA isomerasa mediante ensayos de complementación en cepas de *E. coli* mutantes en *trpF*, la TrpF de *C. trachomatis* complementó dicha mutación. Lo que sugiere que éste gen conserva la actividad TrpF, se intentó obtener proteína pura para obtener los parámetros cinéticos, desafortunadamente, no fuimos capaces de obtener la proteína pura, por lo que llevamos a cabo ensayos con extractos libres de células, y logramos detectar una mínima actividad de TrpF isomerasa, más baja que la detectada en mutantes de PriA de *Streptomyces coelicolor* (Figura 38). En este trabajo además de demostrar cómo la promiscuidad enzimática contribuye a la adaptación de los organismos durante la pérdida de genes, también demuestra que la retención de esta enzima en el genoma se debe a la plasticidad de estas isomerasas tipo  $\beta\alpha_8$  para evolucionar y catalizar reacciones diferentes para las cuales ha evolucionado. Sugiriendo que es probable que las enzimas tipo TrpF que se han retenido en algunos genomas de *Streptomyces* (Verduzco- Castro et al. 2016), puedan evolucionar una actividad de GTP-hidrolasa.



**Figura 38.** Caracterización funcional del gen *trpF* retenido en *Chlamydia*. A) Producto de PCR obtenido, en donde se muestra la amplificación del gen *trpF* (500 pb). B) Complementación de la actividad PRA isomerasa en células de *E. coli trpF*<sup>-</sup>. C) Caracterización in vitro de la actividad de PRA isomerasa de TrpF.

## Referencias

1. Adams NE, Thiaville JJ, Proestos J, Juárez-Vázquez AL, McCoy AJ, Barona-Gómez F, Iwata-Reuyl D, de Crécy-Lagard V, Maurelli AT. Promiscuous and adaptable enzymes fill "holes" in the tetrahydrofolate pathway in *Chlamydia* species. *MBio*. 2014 Jul 8;5(4):e01378-14. doi: 10.1128/mBio.01378-14.
2. Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS. The 'evolvability' of promiscuous protein functions. *Nat Genet*. 2005 Jan;37(1):73-6. PubMed [citation] PMID: 15568024
3. Albalat R, Cañestro C. Evolution by gene loss. *Nat Rev Genet*. 2016 Jul;17(7):379-91. doi: 10.1038/nrg.2016.39. PubMed [citation] PMID: 27087500
4. Alifano P, Fani R, Liò P, Lazcano A, Bazzicalupo M, Carlomagno MS, Bruni CB. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol Rev*. 1996 Mar;60(1):44-69. Review. No abstract available. PubMed [citation] PMID: 8852895, PMCID: PMC239417
5. Amitai G, Gupta RD, Tawfik DS. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J*. 2007 May;1(1):67-78. doi: 10.2976/1.2739115/10.2976/1. PubMed [citation] PMID: 19404461, PMCID: PMC2645560
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008 Feb 8;9:75. doi: 10.1186/1471-2164-9-75. PubMed [citation] PMID: 18261238, PMCID: PMC2265698
7. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. *Journal of Computational Biology*. 2012 May; 19(5): 455-477 PMC [article] PMCID: PMC3342519, PMID: 22506599, DOI: 10.1089/cmb.2012.0021
8. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk HP, Clément C, Ouhdouch Y, van Wezel GP. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev*. 2015 Nov 25;80(1):1-43. doi: 10.1128/MMBR.00019-15. Review. PubMed [citation] PMID: 26609051, PMCID: PMC4711186
9. Barona-Gómez F, Hodgson DA. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep*. 2003 Mar;4(3):296-300. PubMed [citation] PMID: 12634849, PMCID: PMC1315899



10. Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT. The evolution of metabolic networks of *E. coli*. *BMC Syst Biol*. 2011 Nov 1;5:182. doi: 10.1186/1752-0509-5-182. PubMed [citation] PMID: 22044664, PMCID: PMC3229490
11. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci U S A*. 2007 Oct 23;104(43):17004-9. PubMed [citation] PMID: 17942681, PMCID: PMC2040452
12. Bloom JD, Romero PA, Lu Z, Arnold FH. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol Direct*. 2007 Jun 28;2:17. PubMed [citation] PMID: 17598905, PMCID: PMC1914045
13. Bloom JD, Lu Z, Chen D, Raval A, Venturelli OS, Arnold FH. Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol*. 2007 Jul 17;5:29. PubMed [citation] PMID: 17640347, PMCID: PMC1995189
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed [citation] PMID: 24695404, PMCID: PMC4103590
15. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA 3rd, Stevens R, Vonstein V, Wattam AR, Xia F. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep*. 2015 Feb 10;5:8365. doi: 10.1038/srep08365. PubMed [citation] PMID:25666585, PMCID: PMC4322359
16. Bryant JM, Schürch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, Kremer K, van Hijum SA, Siezen RJ, Borgdorff M, Bentley SD, Parkhill J, van Soolingen D. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect Dis*. 2013 Feb 27;13:110. doi: 10.1186/1471-2334-13-110. PubMed [citation] PMID: 23446317, PMCID: PMC3599118
17. Bukau B, Horwich AL. The Hsp70 and Hsp60 chaperone machines. *Cell*. 1998 Feb 6;92(3):351-66. Review. No abstract available. PubMed [citation] PMID: 9476895
18. Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA. Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. *Mol Syst Biol*. 2008;4:177. doi: 10.1038/msb.2008.15. PubMed [citation] PMID: 18364711, PMCID: PMC2290936

19. Chechik G, Oh E, Rando O, Weissman J, Regev A, Koller D. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat Biotechnol.* 2008 Nov;26(11):1251-9. doi: 10.1038/nbt.1499. PubMed [citation] PMID: 18953355, PMCID: PMC2651818
20. Cooper VS, Lenski RE. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature.* 2000 Oct 12;407(6805):736-9. PubMed [citation] PMID: 11048718
21. Copley SD. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr Opin Chem Biol.* 2003 Apr;7(2):265-72. Review. PubMed [citation] PMID: 12714060
22. Copley SD. Toward a systems biology perspective on enzyme evolution. *J Biol Chem.* 2012 Jan 2;287(1):3-10. doi: 10.1074/jbc.R111.254714. Review. PubMed [citation] PMID: 22069330, PMCID: PMC3249082
23. Copley SD. Moonlighting is mainstream: paradigm adjustment required. *Bioessays.* 2012 Jul;34(7):578-88. doi: 10.1002/bies.201100191. Review. PubMed [citation] PMID: 22696112
24. Couñago R, Chen S, Shamoo Y. In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol Cell.* 2006 May 19;22(4):441-9. PubMed [citation] PMID: 16713575
25. Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, Feldmann J, Barona-Gómez F. Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomyces. *Genome Biol Evol.* 2016 Jul 2;8(6):1906-16. doi: 10.1093/gbe/evw125. PubMed [citation] PMID: 27289100, PMCID: PMC4943196
26. Dasmeh P, Serohijos AW, Kepp KP, Shakhnovich EI. The influence of selection for protein stability on dN/dS estimations. *Genome Biol Evol.* 2014 Oct 28;6(10):2956-67. doi: 10.1093/gbe/evu223. PubMed [citation] PMID: 25355808, PMCID: PMC4224349
27. Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet.* 2007 Sep;8(9):675-88. Review. PubMed [citation] PMID: 17703238, PMCID: PMC2488205
28. DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet.* 2005 Sep;6(9):678-87. Review. PubMed [citation] PMID: 16074985

29. Depristo MA. The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP J.* 2007 Jul;1(2):94-8. doi: 10.2976/1.2754665. PubMed [citation] PMID: 19404414, PMCID: PMC2639842
30. Dewar ML, Arnould JP, Dann P, Trathan P, Groscolas R, Smith S. Interspecific variations in the gastrointestinal microbiota in penguins. *Microbiologyopen.* 2013 Feb;2(1):195-204. doi: 10.1002/mbo3.66. PubMed [citation] PMID: 23349094, PMCID: PMC3584224
31. Due AV, Kuper J, Geerlof A, von Kries JP, Wilmanns M. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc Natl Acad Sci U S A.* 2011 Mar 1;108(9):3554-9. doi: 10.1073/pnas.1015996108. PubMed [citation] PMID: 21321225, PMCID: PMC3048130
32. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004 Aug 19;5:113. PubMed [citation] PMID: 15318951, PMCID: PMC517706
33. Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A.* 2000 May 9;97(10):5528-33. PubMed [citation] PMID: 10805808, PMCID: PMC25862
34. Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem.* 1999 Jun 18;274(25):17410-6. PubMed [citation] PMID: 10364169
35. Conservation and topology of protein interaction networks under duplication-divergence evolution. Evlampiev K, Isambert H. *Proceedings of the National Academy of Sciences of the United States of America.* 2008 Jul 16; 105(29): 9863-9868 PMC [article] PMCID: PMC2481380, PMID: 18632555, DOI: 10.1073/pnas.0804119105
36. Evlampiev K, Isambert H. Conservation and topology of protein interaction networks under duplication-divergence evolution. *Proc Natl Acad Sci U S A.* 2008 Jul 22;105(29):9863-8. doi: 10.1073/pnas.0804119105. PubMed [citation] PMID:18632555, PMCID: PMC2481380
37. Fani R, Brillì M, Liò P. The origin and evolution of operons: the piecewise building of the proteobacterial histidine operon. *J Mol Evol.* 2005 Mar;60(3):378-90. PubMed [citation] PMID: 15871048

38. Fani R, Liò P, Chiarelli I, Bazzicalupo M. The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the *hisA* and *hisF* genes. *J Mol Evol.* 1994 May;38(5):489-95. PubMed [citation] PMID: 8028028
39. Fani R, Liò P, Lazcano A. Molecular evolution of the histidine biosynthetic pathway. *J Mol Evol.* 1995 Dec;41(6):760-74. PubMed [citation] PMID: 8587121
40. Felsenstein J, Churchill GA. A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol Biol Evol.* 1996 Jan;13(1):93-104. PubMed [citation] PMID: 8583911
41. Gao B, Gupta RS. Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int J Syst Evol Microbiol.* 2005 Nov;55(Pt 6):2401-12. PubMed [citation] PMID: 16280504
42. Gonzalez O, Gronau S, Falb M, Pfeiffer F, Mendoza E, Zimmer R, Oesterhelt D. Reconstruction, modeling & analysis of *Halobacterium salinarum* R-1 metabolism. *Mol Biosyst.* 2008 Feb;4(2):148-59. doi: 10.1039/b715203e. PubMed [citation] PMID: 18213408
43. He X, Zhang J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics.* 2005 Feb;169(2):1157-64. PubMed [citation] PMID: 15654095, PMCID: PMC1449125
44. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, Sterner R. Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry.* 2002 Oct 8;41(40):12032-42. PubMed [citation] PMID: 12356303
45. Hommel U, Eberhard M, Kirschner K. Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry.* 1995 Apr 25;34(16):5429-39. PubMed [citation] PMID: 7727401
46. Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci.* 1994 May 23;256(1346):119-24. PubMed [citation] PMID: 8029240
47. Iwasaki W, Takagi T. Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages. *PLoS Genet.* 2009 Mar;5(3):e1000402. doi: 10.1371/journal.pgen.1000402. PubMed [citation] PMID: 19266023, PMCID: PMC2644373

48. James LC, Tawfik DS. The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Sci.* 2003 Oct;12(10):2183-93. PubMed [citation] PMID: 14500876, PMCID: PMC2366915
49. Jensen RA. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol.* 1976;30:409-25. Review. PubMed [citation] PMID: 791073
50. Jürgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R. Directed evolution of a (beta alpha)<sub>8</sub>-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A.* 2000 Aug 29;97(18):9925-30. PubMed [citation] PMID: 10944186, PMCID: PMC27628
51. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000 Jan 1;28(1):27-30. PubMed [citation] PMID: 10592173, PMCID: PMC102409
52. Kelkar YD, Ochman H. Genome reduction promotes increase in protein functional complexity in bacteria. *Genetics.* 2013 Jan;193(1):303-7. doi:10.1534/genetics.112.145656. PubMed [citation] PMID: 23114380, PMCID: PMC3527252
53. Khersonsky O, Roodveldt C, Tawfik DS. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol.* 2006 Oct;10(5):498-508. Review. PubMed [citation] PMID: 16939713
54. Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem.* 2010;79:471-505. doi: 10.1146/annurev-biochem-030409-143718. Review. PubMed [citation] PMID: 20235827
55. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature.* 1977 May 19;267(5608):275-6. No abstract available. PubMed [citation] PMID: 865622
56. Klassen JL. Pathway evolution by horizontal transfer and positive selection is accommodated by relaxed negative selection upon upstream pathway genes in purple bacterial carotenoid biosynthesis. *J Bacteriol.* 2009 Dec;191(24):7500-8. doi: 10.1128/JB.01060-09. PubMed [citation] PMID: 19820094, PMCID: PMC2786601
57. Kuper J, Doenges C, Wilmanns M. Two-fold repeated (beta alpha)<sub>4</sub> half-barrels may provide a molecular tool for dual substrate specificity. *EMBO Rep.* 2005 Feb;6(2):134-9. PubMed [citation] PMID: 15654319, PMCID: PMC1299240

58. Li JW, Bolser D, Manske M, Giorgi FM, Vyahhi N, Usadel B, Clavijo BJ, Chan TF, Wong N, Zerbino D, Schneider MV. The NGS WikiBook: a dynamic collaborative online training effort with long-term sustainability. *Brief Bioinform.* 2013 Sep;14(5):548-55. doi: 10.1093/bib/bbt045. PubMed [citation] PMID: 23793381, PMCID: PMC3771235
59. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 2011 Nov 8;10(1):13-26. doi: 10.1038/nrmicro2670. Review. PubMed [citation] PMID: 22064560
60. Nam H, Lewis NE, Lerman JA, Lee DH, Chang RL, Kim D, Palsson BO. Network context and selection in the evolution to enzyme specificity. *Science.* 2012 Aug 31;337(6098):1101-4. doi: 10.1126/science.1216861. PubMed [citation] PMID: 22936779, PMCID: PMC3536066
61. Näsval J, Sun L, Roth JR, Andersson DI. Real-time evolution of new genes by innovation, amplification, and divergence. *Science.* 2012 Oct 19;338(6105):384-7. doi: 10.1126/science.1226521. PubMed [citation] PMID: 23087246, PMCID: PMC4392837
62. Noda-García L, Barona-Gómez F. Enzyme evolution beyond gene duplication: A model for incorporating horizontal gene transfer. *Mob Genet Elements.* 2013 Sep 1;3(5):e26439. PubMed [citation] PMID: 24251070, PMCID: PMC3827091
63. Noda-García L, Camacho-Zarco AR, Medina-Ruíz S, Gaytán P, Carrillo-Tripp M, Fülöp V, Barona-Gómez F. Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Mol Biol Evol.* 2013 Sep;30(9):2024-34. doi: 10.1093/molbev/mst115. PubMed [citation] PMID: 23800623
64. Noda-García L, Juárez-Vázquez AL, Ávila-Arcos MC, Verduzco-Castro EA, Montero-Morán G, Gaytán P, Carrillo-Tripp M, Barona-Gómez F. Insights into the evolution of enzyme substrate promiscuity after the discovery of ( $\beta\alpha$ )<sub>8</sub> isomerase evolutionary intermediates from a diverse metagenome. *BMC Evol Biol.* 2015 Jun 10;15:107. doi: 10.1186/s12862-015-0378-1. PubMed [citation] PMID: 26058375, PMCID: PMC4462073
65. Noda-García L, Camacho-Zarco AR, Verdel-Aranda K, Wright H, Soberón X, Fülöp V, Barona-Gómez F. Identification and analysis of residues contained on beta --> alpha loops of the dual-substrate (beta alpha)<sub>8</sub> phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci.* 2010 Mar;19(3):535-43. doi: 10.1002/pro.331. PubMed [citation] PMID: 20066665, PMCID: PMC2866278

66. Notebaart RA, Teusink B, Siezen RJ, Papp B. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput Biol*. 2008 Jan;4(1):e26. doi: 10.1371/journal.pcbi.0040026. PubMed [citation] PMID: 18225949, PMCID: PMC2211535
67. Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J Bacteriol*. 2008 Apr;190(8):2790-803. doi: 10.1128/JB.01583-07. PubMed [citation] PMID: 18192387, PMCID: PMC2293256
68. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*. 2009;5:320. doi: 10.1038/msb.2009.77. Review. PubMed [citation] PMID: 19888215, PMCID: PMC2795471
69. Ochman H, Moran NA. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*. 2001 May 11;292(5519):1096-9. Review. PubMed [citation] PMID: 11352062
70. Oró J, Mills T, Lazcano A. Comets and the formation of biochemical compounds on the primitive Earth--a review. *Orig Life Evol Biosph*. 1992;21(5-6):267-77. Review. PubMed [citation] PMID: 11537542
71. Ouzounis CA, Kunin V, Darzentas N, Goldovsky L. A minimal estimate for the gene content of the last universal common ancestor--exobiology from a terrestrial perspective. *Res Microbiol*. 2006 Jan-Feb;157(1):57-68. PubMed [citation] PMID: 16431085
72. Palmer DR, Garrett JB, Sharma V, Meganathan R, Babbitt PC, Gerlt JA. Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase. *Biochemistry*. 1999 Apr 6;38(14):4252-8. PubMed citation] PMID: 10194342
73. Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO. Comparison of network-based pathway analysis methods. *Trends Biotechnol*. 2004 Aug;22(8):400-5. PubMed [citation] PMID: 15283984
74. Identification of a two-component fatty acid kinase responsible for host fatty acid incorporation by *Staphylococcus aureus*. Parsons JB, Broussard TC, Bose JL, Rosch JW, Jackson P, Subramanian C, Rock CO. *Proceedings of the National Academy of Sciences of the United States of America*. 2014 Jul 7; 111(29): 10532-10537 PMC [article] PMCID: PMC4115530, PMID: 25002480, DOI: 10.1073/pnas.1408797111
75. Plach MG, Reisinger B, Sterner R, Merkl R. Long-Term Persistence of Bi-functionality Contributes to the Robustness of Microbial Life through Exaptation. *PLoS Genet*. 2016 Jan

29;12(1):e1005836. doi: 10.1371/journal.pgen.1005836. PubMed [citation] PMID: 26824644, PMCID: PMC4732765

76. Priya VK, Sarkar S, Sinha S. Evolution of tryptophan biosynthetic pathway in microbial genomes: a comparative genetic study. *Syst Synth Biol*. 2014 Mar;8(1):59-72. doi: 10.1007/s11693-013-9127-1. PubMed [citation] PMID: 24592292, PMCID: PMC3933628

77. Reams AB, Kofoid E, Savageau M, Roth JR. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics*. 2010 Apr;184(4):1077-94. doi: 10.1534/genetics.109.111963. PubMed [citation] PMID: 20083614, PMCID: PMC2865909

78. Reams AB, Roth JR. Mechanisms of gene duplication and amplification. *Cold Spring Harb Perspect Biol*. 2015 Feb 2;7(2):a016592. doi: 10.1101/cshperspect.a016592. Review. PubMed [citation] PMID: 25646380

79. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012 May;61(3):539-42. doi: 10.1093/sysbio/sys029. PubMed [citation] PMID: 22357727, PMCID: PMC3329765

80. Rouhbakhsh D, Lai CY, von Dohlen CD, Clark MA, Baumann L, Baumann P, Moran NA, Voegtlin DJ. The tryptophan biosynthetic pathway of aphid endosymbionts (*Buchnera*): genetics and evolution of plasmid-associated anthranilate synthase (*trpEG*) within the aphididae. *J Mol Evol*. 1996 Apr;42(4):414-21. PubMed [citation] PMID: 8642610

81. Setiyaputra S, Mackay JP, Patrick WM. The structure of a truncated phosphoribosylanthranilate isomerase suggests a unified model for evolution of the ( $\beta\alpha$ )<sub>8</sub> barrel fold. *J Mol Biol*. 2011 Apr 29;408(2):291-303. doi: 10.1016/j.jmb.2011.02.048. PubMed [citation] PMID: 21354426

82. Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. *Nat Rev Genet*. 2010 Aug;11(8):572-82. doi: 10.1038/nrg2808. Review. PubMed [citation] PMID: 20634811

83. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014 May 1;30(9):1312-3. doi: 10.1093/bioinformatics/btu033. PubMed [citation] PMID: 24451623, PMCID:

PMC3998144



84. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000 Jan 1;28(1):33-6. PubMed [citation] PMID: 10592175, PMCID: PMC102395
85. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* 2005 Sep;3(9):711-21. Review. PubMed [citation] PMID: 16138099
86. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 2009 Oct;19(5):596-604. doi: 10.1016/j.sbi.2009.08.003. Review. PubMed [citation] PMID: 19765975
87. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. *Science.* 2009 Apr 10;324(5924):203-7. doi: 10.1126/science.1169375. Review. PubMed [citation] PMID: 19359577
88. Treangen TJ, Rocha EP. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 2011 Jan 27;7(1):e1001284. doi: 10.1371/journal.pgen.1001284. PubMed [citation] PMID: 21298028, PMCID: PMC3029252
89. Verduzco-Castro EA, Michalska K, Endres M, Juárez-Vazquez AL, Noda-García L, Chang C, Henry CS, Babnigg G, Joachimiak A, Barona-Gómez F. Co-occurrence of analogous enzymes determines evolution of a novel ( $\beta\alpha$ )8-isomerase sub-family after non-conserved mutations in flexible loop. *Biochem J.* 2016 May 1;473(9):1141-52. doi: 10.1042/BJ20151271. PubMed [citation] PMID: 26929404
90. Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol.* 2006;7(5):R39. PubMed [citation] PMID: 16684370, PMCID: PMC1779518
91. Wagner A. Gene duplications, robustness and evolutionary innovations. *Bioessays.* 2008 Apr;30(4):367-73. doi: 10.1002/bies.20728. PubMed [citation] PMID: 18348184
92. Wagner A. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays.* 2005 Feb;27(2):176-88. PubMed [citation] PMID: 15666345
93. Wagner A, de la Chaux N. Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences. *Mol Genet Genomics.* 2008 Nov;280(5):397-408. doi: 10.1007/s00438-008-0373-y. PubMed [citation] PMID: 18751731

94. Wagner A. Robustness, evolvability, and neutrality. *FEBS Lett.* 2005 Mar 21;579(8):1772-8. PubMed [citation] PMID: 15763550
95. Wernersson R, Pedersen AG. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* 2003 Jul 1;31(13):3537-9. PubMed [citation] PMID: 12824361, PMCID: PMC169015
96. Wilmanns M, Eisenberg D. Three-dimensional profiles from residue-pair preferences: identification of sequences with beta/alpha-barrel fold. *Proc Natl Acad Sci U S A.* 1993 Feb 15;90(4):1379-83. PubMed [citation] PMID: 8433995, PMCID: PMC45876
97. Wright H, Noda-García L, Ochoa-Leyva A, Hodgson DA, Fülöp V, Barona-Gómez F. The structure/function relationship of a dual-substrate (betaalpha)<sub>8</sub>-isomerase. *Biochem Biophys Res Commun.* 2008 Jan 4;365(1):16-21. PubMed [citation] PMID: 17967415
98. Wright H, Barona-Gómez F, Hodgson DA, Fülöp V. Expression, purification and preliminary crystallographic analysis of phosphoribosyl isomerase (PriA) from *Streptomyces coelicolor*. *Acta Crystallogr D Biol Crystallogr.* 2004 Mar;60(Pt 3):534-6. PubMed [citation] PMID: 14993684
99. Xie G, Keyhani NO, Bonner CA, Jensen RA. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev.* 2003 Sep;67(3):303-42, table of contents. Review. PubMed [citation] PMID: 12966138, PMCID: PMC193870
100. Xie G, Bonner CA, Song J, Keyhani NO, Jensen RA. Inter-genomic displacement via lateral gene transfer of bacterial trp operons in an overall context of vertical genealogy. *BMC Biol.* 2004 Jun 23;2:15. PubMed [citation] PMID: 15214963, PMCID: PMC471576
101. Yang Z, Kumar S, Nei M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics.* 1995 Dec;141(4):1641-50. PubMed [citation] PMID: 8601501, PMCID: PMC1206894
102. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol.* 2000 Dec 1;15(12):496-503. PubMed [citation] PMID: 11114436
103. Yang Z. Maximum-Likelihood Models for Combined Analyses of Multiple Sequence Data *J Mol Evol.* 1996 May;42(5):587-96. PubMed [citation] PMID: 8662011

104. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007 Aug;24(8):1586-91. PubMed [citation] PMID: 17483113
105. Yao J, Rock CO. How bacterial pathogens eat host lipids: implications for the development of fatty acid synthesis therapeutics. *J Biol Chem.* 2015 Mar 6;290(10):5940-6. doi: 10.1074/jbc.R114.636241. Review. PubMed [citation] PMID:25648887, PMCID: PMC4358231
106. Yizhak K, Tuller T, Papp B, Ruppin E. Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Mol Syst Biol.* 2011 Mar 29;7:479. doi: 10.1038/msb.2011.11. PubMed [citation] PMID: 21451589, PMCID: PMC3094061
107. Metabolic modeling of endosymbiont genome reduction on a temporal scale. Yizhak K, Tuller T, Papp B, Ruppin E. *Molecular Systems Biology.* 2011 Mar 29; 7: 479 PMC [article] PMCID: PMC3094061, PMID: 21451589, DOI: 10.1038/msb.2011.11
108. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008 May;18(5):821-9. doi: 10.1101/gr.074492.107. PubMed [citation] PMID: 18349386, PMCID: PMC2336801
109. Zhao K, Li W, Kang C, Du L, Huang T, Zhang X, Wu M, Yue B. Phylogenomics and evolutionary dynamics of the family Actinomycetaceae. *Genome Biol Evol.* 2014 Sep 22;6(10):2625-33. doi: 10.1093/gbe/evu211. PubMed [citation] PMID: 25245410,PMCID: PMC4224338
110. Zinser ER, Schneider D, Blot M, Kolter R. Bacterial evolution through the selective loss of beneficial Genes. Trade-offs in expression involving two loci. *Genetics.* 2003 Aug;164(4):1271-7. PubMed [citation] PMID: 12930738, PMCID: PMC1462639
111. Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. Evolution of conformational dynamics determines the conversion of a promiscuous generalista into a specialist enzyme. *Mol Biol Evol.* 2015 Jan;32(1):132-43. doi: 10.1093/molbev/msu281. PubMed [citation] PMID: 25312912

## **Anexos**