



**UNIVERSIDAD DE CHILE**

**FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS**

**DETERMINACIÓN DEL ROL DEL SESGO EN EL USO DE  
CODONES EN LA REGULACIÓN DE LA TRADUCCIÓN DE  
GENES DE RESPUESTA A ESTRÉS OXIDATIVO EN BACTERIAS**

**Memoria para optar al título de Bioquímico**

**JOAQUÍN RODRIGO ROJAS GUTIÉRREZ**

**Profesora patrocinante**

**Dra. Amalia Sapag Muñoz de la Peña**

Laboratorio de Farmacoterapia Génica  
Dpto. de Química Farmacológica y Toxicológica  
Facultad de Ciencias Químicas y Farmacéuticas  
Universidad de Chile

**Director de memoria**

**Dr. Assaf Katz Zondek**

Laboratorio de Biología Molecular Bacteriana  
Instituto de Ciencias Biomédicas  
Facultad de Medicina Norte  
Universidad de Chile

**Santiago de Chile**

**2017**

### **Lugar de realización**

Laboratorio de Biología Molecular Bacteriana  
Programa de Biología Celular y Molecular, ICBM  
Facultad de Medicina Norte  
Universidad de Chile

### **Financiamiento**

Esta memoria de título fue financiada por el proyecto FONDECYT 11140222 y el proyecto CONICYT de Inserción en la Academia 79130044.

### **Presentaciones en congresos científicos**

#### **Rojas J, Katz A**

Determinación del papel del uso de codones en la regulación traduccional de genes de respuesta a estrés oxidativo.

XXXVIII Congreso Chileno de Microbiología. Valdivia, Chile, 25 de noviembre de 2016.

#### **Rojas J**

Determinación del rol del uso de codones en genes de respuesta a estrés oxidativo en bacterias.

Reuniones mensuales del Grupo de RNA de la Universidad de Chile. 7 de octubre de 2016.

## ÍNDICE GENERAL

ÍNDICE GENERAL.....	i
ÍNDICE DE TABLAS .....	iii
ÍNDICE DE FIGURAS.....	iii
ABREVIATURAS.....	iv
RESUMEN .....	1
SUMMARY.....	2
1. INTRODUCCIÓN.....	3
1.1. Especies reactivas de oxígeno.....	3
1.1.1. Anión superóxido.....	3
1.1.2. Peróxido de hidrógeno.....	4
1.1.3 Radical hidroxilo.....	4
1.2. Respuesta a estrés oxidativo.....	5
1.2.1 Regulón SoxRS.....	5
1.2.2. Regulón OxyR.....	6
1.2.3. Regulación de la respuesta a nivel traduccional.....	6
1.3. Síntesis de proteínas en procariontes.....	7
1.4. Sesgo en el uso de codones.....	8
1.4.1. Medición del sesgo en el uso de codones.....	9
1.4.2. Sesgo en el uso de codones y la carencia de nutrientes.....	10
1.4.3. Sesgo en el uso de patrones de codones.....	11
1.4.4. Sesgo en el uso de codones en estrés oxidativo.....	12
1.5. Regulación traduccional de la respuesta a estrés oxidativo.....	13
2. HIPÓTESIS.....	15
3. OBJETIVOS.....	15
3.1. Objetivo general.....	15
3.2. Objetivos específicos.....	15
4. MATERIALES Y MÉTODOS.....	16
4.1. Reactivos generales.....	16

4.2. Cambios en niveles proteicos bajo estrés oxidativo.....	16
4.3. Conteo de codones.....	17
4.4. Regresiones lineales de patrones.....	18
4.5. Selección de patrones.....	19
4.6. Análisis de enriquecimiento de funciones en genes.....	23
4.7. Cepas utilizadas.....	23
4.8. Preparación del plásmido para la ligación.....	24
4.9. Preparación de insertos de patrones dobles y triples.....	25
4.10. Ligación de plásmido e inserto.....	26
4.11. Selección de cepas transformadas.....	28
4.12. Medición de fluorescencia bajo estrés oxidativo.....	29
5. RESULTADOS.....	31
5.1. Distribución de parámetros en regresiones.....	31
5.2. Cambio de niveles de traducción <i>in vivo</i> .....	35
5.3. Funciones génicas en común.....	44
6. DISCUSIÓN.....	48
6.1. Idoneidad de la base de datos.....	49
6.2. Conteo de patrones de codones.....	51
6.3. Velocidad de traducción por razón de fluorescencia.....	52
6.4. Efectos a nivel de metabolismo.....	53
7. CONCLUSIÓN .....	55
8. BIBLIOGRAFÍA.....	56
9. ANEXOS.....	59
9.1. contador123.pl (código).....	59
9.2. regresion.pl (código).....	62
9.3. psfGFP-mCherry.....	63
9.3.1. Partidores para la construcción de pBAD30SFIT.....	64
9.3.2. Sitios relevantes de psfGFP-mCherry.....	65
9.3.3. Secuencia.....	66

## ÍNDICE DE TABLAS

TABLA 1: Filtro para valores de r.....	21
TABLA 2: Oligonucleótidos directos para insertos de patrones.....	27
TABLA 3: Oligonucleótidos inversos para insertos de patrones.....	27
TABLA 4: Patrones dobles de codones seleccionados.....	33
TABLA 5: Patrones triples de codones seleccionados.....	35
TABLA 6: Valores de razón de fluorescencia en tiempos seleccionados.....	46
TABLA 7: Análisis de enriquecimiento de genes con CTGGGTAAA.....	47

## ÍNDICE DE FIGURAS

FIGURA 1: Conteo de patrones de codones.....	18
FIGURA 2: Regresión entre conteo de ATCGCT y cambios de expresión bajo H <sub>2</sub> O <sub>2</sub> .....	20
FIGURA 3: Regresión lineal sin datos de conteo cero.....	20
FIGURA 4: Filtro para número de puntos por regresión de patrones dobles	22
FIGURA 5: Pendientes de regresiones obtenidas.....	34
FIGURA 6: Curva de crecimiento bajo estrés con H <sub>2</sub> O <sub>2</sub> .....	38
FIGURA 7: Cambios de síntesis de sfGFP respecto al crecimiento .....	39
FIGURA 8: Cambios de síntesis de mCherry respecto al crecimiento.....	40
FIGURA 9: Cambios de síntesis de sfGFP en dos condiciones.....	42
FIGURA 10: Cambio en velocidad de traducción inducido por estrés.....	45

## ABREVIATURAS

ATF4	Factor transcripcional activador 4 ( <i>activating transcription factor 4</i> )
ATP	Adenosina trifosfato ( <i>adenosine triphosphate</i> )
CAI	Índice de adaptación de codones ( <i>codon adaptation index</i> )
GFP	Proteína fluorescente verde ( <i>green fluorescent protein</i> )
Gly	Glicina ( <i>glycine</i> )
Leu	Leucina ( <i>leucine</i> )
mRNA	RNA mensajero ( <i>messenger RNA</i> )
NADH	Dinucleótido de nicotinamida y adenina ( <i>nicotinamide adenine dinucleotide</i> )
NADPH	Nicotinamida adenina dinucleótido fosfato ( <i>nicotinamide adenine dinucleotide phosphate</i> )
RNA-seq	Secuenciación de RNA ( <i>RNA sequencing</i> )
ROS	Especies reactivas de oxígeno ( <i>reactive oxygen species</i> )
RSCU	Uso relativo de codones sinónimos ( <i>relative synonymous codon usage</i> )
sfGFP	Proteína fluorescente verde de plegamiento rápido ( <i>super folder green fluorescent protein</i> )
SILAC	Marcado isotópico estable con aminoácidos en cultivo celular ( <i>stable isotope labeling with amino acids in cell culture</i> )
SOD	Superóxido dismutasa ( <i>superoxide dismutase</i> )
TAI	Índice de adaptación a tRNA ( <i>tRNA adaptation index</i> )
tRNA	RNA de transferencia ( <i>transfer RNA</i> )

## RESUMEN

La respuesta a estrés oxidativo en bacterias es un proceso complejo que involucra el cambio en los niveles de varias proteínas relacionadas con el aumento de la supervivencia frente a esta condición adversa. Sin embargo, a pesar de que la respuesta a nivel transcripcional a través de los regulones *soxRS* y *oxyR* ha sido descrita en detalle en enterobacterias, ésta no explica completamente la adaptabilidad de una cepa a estas condiciones. En este trabajo se analiza la relación entre el uso de codones y los cambios de expresión génica inducidos en respuesta a estrés oxidativo en *Salmonella enterica* serovar Enteritidis con el fin de encontrar un mecanismo post-transcripcional de regulación génica frente a estrés oxidativo en bacterias. Se tomaron datos de cambio en los niveles de 102 proteínas de *S. Enteritidis* bajo estrés oxidativo y se realizaron regresiones lineales entre ellos y la frecuencia de uso de codones, de patrones dobles de codones y de patrones triples de codones en los mRNA que las codifican. Luego se seleccionaron cinco patrones de codones asociados a regresiones con valores extremos de pendiente en conjunto con un buen valor de coeficiente  $r^2$ . Finalmente, se analizó el efecto *in vivo* de la fusión de estos patrones a un gen reportero *gfp*, midiendo los cambios en sus niveles de traducción inducidos por estrés al añadir  $H_2O_2$  1 mM al medio. Los resultados muestran dos patrones triples de codones capaces de aumentar la traducción de *gfp*, lo cual sugiere que existe una regulación de la traducción en la respuesta a estrés oxidativo.

## SUMMARY

### DETERMINING THE ROLE OF CODON USAGE BIAS ON THE REGULATION OF TRANSLATION OF OXIDATIVE STRESS RESPONSE GENES IN BACTERIA

Bacterial response to oxidative stress is a complex process that involves changes in expression level for many genes required to increase survival rate under conditions of stress. However, what is known about the transcriptional oxidative stress response systems *soxRS* and *oxyR* is not enough to explain the whole response. This thesis is based on the analysis of the relationship between codon usage and gene expression changes associated with oxidative stress response in *Salmonella enterica* serovar Enteritidis. For this purpose, the stress-induced changes in the level of 102 different *S. Enteritidis* proteins were compared to the frequency of codons, dicodons and tricodons present in the corresponding genes through linear regression. Next, five codon pattern regressions associated with the most extreme values of slope and a good  $r^2$  value were introduced into the 5' region of a reporter *gfp* gene. Based on *in vivo* changes of translation levels induced by adding 1 mM  $H_2O_2$ , two tricodon patterns were found to increase *gfp* translation, suggesting a role for codon usage in regulating translation as part of the response to oxidative stress.

## 1.- INTRODUCCIÓN

### 1.1.- Especies reactivas de oxígeno

Las bacterias se ven enfrentadas constantemente a la acumulación de especies reactivas de oxígeno (ROS), las cuales corresponden a una familia de compuestos de oxígeno parcialmente reducido tales como el anión superóxido ( $\cdot\text{O}_2^-$ ), el peróxido de hidrógeno ( $\text{H}_2\text{O}_2$ ) y el radical hidroxilo ( $\cdot\text{OH}$ ). Estas ROS pueden surgir del medio extracelular o como productos de reacciones intracelulares, por ejemplo, durante la autooxidación accidental de flavoproteínas en estado reducido. También se caracterizan por su alto nivel de reactividad, lo cual produce daño por la oxidación de ácidos nucleicos, proteínas y, en algunos casos, lípidos (Farrugia y Balzan, 2012).

#### 1.1.1.- Anión superóxido

El  $\cdot\text{O}_2^-$  es producto del traspaso de un único electrón a la molécula de  $\text{O}_2$ . Debido a su carga negativa, es incapaz de difundir por las membranas de la célula y sólo se genera a través de la oxidación de flavoproteínas o de agentes ciclantes redox como paraquat en el medio intracelular. Su principal blanco de acción son los centros hierro-azufre de enzimas deshidrogenasas y, en el caso de las mutantes  $\Delta\text{sodA}$  y  $\Delta\text{sodB}$  de *Escherichia coli*, que son incapaces de protegerse de estas especies y evitar su acumulación, se observa un alto nivel de mutagénesis debido a los mayores niveles intracelulares de ion ferroso que se liberan de los centros hierro-azufre por efecto del  $\cdot\text{O}_2^-$  acumulado. Este ion ferroso cataliza la reacción de Fenton, la cual genera radical hidroxilo que

finalmente oxida el DNA (Farr *et al.*, 1986). A su vez, las mutantes  $\Delta sodA$  y  $\Delta sodB$  generan auxotrofia para aminoácidos ramificados, azufrados y aromáticos, lo cual se explica por la inactivación por oxidación de enzimas involucradas en la producción de estos aminoácidos (Carlioz y Touati, 1986).

### **1.1.2.- Peróxido de hidrógeno**

El  $H_2O_2$  puede provenir del traspaso de un segundo electrón al anión superóxido o directamente del medio extracelular, cruzando la membrana plasmática, ya que su falta de carga lo permite. Además de atacar centros de hierro-azufre como lo hace el  $\cdot O_2^-$ , el  $H_2O_2$  puede también oxidar residuos de cisteína, formando así ácido sulfénico, enlace disulfuro, ácido sulfinico y ácido sulfónico sucesivamente (Imlay, 2003).

### **1.1.3.- Radical hidroxilo**

El  $\cdot OH$  es la especie más reactiva de las tres y se forma a través de la reacción de Fenton, en la cual el  $H_2O_2$  recibe un electrón de un ion férrico. Su mayor impacto es sobre la mutagénesis del DNA (Imlay, 2003), pero también es capaz de oxidar aminoácidos y generar a partir de ellos especies radicalarias, cuya propagación en la cadena polipeptídica puede desestructurar completamente una proteína (Hawkins y Davies, 1998). Debido a su alta reactividad y su corta vida media en el ambiente intracelular, no viaja muy lejos de su punto de origen y no se han identificado compuestos dedicados a combatir su presencia directamente, como ha sido el caso de la superóxido dismutasa (SOD) para  $\cdot O_2^-$  o la catalasa para  $H_2O_2$ .

## **1.2.- Respuesta a estrés oxidativo**

Al estar expuesta a niveles normales de oxígeno, una bacteria es capaz de lidiar con los niveles intracelulares de ROS a través de la producción de enzimas como la SOD y la catalasa, o de compuestos no enzimáticos como el glutatión o las vitaminas A, C y E. Sin embargo, cuando estos mecanismos no son suficientes para neutralizar las ROS, se activan también diversas vías de señalización que pueden promover la supervivencia o la muerte celular dependiendo de la intensidad y duración del estímulo. Esto es lo que se conoce como respuesta a estrés oxidativo e incluye mecanismos como la activación de los regulones *soxRS* y *oxyR* en *E. coli* o *perR* en *Bacillus subtilis*. Esta respuesta puede implicar la reducción de la motilidad de la cepa, la inducción de la formación de biopelícula (Gambino y Cappitelli, 2016) y la inhibición de la replicación celular (Imlay, 2013).

### **1.2.1.- Regulón *soxRS***

SoxR es una proteína homodimérica de expresión constitutiva que se une a la región 5' del gen *soxS* y sólo es capaz de activar la transcripción de este blanco en respuesta a un estímulo con agentes ciclantes redox, los cuales son normalmente secretados por plantas y bacterias, generando estrés oxidativo con el fin de inhibir el crecimiento de organismos en competencia (Gu e Imlay, 2011). Cuando los dos centros hierro-azufre en su estructura se oxidan, se genera un cambio conformacional en SoxR que permite disminuir la distancia entre los dos hexámeros en la zona promotora de *soxS*, lo cual permite la unión

más eficiente de la RNA polimerasa y el aumento de su transcripción (Hidalgo *et al.*, 1997). El factor transcripcional SoxS, a su vez, activa la transcripción de genes de respuesta a estrés oxidativo. Entre ellos destacan *sodA* y *sodB*, que codifican proteínas SOD que degradan el  $\cdot\text{O}_2^-$  del medio, generando  $\text{H}_2\text{O}_2$  y  $\text{O}_2$ .

### **1.2.2.- Regulón *oxyR***

OxyR es un factor de transcripción de la familia LysR. Se activa por la oxidación del residuo cisteína-199 de su cadena polipeptídica (numeración de la enzima en *E. coli*) a su forma ácido sulfénico por acción de  $\text{H}_2\text{O}_2$ , lo cual genera un enlace disulfuro con la cisteína-208 y permite su unión a genes blanco (Zheng *et al.*, 1998). Un ejemplo de estos genes es *katG*, que codifica la enzima catalasa, cuya función es degradar el  $\text{H}_2\text{O}_2$  en  $\text{H}_2\text{O}$  y  $\text{O}_2$ .

### **1.2.3.- Regulación de la respuesta a nivel traduccional**

A pesar de que se han dilucidado diversos mecanismos transcripcionales de respuesta a estrés oxidativo en bacterias, como la activación de los regulones *soxRS* y *oxyR* en *E. coli*, el único mecanismo traduccional de regulación de estrés oxidativo en bacterias es el descrito por Zhong *et al.* (2015), en el cual se observa una disminución global de la traducción en *E. coli* bajo estrés con  $\text{H}_2\text{O}_2$  causada por la degradación enzimática de los tRNAs. Sin embargo, no se sabe si esto permite actuar sobre genes específicos. En eucariontes se ha descrito un ejemplo en *Schyzosacharomyces pombe* en respuesta a un estímulo con  $\text{H}_2\text{O}_2$ : la acción de las quinasas Gcn2 y Hri2 permite una disminución global de

la traducción y, a la vez, un aumento en la síntesis del factor ATF4, activando así la expresión de genes de respuesta a estrés (Zhan *et al.*, 2004).

### **1.3.- Síntesis de proteínas en procariontes**

La información genética contenida en el DNA es decodificada primero a través de la síntesis de una hebra de RNA mensajero (mRNA) en el proceso de transcripción, la que luego se utiliza para sintetizar una cadena polipeptídica en lo que se conoce como traducción. La traducción es un proceso complejo que involucra el paso del mRNA a través del ribosoma, el cual está compuesto por una subunidad mayor y una subunidad menor, ambas compuestas por proteínas y RNA ribosomal. El proceso comienza con la unión del RNA de transferencia (tRNA) de inicio al mRNA y a la subunidad menor del ribosoma, lo que forma el complejo de preinicio. Éste se une a la subunidad mayor para terminar la fase de inicio, generando los sitios E, P y A en sentido 5' a 3' entre ambas subunidades del ribosoma completo. Luego, la fase de elongación consta de la selección de un tRNA aminoacilado y su unión al sitio A del ribosoma. Este tRNA se une a la secuencia de tres nucleótidos o codón presente en la secuencia del mRNA después del codón AUG donde se formó el complejo de preinicio. Esta unión permitirá la formación posterior de un enlace peptídico entre el aminoácido del tRNA recién unido y el aminoácido del tRNA unido previamente en el sitio P. Esto a su vez permite la translocación de la cadena polipeptídica en conjunto con los tRNA desde los sitios A y P a los sitios P y E. La salida del tRNA sin aminoácido será entonces posible a través del sitio E y la entrada de un nuevo tRNA aminoacilado en el sitio A. De esta forma, a

medida que el ribosoma avanza por el mRNA, se genera una cadena polipeptídica que no entrará a la fase de término de traducción hasta que se encuentre con un codón de término (UAA, UAG o UGA) en la secuencia del mRNA, lo cual permitirá liberar la cadena polipeptídica y disociar las subunidades del ribosoma (Aitken *et al.*, 2010).

#### **1.4.- Sesgo en el uso de codones**

El reconocimiento entre un tRNA aminoacilado y el mRNA se basa en el código genético, el cual indica qué aminoácido corresponde a qué triplete de nucleótidos o codones (Woese *et al.*, 1964). Esta correspondencia depende de los tRNA, los cuales llevan unido un aminoácido específico en un extremo y poseen una secuencia de 3 nucleótidos o anticodón que reconoce el codón del mRNA mediante apareamiento de bases. Dado que los cuatro nucleótidos A, C, U y G definen 64 posibles codones y que tres de ellos codifican posiciones de término, existen 61 tripletes que codifican 20 aminoácidos distintos. Esto implica entonces que hay codones distintos que codifican el mismo aminoácido, los que se conocen como codones sinónimos. Sin embargo, que dos codones codifiquen el mismo aminoácido no implica necesariamente que se produzcan proteínas con estructura y concentración idénticas. Por ejemplo, se han observado casos en que al generarse mutantes sinónimas, se vieron cambios en los niveles y/o en la estructura de la proteína codificada (Kimchi-Sarfaty *et al.*, 2007; Bali y Bebok, 2015). Esto se debe a que la velocidad de traducción puede variar entre dos codones sinónimos durante la fase de elongación, por

ejemplo, debido a diferencias en la disponibilidad de los distintos tRNA aminoacilados o diferencias en la eficacia con la que éstos se unen a los codones. Como se mencionó, esta diferencia de velocidad al momento de traducir un codón sinónimo puede afectar tanto la estructura como la concentración de la proteína obtenida (Yu *et al.*, 2015). Esto explica entonces que existan codones más o menos óptimos para la traducción que sus equivalentes sinónimos. Esta preferencia por ciertos codones sinónimos sobre otros, ya sea a nivel de genoma o a nivel de gen respecto al resto del genoma, se denomina sesgo de uso de codones.

#### **1.4.1.- Medición del sesgo en el uso de codones**

Al determinarse una diferencia entre la traducción de codones y sus equivalentes sinónimos, fue necesario desarrollar una forma de evaluar cuantitativamente cómo el sesgo de uso de codones incide en la expresión génica. Esto se logró, en primer lugar, a través del índice de adaptación de codones o CAI (*codon adaptation index*). Este índice se obtiene a partir de los valores del uso relativo de codones sinónimos (RSCU; *relative synonymous codon usage*) calculados para los 59 codones con al menos un codón sinónimo (descontando los tres codones de término, el codón único de triptofano y el de metionina). El RSCU se obtiene a partir del conteo de codones de los genes de alta expresión para una cepa específica (27 genes de alta expresión en *E. coli*) y corresponde a la razón entre el número de apariciones de un codón en este grupo de genes y el número de apariciones esperadas para ese mismo codón si todos sus equivalentes sinónimos estuviesen en la misma proporción. Ya que

los codones utilizados en genes de alto uso se asocian a los tRNAs más abundantes, es más difícil que se utilicen tRNAs incorrectos en su traducción, lo cual implica una menor tasa de degradación de proteínas mal sintetizadas y, por lo tanto, un gasto energético considerablemente menor para la célula. Es por esto que, en relación al resto del genoma, los genes de alta expresión tienen una preferencia más marcada por codones sinónimos asociados a tRNAs abundantes. Esto significa que los valores de RSCU calculados a partir de los genes de alta expresión son más extremos que los calculados a partir de cualquier otro grupo de genes y esto permite que sean utilizados en el cálculo del CAI para determinar de la forma más precisa qué tan preferidos son, en promedio, los codones de cada gen (Sharp y Li, 1987). En el caso del índice de adaptación de tRNA o TAI (*tRNA adaptation index*), el cálculo es similar, pero se utilizan valores de número de copias de genes de tRNA en vez de los valores del RSCU para generar el índice (dos Reis *et al.*, 2003), ya que este método supone que la concentración de cualquier tRNA en el medio siempre será proporcional al número de genes que lo codifican. Además, este método supone que el uso de codones en los genes de alta expresión se correlaciona con la concentración de los tRNA que los decodifican. Ambas suposiciones se basan en observaciones realizadas en *Escherichia coli* y otras bacterias (Ikemura, 1985).

#### **1.4.2.- Sesgo en el uso de codones y la carencia de nutrientes**

A diferencia de la respuesta al estrés oxidativo, en respuesta a estrés por carencia de nutrientes se han evidenciado algunos mecanismos basados en el

sesgo de uso de codones. En éstos, un número mayor de codones “raros” en genes de respuesta a estrés permite una traducción más eficiente de sus mRNA bajo condiciones de carencia de aminoácidos con respecto al resto del genoma. Esto se debe a que el mismo estrés es capaz de cambiar los niveles de aminoacilación de los diversos tRNA y así permitir que los codones considerados raros vean un aumento relativo en su eficiencia traduccional basado simplemente en una mayor disponibilidad del tRNA aminoacilado correspondiente respecto al resto de los tRNA (Ditmar *et al.*, 2005). Un ejemplo similar es el de los codones UCN de serina (con N = A, C, U o G) presentes en el gen *sinR* de *Bacillus subtilis*. La serina es codificada por 6 codones entre los cuales UCA, UCC, UCU y UCG reducen su eficiencia traduccional al disminuir la disponibilidad de serina, mientras que AGC y AGU parecen no reaccionar frente a esta limitación. Debido a la mayor proporción de codones UCN respecto a AGC y AGU en el gen *sinR*, se promueve entonces una menor traducción del mRNA correspondiente en condiciones de carencia de serina (Subramaniam *et al.*, 2013). Esto permite activar la formación de biopelículas y mejorar la adaptación de la población a la carencia de nutrientes.

#### **1.4.3.- Sesgo en el uso de patrones de codones**

El sesgo en el uso de codones no se limita únicamente a la distribución desigual de codones dentro del genoma. Se ha observado que la distribución de pares de codones también es sesgada, y que este sesgo es adicional al producido por el sesgo de uso de codones o por sesgo en el par de aminoácidos codificados

(Gutman y Hatfield, 1989). Es decir, el efecto de dos mutaciones sinónimas en codones adyacentes sobre la expresión de un gen no es la suma de los efectos de las mutaciones por separado. Esto se puede explicar en parte debido al efecto que tienen estos pares de codones sobre la velocidad de traducción del transcrito, pues se ha visto que los pares de codones mayormente representados (o que aparecen en mayor proporción en el genoma respecto a lo esperado por azar) se traducen más lento que los pares de codones menos representados según ensayos realizados en *E. coli* (Irwin *et al.*, 1995). Este sesgo en el uso de pares de codones es además direccional, lo que implica que el efecto del mismo par de codones es distinto al observado en el par inverso (CUGACC se traduce 2,75 veces más lento que ACCCUG en *E. coli*) y se limita a codones adyacentes. Estos efectos podrían deberse a interacciones directas del mRNA con los rRNA o bien a interacciones entre los tRNA que reconocen los codones en los sitios A, P y E. Adicionalmente, en algunos casos estos pares de codones podrían ser reconocidos por alguna proteína o RNA. Es por esto que al momento de analizar los efectos traduccionales del uso de codones en un genoma cualquiera, es necesario analizar el efecto no sólo del uso de codones, sino también de los pares o incluso tripletes de codones en las secuencias codificantes.

#### **1.4.4.- Sesgo en el uso de codones en estrés oxidativo**

Hasta el momento, sólo se ha publicado una instancia en que el uso de codones afecte la respuesta a estrés oxidativo y es el caso del aumento del

tRNA<sup>Leu</sup><sub>(CAA)</sub> metilado en la tercera posición en *Saccharomyces cerevisiae* cuando crece en presencia de H<sub>2</sub>O<sub>2</sub>, lo cual permite un aumento en la traducción de genes enriquecidos en el codón UUG de leucina (Chan *et al.*, 2012). En nuestro laboratorio, se ha observado un cambio en los niveles de tRNA aminoacilado en condiciones de estrés oxidativo. Específicamente, se observó que a pesar de no disminuirse los niveles totales de tRNA<sup>Gly</sup> en *E. coli* K-12 MG1655 bajo estrés inducido por paraquat, una fracción de ellos se modifica y se vuelven incapaces de cargar glicina, por lo que no pueden participar en la traducción (Leiva *et al.*, comunicación personal). Esto concuerda a su vez con la oxidación observada en aminoacil tRNA sintetasas, proteínas ribosomales y otras proteínas asociadas a la traducción en condiciones de estrés oxidativo (Katz y Orellana, 2012). En conjunto, esto podría implicar que existen codones o patrones de codones cuya eficiencia de traducción cambie frente a estrés oxidativo. En este trabajo se espera encontrar ese tipo de patrones de codones.

### **1.5.- Regulación traduccional de la respuesta a estrés oxidativo**

Se ha descrito que la respuesta a estrés oxidativo en bacterias se realiza principalmente a nivel de regulación de transcritos (Gu e Imlay, 2011). Sin embargo, los efectos ya observados respecto a la disminución global de la traducción frente a este estrés, en conjunto con los efectos diferenciales del estrés sobre los tRNAs isoaceptores de glicina apuntan a la posibilidad de que también exista una regulación de la respuesta al estrés oxidativo a nivel

traduccional. Específicamente, se propone un mecanismo de regulación basado en el uso de codones. Es decir, que la presencia de ciertos codones en la secuencia codificante de genes sería capaz de activar o inhibir funciones metabólicas que, en condiciones de estrés oxidativo, permitirían promover la supervivencia de la cepa.

Si se analiza entonces la correlación entre los cambios de niveles proteicos inducidos por estrés en una cepa bacteriana y los codones presentes en sus secuencias codificantes, se podrá cuantificar el posible efecto que tenga un codón o patrón de codones cualquiera sobre el cambio de expresión inducido por estrés y será posible identificar los codones o patrones de codones más influyentes en la respuesta al estrés. Sin embargo, es importante notar que este método es sólo una aproximación y no es capaz de discernir si el efecto sobre el cambio en los niveles de proteínas es a nivel transcripcional o traduccional. Por lo mismo, es necesario corroborar experimentalmente que el cambio en niveles de proteínas efectivamente se debe a un aumento en la velocidad de traducción, lo cual en este trabajo se logró mediante el estudio de fusiones de los patrones de codones identificados al gen que codifica la proteína fluorescente GFP. En esta tesis se utilizaron los datos publicados por Kim *et al.* (2010), quienes midieron los cambios de expresión de 102 genes de *Salmonella enterica* serovar Enteritidis frente a un estímulo con H<sub>2</sub>O<sub>2</sub>, por lo que la hipótesis y objetivos se limitan a esa cepa en esa condición específica.

## **2.- HIPÓTESIS**

El uso de codones en genes de *Salmonella enterica* serovar Enteritidis permite regular la traducción de sus RNA mensajeros frente a un estímulo con H<sub>2</sub>O<sub>2</sub>.

## **3.- OBJETIVOS**

### **3.1.- Objetivo general**

Encontrar y caracterizar un sistema de regulación de la respuesta a estrés oxidativo en *Salmonella enterica* serovar Enteritidis basado en cambios de velocidad de traducción debido a un uso diferencial de codones.

### **3.2.- Objetivos específicos**

1. Generar una base de datos de uso de codones y combinaciones de codones en genes de *Salmonella enterica* serovar Enteritidis P125109, cuyos productos proteicos estén cuantificados en presencia y ausencia de estrés oxidativo inducido por H<sub>2</sub>O<sub>2</sub>.
2. Encontrar patrones de uso de codones asociados a la respuesta a estrés oxidativo a través del análisis de los datos obtenidos en el Objetivo 1.
3. Corroborar experimentalmente el papel del uso de los codones o patrones de codones identificados en el Objetivo 2 frente a un estímulo con un inductor de estrés oxidativo.

## 4.- MATERIALES Y MÉTODOS

### 4.1.- Reactivos generales

- **BD Bacto** (Franklin Lakes, NJ, EE.UU.): triptona, extracto de levadura, agar.
- **Merck** (Darmstadt, Alemania):  $\text{Na}_2\text{HPO}_4$ ,  $\text{KH}_2\text{PO}_4$ ,  $\text{NaCl}$ ,  $\text{CaCl}_2$ , glicerol 85%,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $\text{H}_2\text{O}_2$  30% (Catálogo 107209).
- **Sigma-Aldrich** (Saint Louis, MO, EE.UU.):  $\text{NH}_4\text{Cl}$ , isoleucina, leucina, valina, glucosa.
- **US Biological** (Salem, MA, EE.UU.): ampicilina, arabinosa.

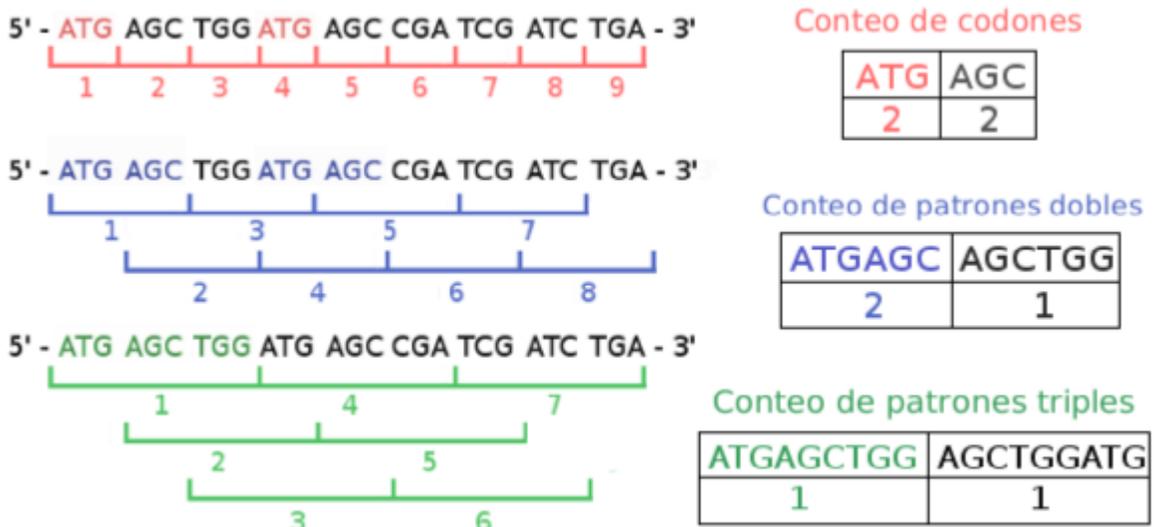
### 4.2.- Cambios en niveles proteicos bajo estrés oxidativo

Los datos cuantitativos para los niveles pre y post inducción de estrés oxidativo en *Salmonella enterica* serovar Enteritidis se obtuvieron del trabajo de Kim *et al.* (2010), en el cual se midieron cambios en los niveles de 102 proteínas luego de tratar la cepa clínica *S. Enteritidis* SE2472 con  $\text{H}_2\text{O}_2$  5 mM durante 2 horas. Esto se logró a través de la separación de las proteínas por electroforesis en dos dimensiones, seguido de su cuantificación a través del método de marcado isotópico estable con aminoácidos en cultivo celular (SILAC). Estos datos se publicaron como porcentaje de cambio de expresión asociado a un valor de desviación estándar para cada gen. Estos valores son los que se utilizaron en análisis posteriores.

### **4.3.- Conteo de codones**

Se escribió un programa en el lenguaje Perl 5 llamado contador123.pl (Anexo1) que se utilizó para realizar el conteo de codones, patrones dobles y patrones triples de codones en los 102 genes de la base de datos de *S. Enteritidis*. Sin embargo, *S. Enteritidis* SE2472 corresponde a una cepa clínica cuyo genoma no se encuentra secuenciado y fue necesario utilizar el genoma de *S. Enteritidis* P125109 como reemplazo, ya que esta es la única cepa de *Enteritidis* cuyo genoma está secuenciado. El programa contador123.pl toma todas las secuencias codificantes del genoma de *S. Enteritidis* P125109 (RefSeq NC\_011294.1), las separa en ventanas de 3, 6 y 9 nucleótidos dentro del marco de lectura e imprime en una tabla la cantidad de veces que aparece cada codón o patrón de codones en cada gen (Figura 1). Se cuentan cada uno de los posibles 64 codones, 3.904 patrones dobles y 238.144 patrones triples en cada gen, sin tomar en cuenta patrones dobles que incluyan un codón de término en la primera posición o patrones triples que lo tengan en su primera o segunda posición. Este conteo genera tres archivos separados según el tamaño del patrón de codones, incluyendo codones, patrones dobles de codones y patrones triples. También incluye otros tres archivos que reemplazan el conteo total de codones/patrones por el mismo conteo normalizado por el número total de patrones dentro del gen, referido de aquí en adelante como frecuencia del patrón. Cabe mencionar que existen otros programas de conteo de codones disponibles de forma pública. Simplemente se creó uno nuevo debido a que la mayoría de los programas disponibles no permiten contar patrones de codones,

además de la comodidad que implica obtener conteos de codones, patrones dobles y triples en tablas con el mismo formato.



**Figura 1. Conteo de patrones de codones.** El programa contador123.pl permite dividir las secuencias codificantes de un genoma, provenientes de un archivo en formato FASTA, en codones, patrones dobles y patrones triples de codones, generando así tablas con las cantidades exactas de cada codón/patrón en cada gen.

#### 4.4.- Regresiones lineales de patrones

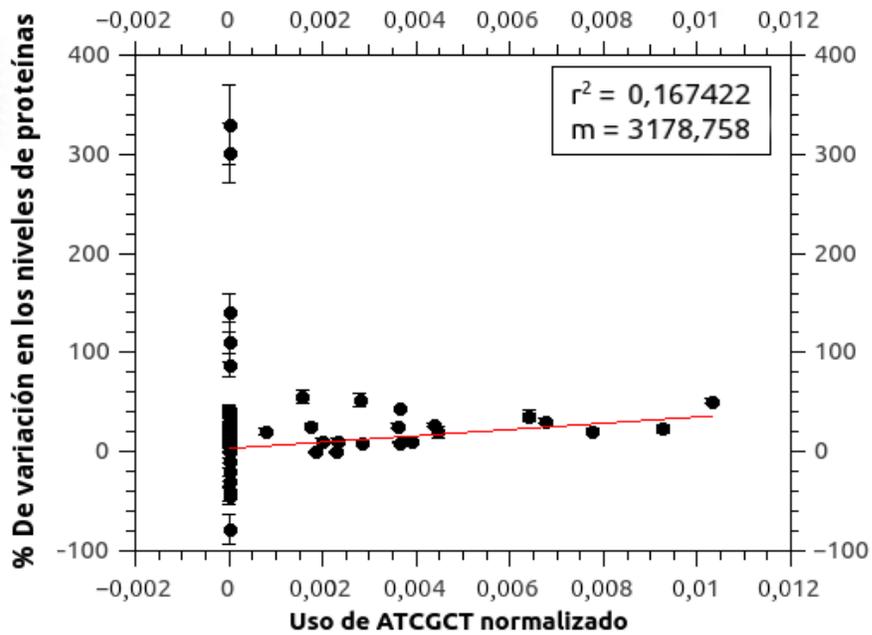
A partir de los datos de cambio de expresión génica bajo estrés oxidativo y conteo de patrones de codones, se realizaron múltiples regresiones lineales en las cuales se relaciona el conteo de un codón o patrón de codones en todos los genes de la base de datos (eje X) con el cambio de expresión de estos mismos genes bajo estrés (eje Y). Para realizar estas regresiones se escribió un programa en el lenguaje Perl 5 llamado regresion.pl (Anexo 2), el cual toma los datos de ambas tablas y realiza una regresión lineal entre la frecuencia de cada codón/patrón y los cambios de expresión respectivos. La regresión toma en cuenta los valores de desviación estándar para los 102 genes analizados,

generando así una lista que contiene los valores del coeficiente de determinación  $r^2$  y los valores de pendiente para cada regresión lineal. Una de estas regresiones se observa en la Figura 2. Es importante notar que para todos los casos se utilizó el conteo normalizado y no el conteo total, ya que permite una mayor dispersión de datos en el eje X y además considera el efecto del tamaño del gen en la velocidad de traducción de su mRNA.

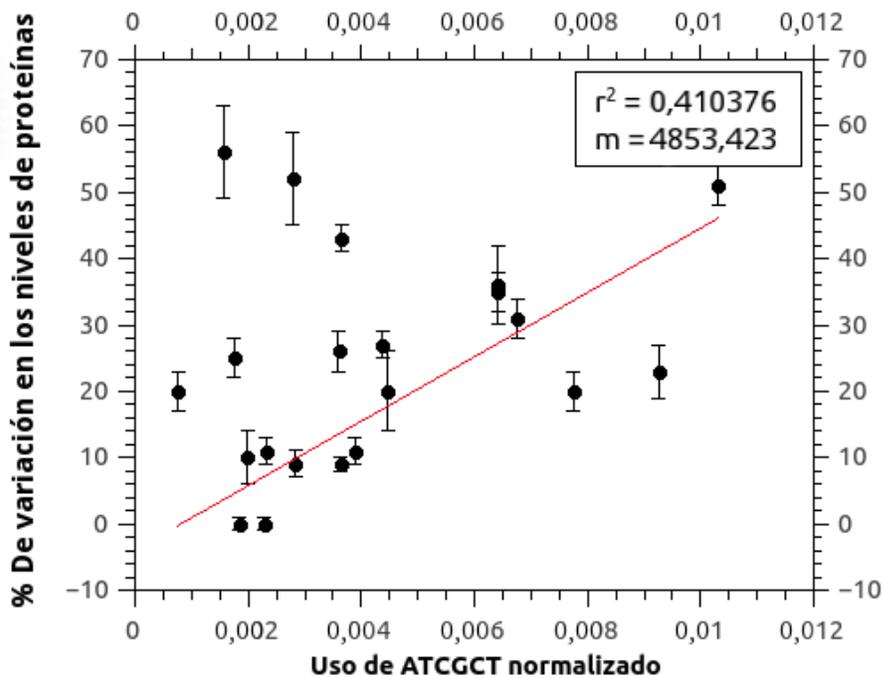
#### **4.5.- Selección de patrones**

Lo primero que se realizó a partir de los datos de regresión fue eliminar todos los datos correspondientes a genes con cero copias del patrón específico. Esto debido a que, tanto en regresiones de conteo de patrones dobles como de patrones triples, se genera ruido debido a la gran cantidad de puntos en el eje Y. Es decir, no se generan regresiones comparables debido a que la mayor parte de la dispersión de los datos se genera en un solo punto del eje X. Esto genera regresiones con menor número de datos, pero con valores mayores de  $r^2$  y de pendiente (Figura 3).

Con el fin de determinar los codones o patrones más influyentes en la respuesta contra el estrés oxidativo, se aplicaron tres filtros a los datos de regresión lineal. En primer lugar, se utilizaron como referencia los valores críticos al 5% de significancia del coeficiente de Pearson o coeficiente de correlación  $r$  obtenidos de una prueba T de Student según el número de puntos por regresión. Si el valor de  $r$  en una regresión es mayor al valor crítico positivo o menor al valor



**Figura 2. Regresión entre conteo de ATCGCT y cambios de expresión bajo H<sub>2</sub>O<sub>2</sub>.** La figura grafica los datos de cambio en los niveles de proteínas producido por un estímulo con H<sub>2</sub>O<sub>2</sub> según el conteo del patrón doble ATCGCT en cada uno de los 102 genes de la base de datos. El recuadro marca el valor de pendiente (m) y de coeficiente de determinación r<sup>2</sup> obtenidos de la regresión lineal calculada por el programa regresion.pl.



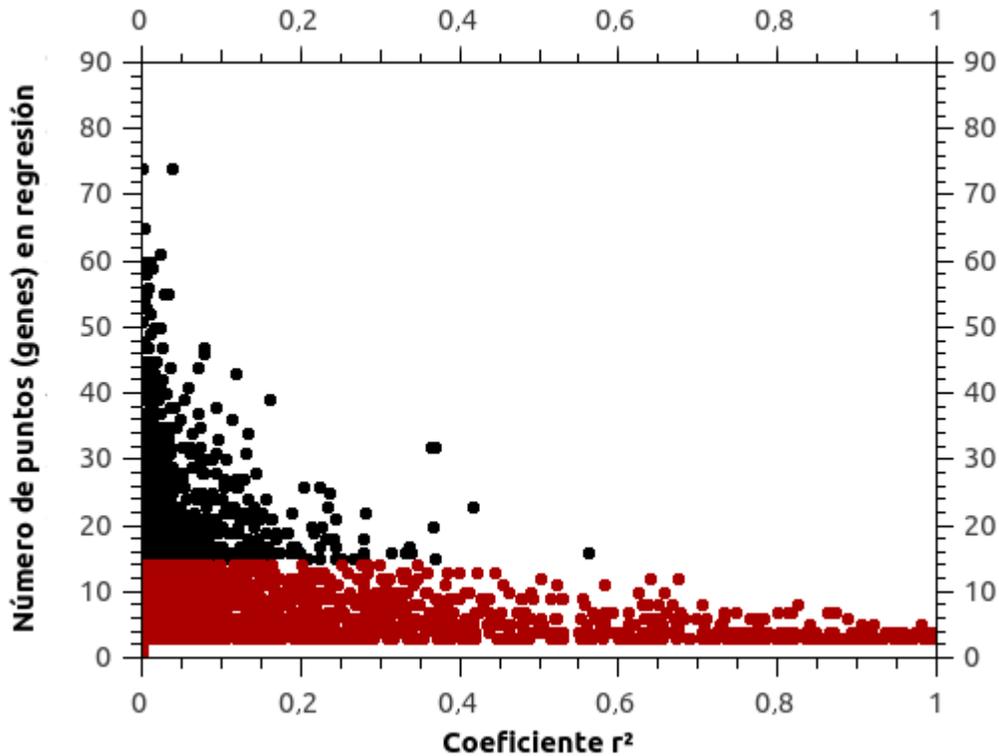
**Figura 3. Regresión lineal sin datos de conteo cero.** El gráfico corresponde a la misma regresión de la Figura 2, pero habiendo eliminado todos los puntos correspondientes a genes con cero copias del patrón ATCGCT para mejorar los valores de pendiente y r<sup>2</sup>.

crítico negativo asociado al número de puntos que incluye, se considera que el valor de  $r$  es significativamente mayor a cero y que la regresión tiene valor predictivo. De esta forma, se eliminaron todas las regresiones cuyos valores de  $r$  se encontraban entre los valores positivo y negativo del valor crítico según su número de puntos. Un ejemplo de esto se observa en la Tabla 1.

**Tabla 1. Filtro para valores de  $r$ .** Se muestran dos ejemplos en los cuales se comparan los valores del coeficiente de correlación de Pearson o " $r$ " con sus respectivos valores límite, los cuales aumentan a medida que la regresión tiene menos puntos. Ya que ambos valores de  $r$  son positivos, si  $r$  es menor al valor límite, la regresión se elimina del conjunto de datos, pero si es mayor o igual, se mantiene.

Patrón	Pendiente	$r^2$	$r$	Puntos	Valor límite	Filtro
AAATAA	-4877,33	0,4067	0,6377	10	0,632	Aceptada
AAATAC	-1207,76	0,0251	0,1586	22	0,423	Rechazada

Como segundo filtro, se eliminaron todas las regresiones con un número muy bajo de puntos. Esto se hizo debido a la tendencia por azar de generarse regresiones de pendientes muy altas o muy bajas, las cuales pueden no ser representativas de lo que ocurre a nivel de genoma debido a los pocos genes incluidos en la base de datos utilizada. Para guiar la elección del número mínimo de genes requeridos, se utilizó un gráfico que relaciona número de puntos dentro de cada regresión (genes dentro de los 102 con al menos una copia del patrón) con los valores de  $r^2$  asociados a cada una de ellas. En este gráfico se estableció arbitrariamente el punto de corte como el número de puntos donde el valor de  $r^2$  tiende a aumentar de manera más brusca (Figura 4). Luego se eliminaron todas las regresiones con número de puntos menor a este valor, el cual fue 15 para patrones dobles y 5 para patrones triples de codones.



**Figura 4. Filtro para número de puntos por regresión de patrones dobles.** Lo que se observa es una tendencia de los valores del coeficiente de determinación  $r^2$  a aumentar a medida que el número de genes por regresión tiende a cero. Se tomaron solamente regresiones de 15 o más genes en regresiones de patrones dobles para asegurar que los valores de  $r^2$  no se deben al azar. Los valores de  $r^2$  para las regresiones de 0, 1 o 2 genes fueron definidas como cero de manera automática.

Finalmente, de las regresiones disponibles después de los primeros dos filtros, se seleccionaron los patrones asociados a regresiones con los valores mayores y menores de pendiente. Para esto se utilizó como punto de corte el valor de pendiente que permitiría predecir un cambio de expresión de un 80% en un gen de longitud promedio (317 codones), lo que debiese ser medible mediante los experimentos de fusión traduccional de los patrones al gen que codifica GFP que se realizaron (ver más abajo). Estos valores deben ser al menos de 25.280 para patrones dobles y 25.200 para patrones triples (obtenido de la

multiplicación entre el valor esperado de 80% y los largos promedios 316 y 315 para patrones dobles y triples respectivamente, ya que una pendiente de valor 1 corresponde a un cambio de un 1% por cada copia del patrón normalizado por el tamaño del gen). Sin embargo, ya que los valores obtenidos de pendientes no alcanzaron este punto de corte, se decidió insertar cuatro copias de cada patrón doble y triple al gen reportero *gfp*, lo cual permitió utilizar un punto de corte cuatro veces menor en cada caso (6.320 y 6.300 para patrones dobles y triples).

#### **4.6.- Análisis de enriquecimiento de funciones en genes**

Se realizó un análisis de enriquecimiento de funciones para los todos los genes de *S. Enteritidis* P125109, cuyas secuencias incluyesen al menos una copia del patrón seleccionado. Esto se realizó con la base de datos pathways de BioCyc (<https://ecocyc.org>), la cual clasifica todos los genes del grupo ingresado según vía funcional o proceso en el cual participa su producto génico y ordena todas estas funciones génicas de menor a mayor valor  $p$ , el cual representa la probabilidad de que las funciones encontradas en este grupo sean producto de seleccionar genes del genoma al azar. A partir de estos datos se seleccionaron las funciones más relevantes según los valores  $p$  obtenidos (valor  $p < 0,01$ ) y se agregaron a una tabla con el fin de comparar funciones enriquecidas a nivel de genoma entre todos los grupos.

#### **4.7.- Cepas utilizadas**

Para medir los cambios en los niveles de traducción inducidos por estrés

oxidativo se utilizó una cepa *S. Enteritidis* NCTC 13349 (Betancor *et al.*, 2012) donada gentilmente por el Dr. Carlos Santiviago. Se la transformó con el plásmido pBAD30SFIT, construido por la Dra. Sara Elgamal (Anexo 3), de aquí en adelante referido como psfGFP-mCherry, basado en el plásmido pBAD30XS utilizado por Elgamal *et al.* (2016) que contiene los insertos correspondientes a los patrones de codones en la secuencia codificante de sfGFP. Estas variantes del plásmido se obtuvieron a partir del psfGFP-mCherry que se conserva en una cepa *E. coli* JM109 provista por el Dr. Omar Orellana.

#### **4.8. - Preparación del plásmido para la ligación**

El plásmido psfGFP-mCherry es un derivado del plásmido con promotor inducible por arabinosa pBAD30, al cual se le insertó un gen de la variante termoestable de la proteína fluorescente verde (*super folder* GFP o sfGFP) en fusión transcripcional con el gen que codifica la proteína fluorescente roja mCherry, ambos con su propia secuencia Shine Dalgarno de inicio de la traducción. Este plásmido se extrajo de una cepa *E. coli* JM109 a través de un kit de extracción de DNA Qiagen spin miniprep, siguiendo el protocolo recomendado por el proveedor para la extracción para plásmidos de bajo número de copias, basado en técnicas de extracción por lisis alcalina (Birnboim y Doly, 1979). Este plásmido se concentró en un speedvac DNA 110 marca Savant a 65 °C por 20 minutos y se dializó contra agua ultrapura (conductividad 18,2 MΩ·cm ) durante una hora empleando una membrana de microdiálisis de ésteres mixtos de celulosa con poros de 0,025 μm y 47 mm de diámetro marca Millipore (Burlington, MA, EE.UU.). Se digirió 1 μg de este

plásmido en 20  $\mu$ L de reacción con 10 U de *Xho*I y 10 U de *Spe*I, ambas de marca New England Biolabs (Ipswich, MA, EE.UU.). Esta reacción incluía también tampón comercial NEB 2 (compuesto de NaCl 50 mM, MgCl<sub>2</sub> 10 mM, ditioneitol (DTT) 1 mM y Tris-HCl 10 mM ajustado a pH 7,9 a 25 °C) y albúmina de suero bovino (BSA) para llegar a 0,1 mg/mL. Las enzimas *Xho*I y *Spe*I cortan en el cuarto y noveno codón de la zona codificante de *gfp* respectivamente, generando los extremos cohesivos que permitirán unir un inserto que contiene las 4 copias del patrón de codones que se quiera analizar. La reacción de digestión se dejó a 37°C toda la noche. Se agregaron luego 0,025 U de fosfatasa alcalina marca Promega (Madison, WI, EE.UU.) y se incubó a 37°C durante media hora para desfosforilar los extremos cohesivos del plásmido digerido. Se agregaron otras 0,025 U de fosfatasa alcalina y se incubó nuevamente durante media hora. Luego se incubó a 80°C durante 40 minutos para inactivar la fosfatasa y las enzimas de restricción. Se añadieron 2,5  $\mu$ L de acetato de sodio 3 M pH 5,5 y 75  $\mu$ L de etanol frío y se incubó a -20 °C toda la noche. Luego se centrifugó a 15000 rpm a 4 °C durante 30 minutos seguido de tres lavados con 500  $\mu$ L de etanol frío al 75% v/v con centrifugaciones de 15000 rpm a 4 °C durante 10 minutos entre lavados. Finalmente, el plásmido se secó bajo mechero y se resuspendió en 25  $\mu$ L de agua ultrapura (conductividad 18,2 M $\Omega$ ·cm ), seguido de una incubación de 5 minutos a 60 °C para disolver mejor la pella obtenida.

#### **4.9.- Preparación de insertos de patrones dobles y triples**

Los insertos correspondientes a cada patrón doble y patrón triple se prepararon

por hibridación de dos oligonucleótidos complementarios sintetizados en Integrated DNA Technologies, Inc. (Coralville, IA, EE.UU.) y se insertaron en el plásmido psfGFP-mCherry. Los insertos están conformados por los sitios cohesivos complementarios a los generados en el plásmido por *XhoI* y *SpeI*, con 4 copias del patrón doble o triple específico entre ellos. Los oligonucleótidos directo e inverso se diluyeron con agua para llevarlos a una concentración de 100  $\mu$ M. Luego, 5  $\mu$ L se trataron con 10 U de polinucleótido quinasa T4 marca Thermo Scientific (Waltham, MA, EE.UU.) en una reacción de 20  $\mu$ L con ATP 1 mM en tampón comercial ( $MgCl_2$  10 mM, DTT 5 mM, espermidina 0,1 mM y Tris HCl 50 mM ajustado a pH 6,4 a 25 °C) para fosforilar sus extremos. Cada par directo-inverso se hibridó mezclando partes iguales de 5  $\mu$ L de cada uno, incubando en un termobloque Standard Heatblock I de marca VWR (Radnor, PA, EE.UU.) a 95°C durante media hora y dejando enfriar en el mismo bloque hasta temperatura ambiente. Las secuencias de los oligonucleótidos directos e inversos utilizados para armar los insertos se detallan en las Tablas 2 y 3.

#### **4.10.- Ligación de plásmido e inserto**

Se ligaron 50 ng de plásmido digerido a cada dúplex de oligonucleótidos en una proporción molar 1:3 con 400 U de T4 DNA ligasa marca New England Biolabs en una reacción de 10  $\mu$ L con tampón comercial de ligasa (compuesto por  $MgCl_2$  10 mM, DTT 10 mM, ATP 1 mM y Tris-HCl 50 mM ajustado a pH 7,5 a 25 °C). Esta reacción se dejó a 4°C toda la noche y se dializó durante 1 hora en membranas de microdiálisis con poros de 0,025  $\mu$ m marca Millipore. Se transformaron bacterias *E. coli JM109* electrocompetentes, preparadas a partir



de un inóculo del día anterior, del cual se tomaron 0,5 mL para inocular 50 mL de medio LB fresco. El cultivo se incubó a 37 °C hasta alcanzar una  $DO_{600}$  de 0,6 y se lavó 4 veces con 7 mL de agua de 18,2  $M\Omega \cdot cm$  de resistividad a 25°C (de aquí en adelante referida simplemente como agua) seguido de centrifugaciones de 4000 g a 4 °C durante 5 minutos con una resuspensión final en 1 mL de agua. Se mezclaron 90  $\mu L$  de estas *E. coli* electrocompetentes con los 10  $\mu L$  de reacción de ligación dializada y se traspasaron a una cubeta de electroporación de 1 mm, donde se sometieron a un choque eléctrico de 1,8 kV, 25  $\mu F$  de capacitancia y 200  $\Omega$  de resistencia en un equipo Gene Pulser Xcell marca BioRad (Hercules, CA, EE.UU.). Tras esto, se añadió 1 mL de medio de cultivo LB sin antibiótico y se traspasó a un tubo de microcentrífuga de 1,6 mL. Éste se incubó a 37 °C sin agitación durante una hora, luego de lo cual se sembraron 200  $\mu L$  en placas de medio LB-agar con ampicilina que se incubaron a 37 °C toda la noche.

#### **4.11.- Selección de cepas transformadas**

Se inocularon dos tubos de microcentrífuga con 980  $\mu L$  cada uno de medio mínimo M9 ( $Na_2HPO_4$  47,7 mM,  $KH_2PO_4$  22 mM, NaCl 8,6 mM,  $NH_4Cl$  18,7 mM,  $MgSO_4$  2 mM,  $CaCl_2$  0,11 mM y glicerol 0,4%) suplementado con aminoácidos ramificados (isoleucina, leucina y valina 50  $\mu g/mL$  cada uno) con las colonias transformadas. Se añadieron 20  $\mu L$  de glucosa 20% a un tubo y 20  $\mu L$  de arabinosa 20% al otro y se incubaron a 37°C toda la noche con agitación. La fluorescencia asociada a GFP (excitación a  $480 \pm 4,5$  nm y emisión a  $515 \pm 10$

nm), mCherry (excitación a  $555 \pm 4,5$  nm y emisión a  $600 \pm 10$  nm) y la  $DO_{600}$  de 200  $\mu$ L de estos inóculos se midieron en triplicado a través de un lector de placas multipocillos INFINITE M200PRO marca TECAN (Männedorf, Suiza) con el fin de determinar si existía o no inducción de fluorescencia por arabinosa, utilizando el medio con glucosa como control.

En las colonias en que se detectó una inducción de fluorescencia por arabinosa, se corroboró si efectivamente poseían el inserto a través de reacciones de PCR, utilizando como molde el plásmido obtenido mediante el kit DNA Qiagen spin miniprep. En estas reacciones también se utilizó como partidor directo el oligonucleótido directo que se utilizó para construir el inserto originalmente y un partidor inverso (TGCCCATTAACATCACCATC) que permitiría la amplificación de un fragmento de 214 o 226 pares de bases (patrones dobles y triples respectivamente) en caso de que el inserto se hubiese ligado correctamente al plásmido. Las reacciones se hicieron en 20  $\mu$ L, cada una con 1 U de Paq5000 DNA polimerasa marca Agilent Technologies (Santa Clara, CA, EE.UU.), 0,2  $\mu$ L de cada partidor en concentración 10  $\mu$ M, 0,2  $\mu$ L de una mezcla de dNTPs 100 mM, tampón de Paq (del cual el proveedor sólo especifica que contiene  $Mg^{2+}$  2 mM), 7,3  $\mu$ L de agua y 2  $\mu$ L de molde. Los plásmidos de las cepas cuyos insertos se amplificaron por PCR se secuenciaron para confirmar la ausencia de mutaciones, servicio que se contrató en MacroGen, Inc. (Seúl, Corea del Sur).

#### **4.12.- Medición de fluorescencia bajo estrés oxidativo**

Se transformó la cepa *S. enteritidis* NCTC 13349 con los plásmidos de interés a través de electroporación con el mismo protocolo descrito anteriormente. Se resuspendieron las cepas transformadas en 2 mL de medio mínimo M9 con ampicilina 100 µg/mL y se incubaron a 37 °C toda la noche con agitación. Se tomaron luego 30 µL para inocular 3 mL de medio M9 fresco con ampicilina. Cuando los cultivos alcanzaron una DO<sub>600</sub> de 0,6, se tomaron 50 µL para inocular un pocillo con 150 µL de medio M9 fresco con ampicilina con suficiente arabinosa para alcanzar una concentración final de 0,4% p/v. Los pocillos se incubaron con agitación a 37 °C durante 12 horas en el lector de placas multipocillos, donde se midieron a intervalos regulares la fluorescencia de GFP y mCherry, así como la DO<sub>600</sub> como se describió previamente. Una única placa de 96 pocillos contenía 4 réplicas técnicas de cada cepa, a su vez de incluir condiciones con y sin estímulo de estrés y una cepa con el plásmido psGFP-mCherry sin inserto.

## 5.- RESULTADOS

### 5.1.- Distribución de parámetros en regresiones

Luego de una exhaustiva búsqueda, solamente se encontró publicada una base de datos de cuantificaciones de los niveles de proteína en condiciones de estrés oxidativo. Esta base de datos indica cambios en los niveles de 102 proteínas de la cepa clínica *S. Enteritidis* SE2472 bajo estrés inducido por H<sub>2</sub>O<sub>2</sub> 5 mM comparados a una condición control (Kim *et al.*, 2010). Ya que la cepa SE2472 no está aún secuenciada, el conteo de codones, patrones dobles y patrones triples de codones se realizó en los 102 genes equivalentes en *S. Enteritidis* P125109. A partir de estos conteos se realizaron regresiones lineales con el fin de determinar qué codón o patrón de codones tiene el mayor efecto sobre el cambio en los niveles de proteínas en condiciones de estrés oxidativo. Este análisis incluyó regresiones lineales para el conteo de los 64 codones, 3.904 patrones dobles y 238.144 patrones triples de codones. Se compararon los valores de coeficiente de determinación  $r^2$ , los valores de pendiente y la cantidad de genes con conteo de codón/patrón mayor a cero incluidos en ellas.

De las 64 regresiones lineales realizadas sobre conteo de codones, sólo las de GGC y GCT poseían valores de coeficiente de correlación  $r$  sobre el valor crítico según su número de puntos. Estos codones, que codifican glicina y alanina respectivamente, poseían valores positivos de pendiente en su regresión, pero ninguno alcanzó siquiera un incremento teórico de 5% en la cantidad de proteína asumiendo 4 copias del codón añadidas a un gen de largo promedio

(317 codones). Ya que nuestro cambio mínimo esperado era un 80%, esto significa que no eran suficientemente influyentes sobre la síntesis de proteínas para ser utilizadas en el análisis experimental posterior y, por lo tanto, se descartaron.

De las 3.904 regresiones lineales de patrones dobles de codones, 1.343 de ellas o un 34,4% no incluían suficientes puntos, es decir, correspondían a patrones que sólo aparecían en 0, 1 o 2 genes dentro del conjunto de 102 y no podían ser utilizadas en una regresión lineal. Esto en contraste con el 0,5% de patrones dobles que no generarían regresiones válidas de haber tenido disponibles los datos de cambios de expresión para el genoma completo. De las 2.561 regresiones restantes, 119 poseían un coeficiente  $r$  sobre el valor límite y de éstas sólo 34 incluían más de 15 puntos (En la sección de métodos se pueden encontrar más detalles sobre estos criterios de selección). A partir de éstas, se seleccionaron regresiones con un valor absoluto de pendiente mayor a 6.340, el cual corresponde al valor de pendiente necesario para generar un cambio de expresión de al menos un 80% con 4 copias del patrón en un gen de largo promedio y marca el 7,24% superior de valores de pendiente respecto a las 3.904 regresiones con patrones dobles y un 11,7% superior respecto a las 34 regresiones obtenidas tras aplicar ambos filtros. En el caso de las pendientes negativas, sólo GACCAG tiene una pendiente menor a -6.340, pero se seleccionaron adicionalmente CTGGTT, GTTATC y TATCAG, cuyas pendientes fueron menor a -5.000 y permitirían de todas formas observar un cambio de expresión de un 60% con 4 copias. Estos valores de pendiente

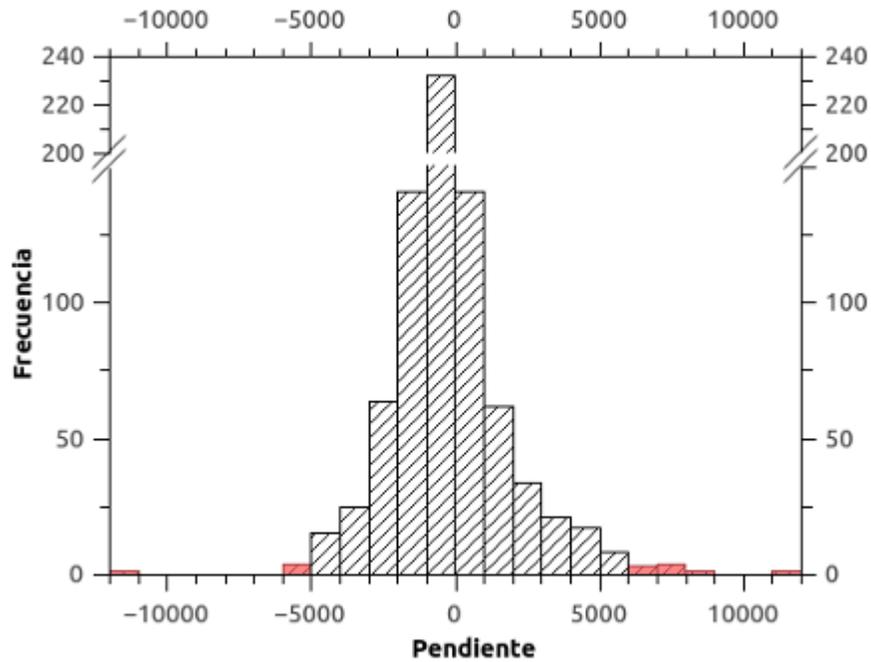
corresponderían entonces al 9,3% menor respecto a las 3.904 regresiones totales y a un 11,7% menor respecto a las 34 regresiones que pasaron ambos filtros. Los histogramas con la distribución de los valores de pendiente se ven en la Figura 5, mientras que los patrones y sus valores de coeficiente  $r^2$  y pendiente asociados se muestran en la Tabla 4.

**Tabla 4. Patrones dobles de codones seleccionados.** La columna “Kim” corresponde al número de genes de los 102 publicados por Kim *et al.* que poseen al menos una copia del patrón, mientras que la columna “Genoma” es la cantidad de genes en todo el genoma de *S. Enteritidis* que cumple con la misma condición. La selección incluye los 4 patrones de mayor pendiente y los 4 patrones de menor pendiente.

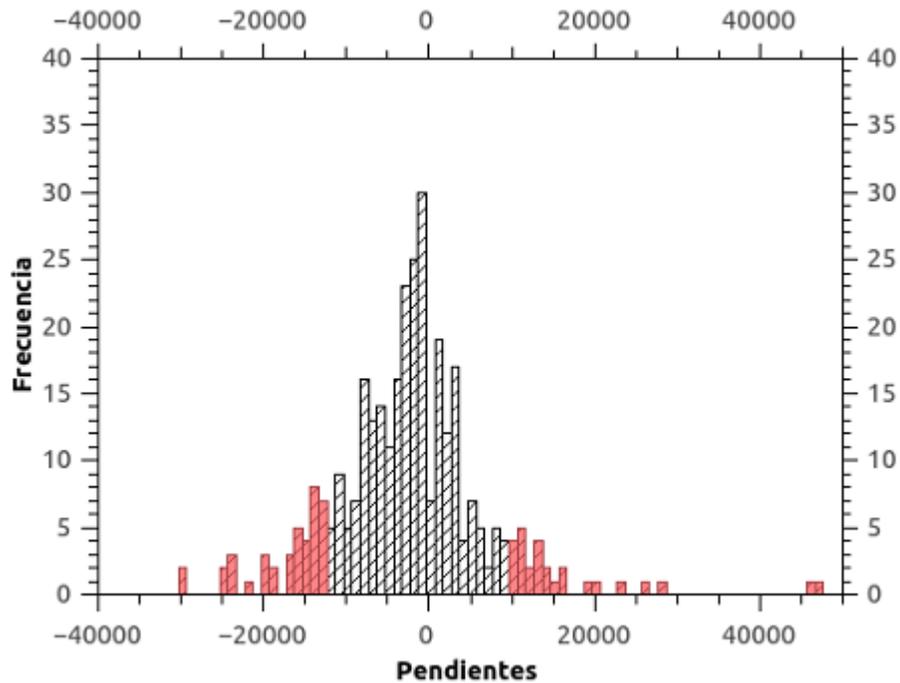
Código	Patrón	Aminoácidos	Pendiente	$r^2$	Kim	Genoma
2p1	AAC TTT	Asn-Phe	11.716	0,24	17	460
2p2	GTC GTC	Val-Val	8.923	0,31	16	549
2p3	GTT GCG	Val-Ala	7.861	0,37	20	526
2p4	GGT ACC	Gly-Thr	8.923	0,56	16	345
2p5	GAC CAG	Asp-Gln	-11.603	0,24	18	471
2p6	CTG GTT	Leu-Val	-5.997	0,36	32	987
2p7	GTT ATC	Val-Ile	-5.668	0,28	22	754
2p8	TAT CAG	Tyr-Gln	-5.361	0,23	18	878

En el caso de las 238.144 regresiones de patrones triples, 236.088 de ellas o un 99,14% no tienen suficientes puntos para generar una regresión válida. Esto en comparación al 52,58% observado a nivel de genoma completo que no podría ser analizado por las mismas razones. Luego, de las 2.056 regresiones disponibles, sólo 95 tenían valor de coeficiente  $r$  mayor a su correspondiente valor crítico y sólo 24 de esas tenían más de 5 puntos. Finalmente, se

A)



B)



**Figura 5. Pendientes de regresiones obtenidas.** Los valores de pendiente para las regresiones lineales de patrones dobles (A) y de patrones triples (B). Las secciones en rojo representan las regresiones cuyos valores de pendiente pasan el filtro.

seleccionaron 2 regresiones de pendiente positiva en conjunto con 3 regresiones de pendiente negativa. Esto debido a que sólo hay 2 regresiones con pendiente positiva en el grupo final y ambas permiten un aumento hipotético de hasta 120% de expresión en un gen de largo promedio (317 codones) con 4 copias del patrón en su secuencia. En la Tabla 5 se detallan los patrones triples seleccionados.

**Tabla 5. Patrones triples de codones seleccionados.** La columna “Kim” corresponde al número de genes de los 102 publicados por Kim *et al.* que poseen al menos una copia del patrón, mientras que la columna “Genoma” es la cantidad de genes en todo el genoma de *S. Enteritidis* que cumple con la misma condición. La selección incluye los 2 patrones de mayor pendiente y los 3 patrones de menor pendiente.

Código	Patrón	Aminoácidos	Pendiente	r <sup>2</sup>	Kim	Genoma
3p9	GAA GTG ACC	Glu-Val-Thr	23.981	0,88	5	38
3p10	CTG GGT AAA	Leu-Gly-Lys	12.854	0,74	7	69
3p11	AAA CCG CTG	Lys-Pro-Leu	-28.855	0,83	5	76
3p12	CTG GAT AAC	Leu-Asp-Asn	-23.214	0,89	5	93
3p13	GCG CCG CTG	Ala-Pro-Leu	-21.391	0,92	5	183

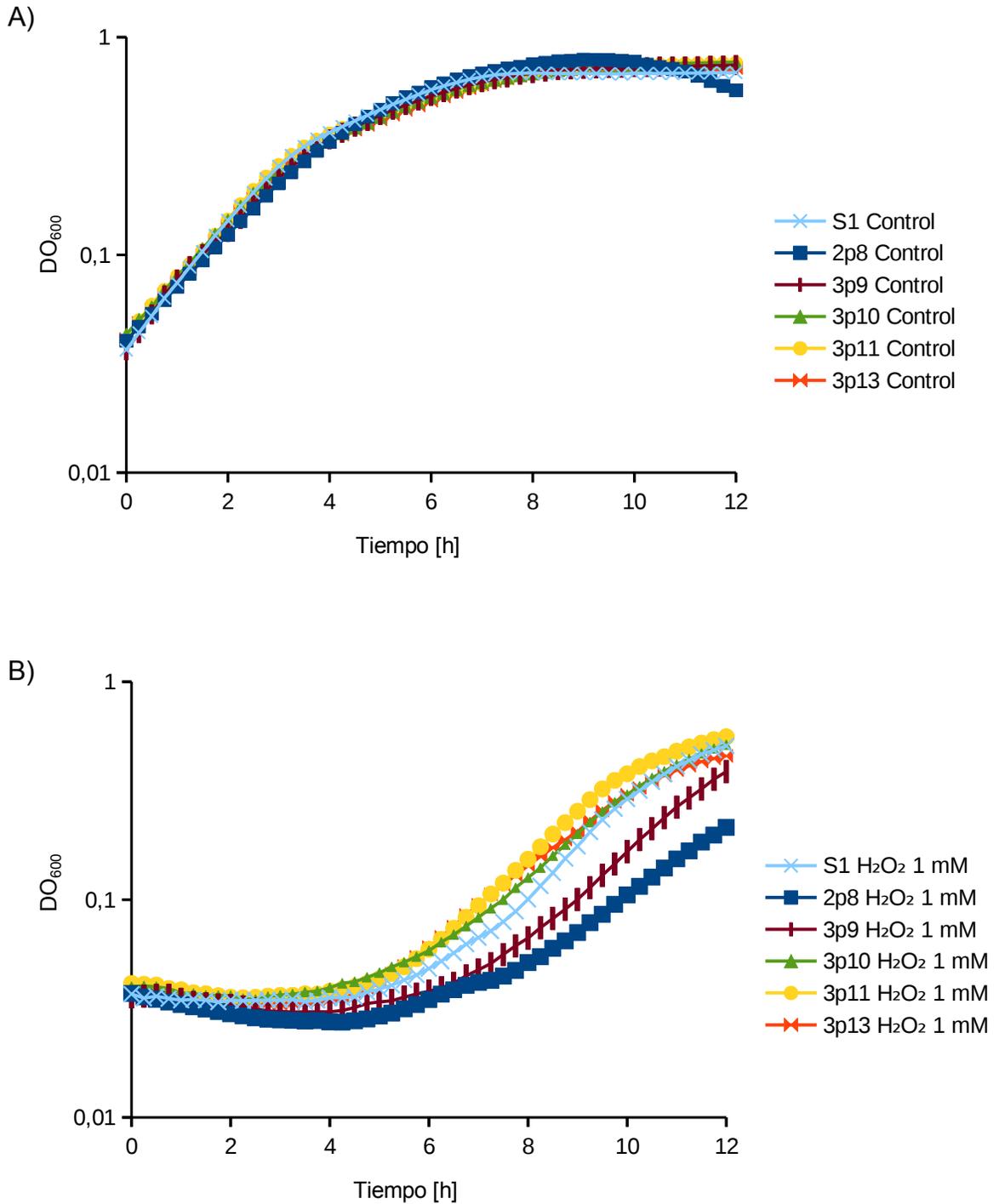
## 5.2.- Cambio en niveles de traducción para un gen reportero *in vivo*

Para confirmar experimentalmente el papel de estos patrones en la regulación de la traducción, se insertaron 4 repeticiones de cada patrón entre el cuarto y noveno codón del gen que codifica sfGFP. Si bien es cierto que en *S. Enteritidis* no se encuentran de forma natural 4 copias consecutivas de estos patrones en un gen, los valores de pendiente calculados para los patrones seleccionados implican un cambio de expresión muy pequeño para ser medido fácilmente, y es

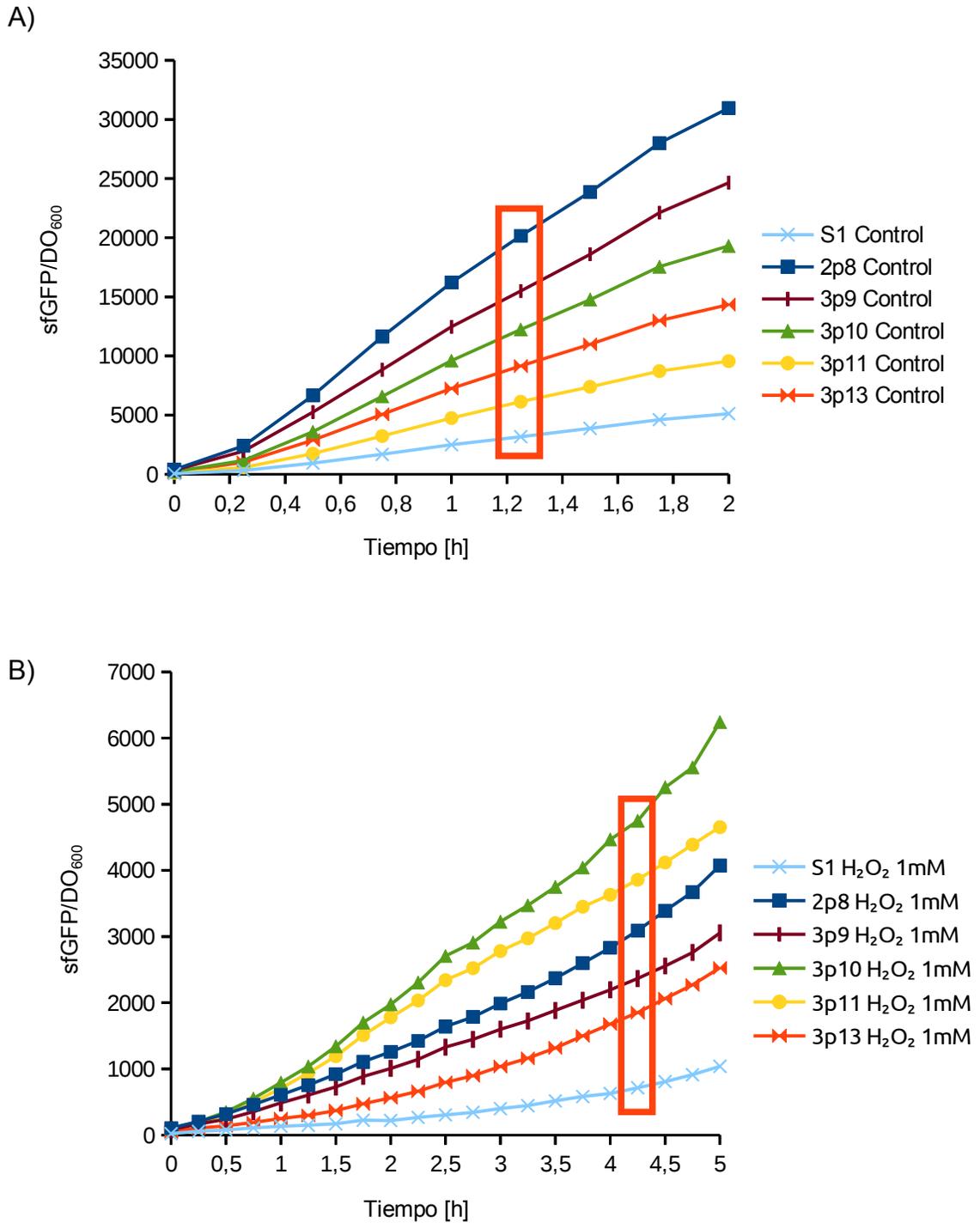
mejor amplificar el efecto de ellos utilizando más de una copia a la vez. De esta forma, se pueden cuantificar los cambios en la síntesis de sfGFP a través de la medición de la razón de fluorescencia sfGFP/mCherry. Al estar mCherry en fusión transcripcional con sfGFP, la razón de fluorescencias entre ambas proteínas permite normalizar los cambios en la traducción de sfGFP por los niveles de transcrito y por los cambios globales en el inicio de la traducción. Esta razón de fluorescencia se analiza entonces en condiciones con y sin inducción de estrés oxidativo, comparando también los valores equivalentes en una cepa con el plásmido psfGFP-mCherry sin inserto, denominada cepa S1. Es importante destacar que las mediciones se analizaron en condiciones iniciales de expresión donde aún existe un aumento lineal en el tiempo de la razón entre la fluorescencia de sfGFP o mCherry respecto a la  $DO_{600}$ , ya que pasado ese punto es necesario considerar no sólo la síntesis *de novo* de proteínas fluorescentes, sino también su cinética de degradación. En el experimento realizado, este rango de tiempo fue entre las 0,5 y 2 horas para la condición control y entre las 2,5 y 5 horas bajo estrés oxidativo. Esto debido a que existe un retardo importante en el comienzo de la aparición de fluorescencia de mCherry en la condición de estrés, lo cual probablemente se debe a una menor velocidad de plegamiento de mCherry comparada con GFP. En este último caso, también es importante realizar la medición antes de que la bacteria salga del estado de estrés y vuelva a replicarse con normalidad (pasadas las 4,5 horas en nuestras condiciones experimentales).

El análisis de fluorescencia de sfGFP se realizó con 5 de los patrones seleccionados y no con los 13 pensados originalmente debido a problemas al momento de construir los respectivos plásmidos. Finalmente se analizaron 4 de los 5 patrones triples (3p9, 3p10, 3p11 y 3p13) y un único patrón doble de codones (2p8). Las curvas de crecimiento de todas las cepas son similares en condición control (Figura 6A), pero se observa un retraso en el comienzo de la etapa de crecimiento exponencial para la condición de estrés oxidativo (Figura 6B). Específicamente, frente a un estímulo con H<sub>2</sub>O<sub>2</sub> 1 mM, las cepas comienzan su crecimiento pasadas las 4,5 horas. Aún así, durante estas 4,5 horas se observó la expresión de sfGFP y mCherry (Figuras 7B y 8B), lo que implica que aún en condiciones de no replicación, la transcripción y traducción de *gfp* es activa.

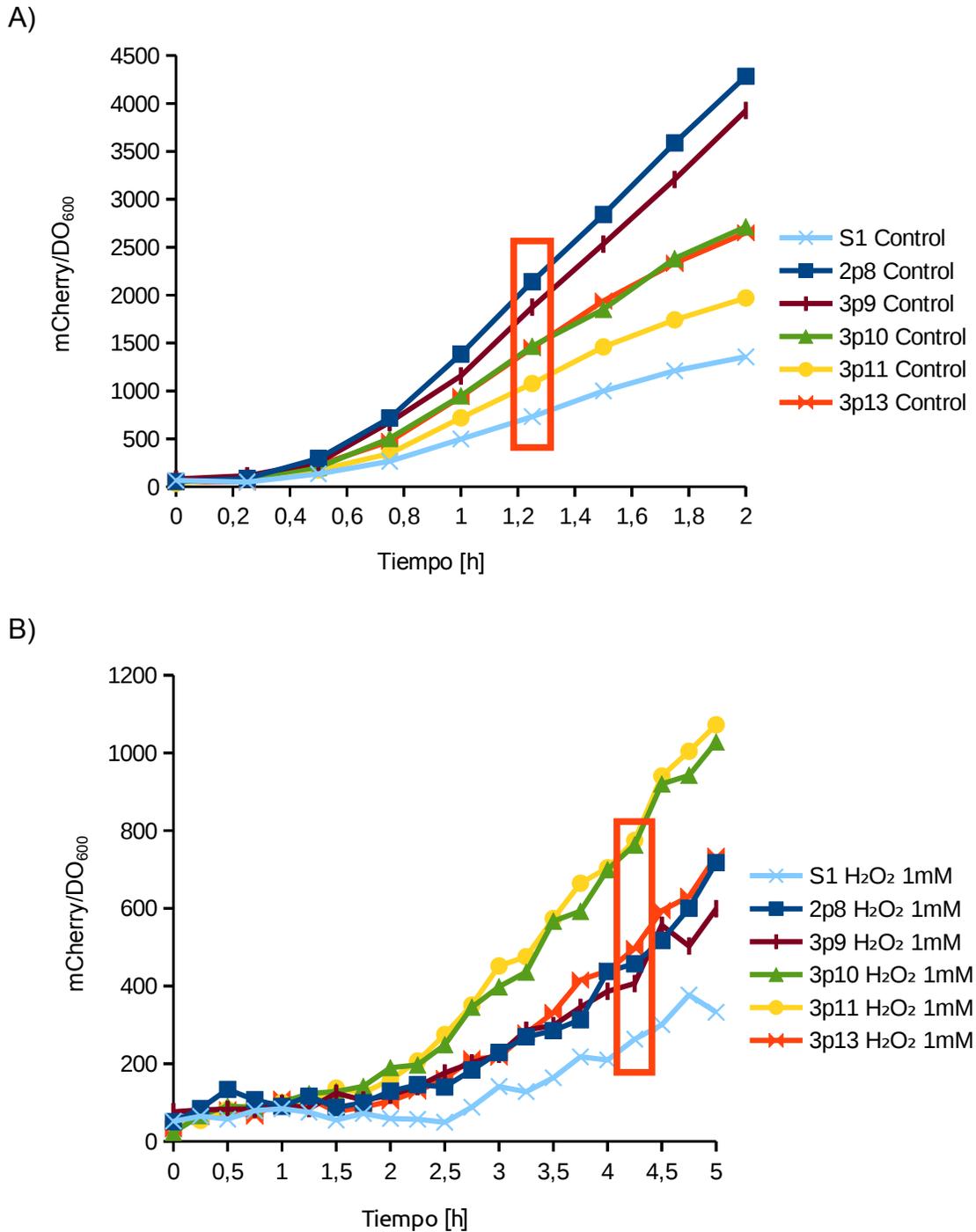
En condiciones control, sfGFP mantiene un aumento lineal de razón fluorescencia/DO<sub>600</sub> respecto al tiempo desde las 0,5 hasta aproximadamente las 2 horas de incubación (Figura 7A). Este rango es más amplio para las cepas incubadas con H<sub>2</sub>O<sub>2</sub> 1 mM, llegando hasta las 5 horas sin perder su linealidad (Figura 7B). En el caso de mCherry, la cepa control no comienza su aumento lineal hasta aproximadamente los 45 minutos de incubación, el cual persiste hasta las 2 horas (Figura 8A) y, en el caso de las cepas con H<sub>2</sub>O<sub>2</sub> 1 mM, este aumento lineal se ve a partir de las 2,5 hasta las 5 horas (Figura 8B). Cabe mencionar que la fluorescencia de mCherry bajo estrés es mucho menor que la de sfGFP, debido en parte a la incapacidad del equipo de medir en todo el



**Figura 6. Curva de crecimiento bajo estrés con  $H_2O_2$ .** Se observa que el crecimiento de *S. Enteritidis* con  $H_2O_2$  1 mM (B) se atrasa aproximadamente 4,5 horas respecto a la condición control (A). S1 corresponde a la cepa con el plásmido sfGFP-mCherry sin inserto. Las barras de desviación estándar se omitieron con el fin de facilitar el análisis.



**Figura 7. Cambios de síntesis de sfGFP respecto al crecimiento.** El aumento de la razón GFP/DO<sub>600</sub> en el tiempo es lineal entre los 30 minutos y dos horas de incubación para S. Enteritidis en condición sin estrés (A) y entre las 1 y 5 horas con H<sub>2</sub>O<sub>2</sub> 1 mM (B). S1 corresponde a la cepa con el plásmido sfGFP-mCherry sin inserto. Las barras de desviación estándar se omitieron con el fin de facilitar el análisis. Los rectángulos rojos marcan los puntos usados en el análisis final.

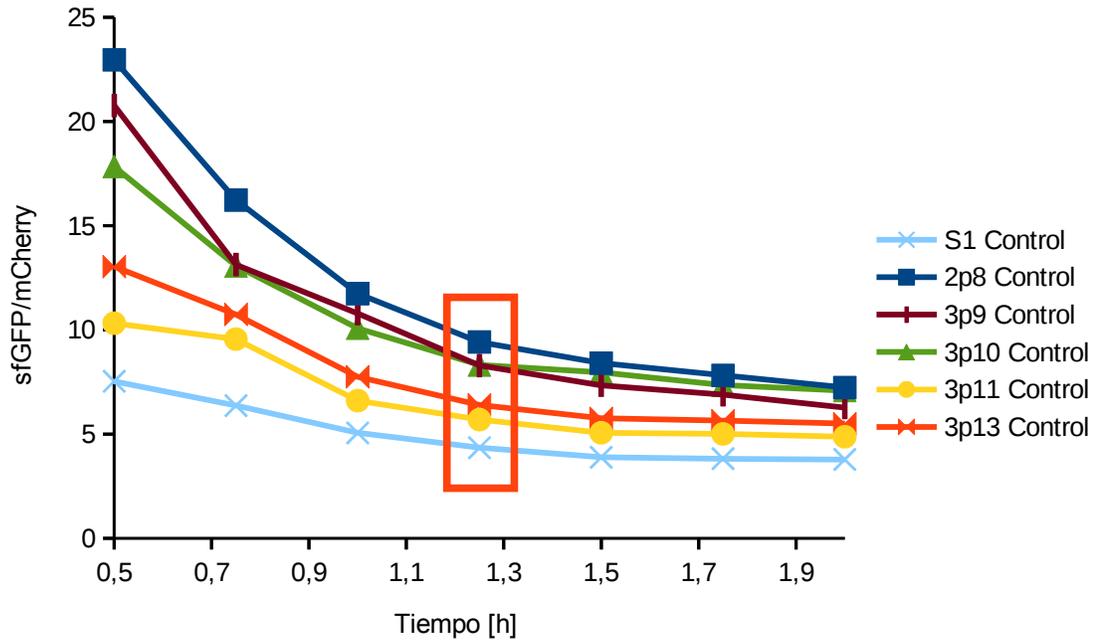


**Figura 8. Cambios de síntesis de mCherry respecto al crecimiento.** El aumento de la razón mCherry/DO<sub>600</sub> en el tiempo es lineal entre los 30 minutos y las dos horas para *S. Enteritidis* en condición sin estrés (A) y entre las 2,5 y 5 horas en medio con H<sub>2</sub>O<sub>2</sub> 1 mM (B). S1 corresponde a la cepa con el plásmido sfGFP-mCherry sin inserto. Las barras de desviación estándar se omitieron con el fin de facilitar el análisis. Los rectángulos rojos marcan los puntos usados en el análisis final.

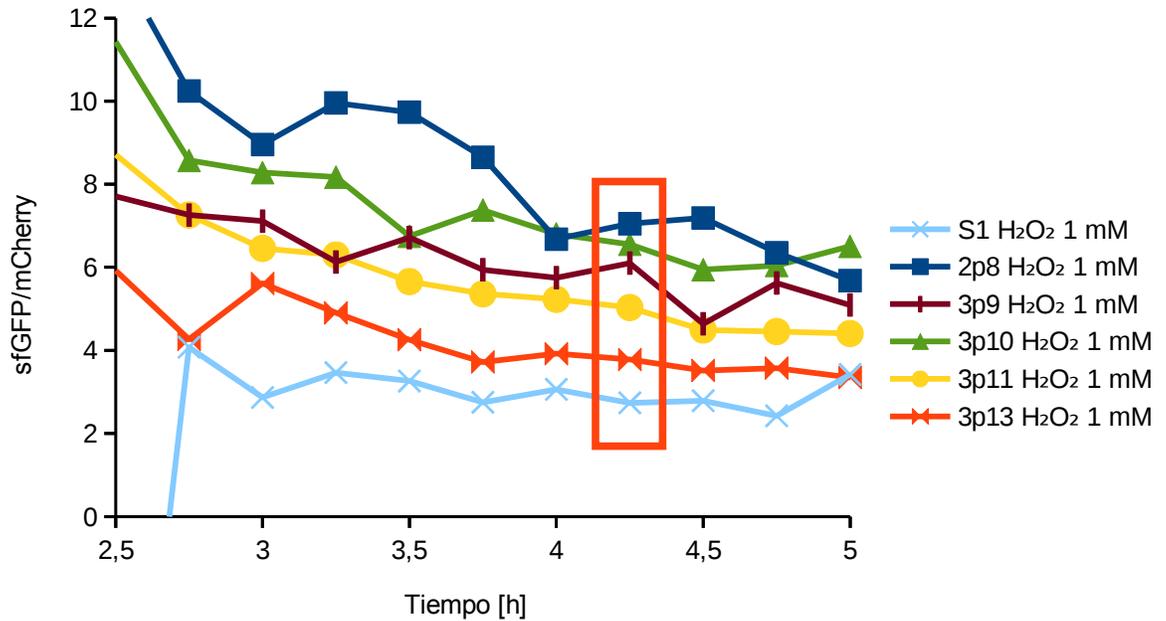
espectro de emisión de mCherry. A su vez, fue necesario eliminar todos los datos de medición previo a los 30 minutos de incubación debido a la gran cantidad de ruido en esos puntos para todas las cepas, lo cual se asocia a los bajos niveles de fluorescencia observados. Probablemente, parte de esta baja fluorescencia inicial se debe al tiempo que demoran la transcripción del gen y la traducción y plegamiento de los productos génicos. Además, bajo estrés oxidativo, no se consideraron los puntos medidos después de las 4,5 horas, cuando la bacteria empieza a replicarse, ya que a pesar de mantenerse dentro del rango lineal de las producciones de sfGFP y mCherry normalizadas por las  $DO_{600}$ , el estado fisiológico de la bacteria ha cambiado y no se puede esperar el mismo comportamiento de su maquinaria de traducción.

En las condiciones discutidas, la razón sfGFP/mCherry refleja la síntesis de la proteína sfGFP independientemente de los niveles de mRNA. Al analizar sólo el rango de tiempo donde la razón sfGFP o mCherry sobre  $DO_{600}$  aumenta linealmente (entre las 0,5 y 2 horas de incubación para la condición control y entre las 2,5 y 5 horas para la condición de estrés), se observa que tanto en condición control como en estrés, la razón sfGFP/mCherry es asintótica con una disminución inicial que tiende a un valor fijo (Figura 9). Esto es esperable para cualquier división de dos rectas con pendiente positiva e intercepto distinto de cero donde los valores de la primera recta son mayores a la segunda. En este caso, el valor fijo al que tiende la relación sfGFP/mCherry corresponde a la razón entre las pendientes de sfGFP y mCherry en el tiempo. Es importante

A)



B)



**Figura 9. Cambios de síntesis de sfGFP en dos condiciones.** En el rango lineal de aumento de la razón  $GFP/DO_{600}$  y  $mCherry/DO_{600}$ , se observan comportamientos similares de razón de fluorescencia sfGFP/mCherry a través del tiempo de incubación en las condiciones control (A) y estrés con  $H_2O_2$  1 mM (B). Las barras de desviación estándar se omitieron con el fin de facilitar el análisis. Los rectángulos rojos muestran la hora de incubación utilizada para calcular la razón estrés/control que permitió comparar niveles de traducción entre cepas.

notar que el rango de valor fijo comienza al final del rango lineal del aumento de sfGFP/DO<sub>600</sub> o mCherry/DO<sub>600</sub> para ambas condiciones.

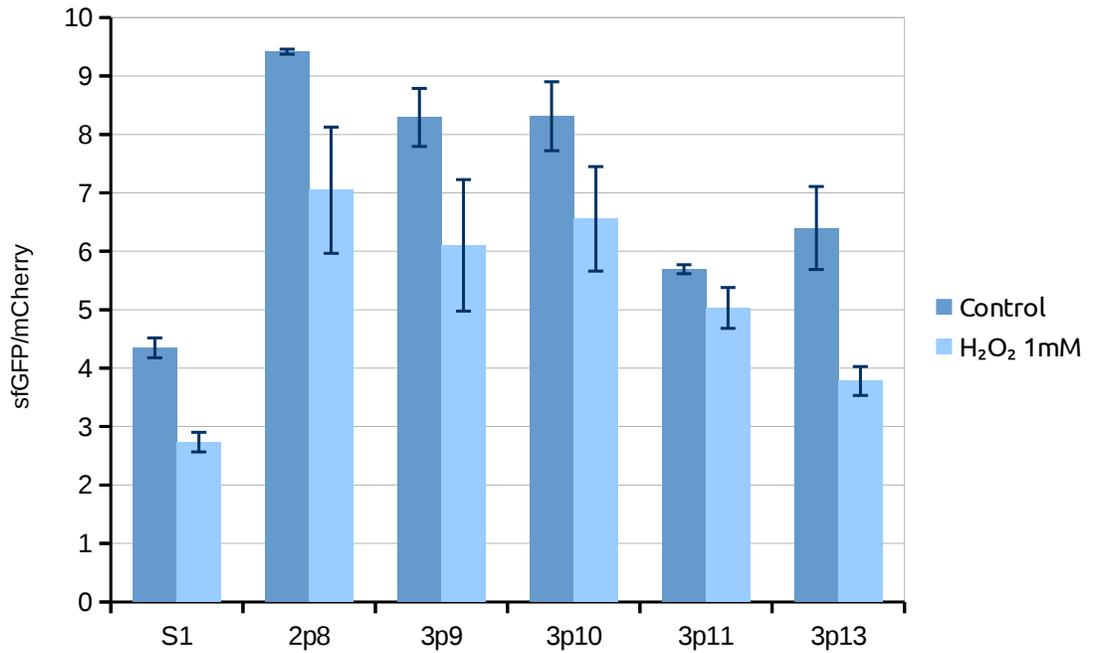
Normalmente sería posible comparar la traducción de sfGFP entre una condición control y una de estrés a través del análisis de los cambios en la razón  $(\text{sfGFP}/\text{mCherry})_{\text{estrés}}/(\text{sfGFP}/\text{mCherry})_{\text{control}}$  en el tiempo. Sin embargo, la expresión de mCherry en las células incubadas en presencia de H<sub>2</sub>O<sub>2</sub> se observa en el rango de tiempo donde la expresión de sfGFP ya no es lineal para la condición control, lo que hace imposible comparar los cambios en el tiempo de la razón  $(\text{sfGFP}/\text{mCherry})_{\text{estrés}}/(\text{sfGFP}/\text{mCherry})_{\text{control}}$ . Es por esto que finalmente se tomó un único punto representativo de la razón sfGFP/mCherry para cada condición dentro del rango de aumento lineal fluorescencia/DO<sub>600</sub> y se analizó directamente la razón entre ambos. Los puntos escogidos fueron 1,25 horas para la condición control y 4,25 horas para la condición de estrés. Estos corresponden a los puntos de mayor fluorescencia y menor desviación estándar dentro del rango de aumento lineal tanto de sfGFP/DO<sub>600</sub> como para mCherry/DO<sub>600</sub> y corresponden a la etapa de crecimiento exponencial en la condición control y previo a la recuperación del crecimiento en estrés. En estos tiempos la razón sfGFP/mCherry se ha estabilizado y, por lo tanto, se espera que representen fielmente la razón entre la velocidad de síntesis de sfGFP y mCherry, es decir, la velocidad de traducción de sfGFP normalizada por la cantidad de mRNA en cada condición. Además, como punto de referencia se compararon estos valores a los observados en la cepa S1, que lleva el plásmido sin inserto en la secuencia de *gfp*.

En primer lugar, es importante mencionar que los valores de la razón sfGFP/mCherry son siempre mayores en la condición control respecto a la condición de estrés. Esto implica que la síntesis de sfGFP disminuye más de lo que disminuye mCherry frente al estrés y esta condición se mantuvo igual en todas las cepas. A su vez, el análisis permitió encontrar dos patrones cuya inclusión en *gfp* evitó una fuerte disminución en la síntesis de sfGFP, es decir, aumentó los niveles relativos de su traducción bajo estrés (Tabla 6 y Figura 10). Estos fueron los patrones triples 3p10 (CTGGGTAAA) y 3p11 (AAACCGCTG). Si bien ambos dieron pendientes positivas, en el análisis de regresiones sólo el patrón 3p10 dio pendiente positiva mientras que 3p11 dio pendiente negativa. Esta contradicción entre la predicción informática y el resultado experimental será analizada en la discusión. Es importante destacar que no todos los patrones produjeron cambios en la expresión de sfGFP bajo estrés oxidativo. En particular, el efecto del estrés oxidativo sobre los niveles de traducción de 3p13 es muy similar al que tiene sobre la cepa control S1. Esto confirma que los cambios observados no corresponden a un cambio inespecífico producido por la introducción de codones adicionales a la secuencia de sfGFP.

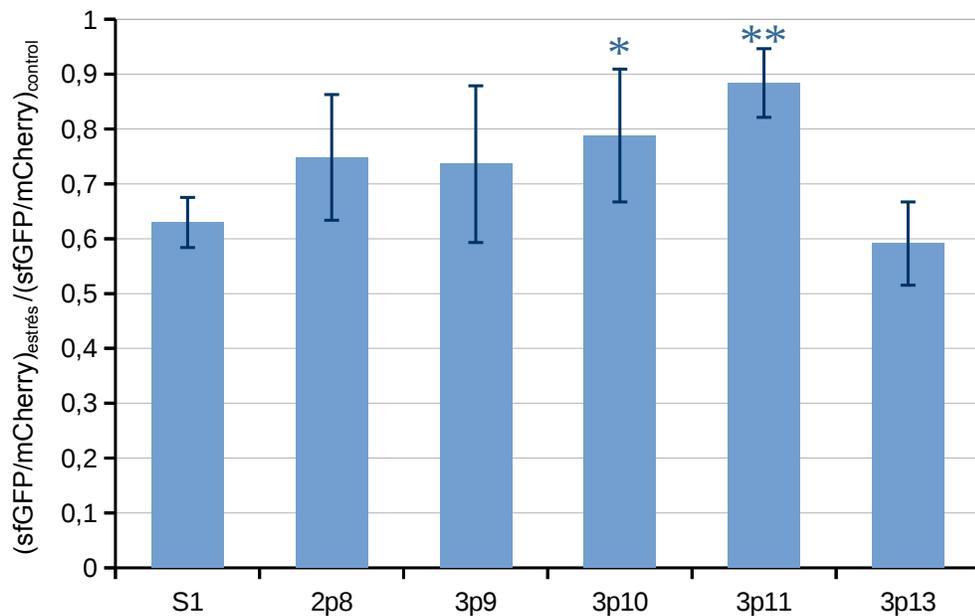
### **5.3.- Funciones génicas en común**

Basados en los resultados expuestos, decidimos analizar en mayor profundidad el posible papel fisiológico del patrón 3p10 (CTGGGTAAA), que es el único de los patrones que se comporta de forma similar en las predicciones y en el experimento. Para ello se creó una lista con el 50% de los genes de mayor frecuencia de uso del patrón 3p10 en el genoma respecto al tamaño del gen. Es

A)



B)



**Figura 10. Cambio en velocidad de traducción inducido por estrés.** A) Se muestran los valores de la razón entre las fluorescencias de sfGFP y mCherry de cada cepa en condición de estrés por H<sub>2</sub>O<sub>2</sub> 1 mM a las 4,25 horas de incubación y en la condición control a las 1,25 horas. B) Se muestran las razones entre los valores mostrados en A) para los 5 insertos de patrones de codones analizados. En B), cada valor se compara a la cepa S1 que incluye el mismo plásmido sin inserto. El símbolo \* indica un valor p menor a 0,05 y \*\* un valor p menor a 0,01 para un test t de Student.

decir, los genes que tienen un uso mayor a la mediana de la distribución dentro de todos los genes que tienen al menos una copia del patrón. A partir de esta lista fue posible realizar un análisis de enriquecimiento de funciones para determinar cuáles son las posibles funciones génicas afectadas al aumentar la eficiencia traduccional de este patrón. Los resultados de este análisis se muestran en la Tabla 7.

**Tabla 6. Valores de razón de fluorescencia en tiempos seleccionados.** “Razón fluorescencia control” corresponde a los valores de la razón de fluorescencia sfGFP/mCherry para la condición control a las 1,25 horas de incubación. “Razón fluorescencia estrés” corresponde a la misma razón de fluorescencia en condiciones de estrés inducido por H<sub>2</sub>O<sub>2</sub> 1mM a las 4,25 horas de incubación. “Razón entre ambas condiciones” es la razón entre las razones de fluorescencia entre la condición estresada y control. “Razón normalizada por S1” equivale a la razón entre cada valor de “Razón entre ambas condiciones” con el mismo valor en la cepa S1.

Cepa	Razón fluorescencia control	Razón fluorescencia estrés	Razón entre ambas condiciones	Razón normalizada por S1
S1	4,346 ± 0,17	2,736 ± 0,17	0,629 ± 0,04	1
2p8	9,418 ± 0,04	7,046 ± 1,08	0,748 ± 0,11	1.189 ± 0,20
3p9	8,290 ± 0,50	6,102 ± 1,13	0,736 ± 0,14	1.170 ± 0,24
3p10	8,314 ± 0,59	6,554 ± 0,89	0,788 ± 0,12	1.253 ± 0,21
3p11	5,693 ± 0,08	5,032 ± 0,35	0,884 ± 0,06	1.405 ± 0,14
3p13	6,399 ± 0,71	3,781 ± 0,25	0,591 ± 0,07	0.939 ± 0,14

**Tabla 7. Análisis de enriquecimiento de funciones en genes con CTGGGTAAA.** La columna “Genes vía” corresponde a la cantidad de genes dentro de la vía metabólica específica que contienen un conteo normalizado por tamaño de CTGGGTAAA mayor a la frecuencia de uso promedio de este patrón. La sección “Función estrés” se refiere a la posible utilidad de aumentar la función determinada en condiciones de estrés oxidativo.

Función	Valor p	Genes	Genes vía	Función estrés
Gluconeogénesis	$5,30 \cdot 10^{-4}$	<i>mdh, eno, pgk</i>	3 de 14	Recuperar función
Glicólisis	0,012	<i>eno, pgk</i>	2 de 14	Recuperar función
Síntesis de alanina	0,013	<i>sufS</i>	1 de 1	Recuperar función
Degradación de asparragina	0,013	<i>ansA</i>	1 de 1	Activar el ciclo de Krebs
Degradación de fosfonatos	0,038	<i>phnX</i>	1 de 3	Activar el ciclo de Krebs, síntesis de Leu, Cys y Arg
Síntesis de centros hierro-azufre	0,050	<i>sufS</i>	1 de 3	Deshacer el daño oxidativo
Síntesis de serina	0,050	<i>serB</i>	1 de 4	Síntesis de tiazol, síntesis de purinas
Fermentación ácido mixta	0,06	<i>mdh, ackA</i>	2 de 33	Respiración anaeróbica
Síntesis de tiazol	0,06	<i>sufS</i>	1 de 4	Recuperar la síntesis de tiamina
Ciclo de la urea	0,06	<i>argG</i>	1 de 5	N/A

## 6.- DISCUSIÓN

En esta memoria se utilizó un método bioinformático basado en la regresión lineal entre cambios de niveles de proteínas bajo estrés oxidativo y el conteo de codones de sus genes correspondientes con el fin de analizar el efecto del uso de codones en la regulación de la respuesta contra el estrés oxidativo en bacterias. Este análisis, en conjunto con la medición *in vivo* de los niveles de traducción en un gen reportero *gfp*, permitieron encontrar dos patrones triples de codones capaces de aumentar la síntesis de proteínas específicas en condiciones de estrés oxidativo en la cepa *Salmonella enterica* serovar Enteritidis P125109.

El aumento de traducción de *gfp* dependiente de H<sub>2</sub>O<sub>2</sub> en la cepa con el patrón CTGGGTAAA (3p10), que codifica la secuencia Leu-Gly-Lys de aminoácidos, es aproximadamente 25% mayor que el aumento observado en una cepa S1 con un plásmido sin el patrón de codones insertado en su secuencia (Tabla 6). Este aumento es considerablemente menor al 160% esperado según el método de selección utilizado (o 200% si se consideran los 257 codones del gen *gfp* con los insertos en su secuencia), pero es necesario considerar los efectos de la transcripción y la degradación de proteínas que también afectan el valor de pendiente calculado. Un punto de referencia importante a considerar es el patrón GCGCCGCTG (3p13), el cual a pesar de ser del mismo largo que el resto de los patrones triples, presenta una traducción considerablemente menos

sensible al estrés oxidativo. La razón de fluorescencia estrés/control de 3p13 es más cercana a la observada en S1, lo cual implica que la sensibilidad de la traducción frente al estrés oxidativo observada en las cepas con inserto no se debe a una mayor longitud de los mensajeros.

Se realizó luego un análisis de enriquecimiento de función y se observó que los genes del genoma de *S. Enteritidis* que poseen CTGGGTAAA en su secuencia están involucrados en 1) la síntesis de centros hierro-azufre, los cuales se oxidan en condiciones de estrés (Farr *et al.*, 1986), 2) la formación de ácido corísmico y la posterior transformación al ácido isocorísmico que permite sintetizar enterobactina, la cual se ha evidenciado protege a *Salmonella enterica* serovar Typhimurium contra el estrés por H<sub>2</sub>O<sub>2</sub> (Achard *et al.*, 2013), y finalmente 3) la síntesis de tiamina o vitamina B1, la cual se vuelve necesaria para mejorar la supervivencia de *Salmonella* al inducir estrés oxidativo con paraquat (Thorgersen y Downs, 2009). En conjunto, esto sugiere que la mayor eficiencia de traducción de CTGGGTAAA podría tener un efecto en la supervivencia de la cepa bajo estrés oxidativo.

### **6.1- Idoneidad de la base de datos**

El trabajo se basó en los datos cuantitativos de cambio de niveles de proteínas observados por Kim *et al.* (2010) al inducir estrés oxidativo en una cepa clínica de *S. Enteritidis* SE2472 mediante la adición de H<sub>2</sub>O<sub>2</sub> 5 mM. Este conjunto de datos incluye cambios en los niveles de 102 proteínas y permitió realizar un

análisis de regresión para 2.561 de los 3.904 posibles patrones dobles de codones. Sin embargo, si la base de datos incluyese todos los genes del genoma, sería posible analizar hasta 3.728 patrones a la vez, lo que implica un 29,9% de regresiones perdidas debido únicamente a un menor tamaño muestral de los datos disponibles. En el caso de los patrones triples se analizaron 2.056 de 238.144 patrones posibles, perdiéndose un 46,56% de las regresiones posibles a nivel de genoma completo. Esto sin contar todos los patrones extra que no se pueden analizar debido a su baja frecuencia a nivel de genoma, los cuales de todas formas se eliminaron de la selección final. Es necesario también considerar que sólo 7 de los 102 genes de la base de datos disminuían su expresión frente a estrés oxidativo, por lo que en general no era esperable encontrar regresiones de pendiente negativa. Esto posiblemente se debe a un sesgo de los autores del trabajo al seleccionar las manchas de la electroforesis que fueron analizadas por espectrometría de masas. La selección de las manchas de mayor intensidad pudo haber afectado la calidad de nuestro análisis y explicar la selección de patrones relacionados a pendientes negativas que finalmente aumentan la traducción bajo condiciones de estrés.

Aún con estas limitaciones, utilizamos la base de datos de Kim *et al.* (2010), ya que fue la única disponible al momento de realizar este análisis, que incluía valores cuantitativos de cambios en niveles de proteínas bajo estrés oxidativo en bacterias. Sin embargo, es importante considerar que una forma de haber obtenido resultados más precisos hubiese sido utilizar una base de datos con

más genes, la cual además incluyese una mayor proporción de proteínas cuyos niveles disminuyan con el estrés. Esto se volvería aún más preciso si se utilizaran a su vez datos de cambios de mRNA bajo estrés, ya sea por secuenciación de RNA (RNAseq) o ensayos por microarreglo para normalizar los cambios de expresión y asegurar que los niveles de proteínas que cambian sean por cambios en la traducción y no en la transcripción. Sin embargo, por el momento no existen bases de datos de *Salmonella* u otras bacterias con este tipo de información en condiciones de crecimiento con y sin H<sub>2</sub>O<sub>2</sub>.

Es importante considerar que algunas de las diferencias observadas entre los efectos de los patrones sobre la traducción en nuestros experimentos y en el análisis de regresión pueden deberse a las diferencias entre los medios de cultivo tipo LB utilizados por Kim *et al.* (2010) y el medio mínimo M9 que utilizamos nosotros. El uso de su grupo de un medio rico en aminoácidos pudo haber alterado la traducción de los patrones, pero debido a que nuestro método utiliza la medición de fluorescencia, no es recomendable utilizar medio LB debido a su elevada fluorescencia basal.

## **6.2.- Conteo de patrones de codones**

El análisis de uso de codones se realizó finalmente a través del conteo de patrones dobles y triples de codones, ignorando completamente el conteo de codones debido a su baja validez predictiva dadas las regresiones obtenidas. Sin embargo, el uso de conteo de patrones dobles y triples de codones sí tiene

una utilidad en el análisis experimental de uso de codones, ya que el efecto que tenga cualquier sustrato sobre ácidos nucleicos, sea proteína o incluso ROS, depende del contexto espacial del codón dentro de la secuencia. Por ejemplo, la participación en o cercanía a estructuras secundarias en el mRNA (Chamary y Hurst, 2005), la conformación de blancos específicos para RNA pequeños o incluso la afinidad entre moléculas de tRNA localizadas en los sitios A, P y E del ribosoma durante el proceso de traducción (Bossi y Roth 1980) son todas posibles razones por las cuales la traducción de un codón puede cambiar según el contexto donde se encuentre. Es por esto que analizar el entorno directo de un codón a través del conteo de patrones dobles y triples implica un análisis más profundo en cuanto a su afinidad como blanco y a su efecto sobre la velocidad de traducción.

### **6.3.- Velocidad de traducción por razón de fluorescencia**

Con el fin de comprobar experimentalmente la validez de los patrones escogidos a través de la comparación de regresiones lineales, se insertaron todos los patrones en la región 5' de un gen *gfp* en fusión transcripcional con el gen de mCherry. De los 5 patrones analizados, la razón de fluorescencia sfGFP/mCherry comparando condición control y estrés por H<sub>2</sub>O<sub>2</sub> 1 mM aumentó significativamente para 2 de ellos, lo cual concuerda con la hipótesis, ya que existe un cambio en los niveles de traducción dependiente de condiciones de estrés con H<sub>2</sub>O<sub>2</sub>. Sin embargo, de los 2 patrones observados, sólo CTGGGTAAA tenía pendiente positiva según el método anterior, lo que sugiere

que es el único patrón analizado que mostró un resultado acorde con los datos obtenidos previamente.

A su vez, debido a la significancia del resultado obtenido para el patrón 3p11, en el cual se observó un aumento significativo y mayor de la síntesis de sfGFP, se considera necesario hacer más pruebas a CTGGGTAAA y probar su verdadero efecto sobre la traducción en condiciones de estrés. Por ejemplo, es posible tomar uno de los genes con CTGGGTAAA, cuya expresión haya aumentado de mayor forma según Kim *et al.* (2010), como lo fueron *phnX* o *mdh*, mutar su única copia de CTGGGTAAA por codones sinónimos y ver si existe algún efecto sobre los niveles de la proteína generada bajo estrés. Otra opción sería realizar un perfil de ribosomas (*ribosome profiling*), el cual consiste en extraer complejos mRNA-ribosomas de una célula e inmovilizarlos con el fin de digerir y eliminar el mRNA no protegido por ribosomas (Ingolia, 2014). Esto permite identificar zonas de pausa en el proceso de traducción, ya que los fragmentos asociados a las zonas lentas estarán en mayor proporción respecto a las rápidas. Si se utiliza esta técnica en genes con CTGGGTAAA, se podría determinar si realmente se observa un aumento de velocidad relativa de traducción específicamente en la zona donde se encuentra el patrón en condiciones de estrés oxidativo.

#### **6.4.- Efectos a nivel de metabolismo**

Se ha observado que la respuesta a estrés oxidativo se refleja en diversos

cambios a nivel de metabolismo en bacterias. Por ejemplo, inhibiendo la replicación celular, disminuyendo los niveles totales de traducción (Zhong *et al.*, 2015), promoviendo la síntesis de proteínas de reparación del daño oxidativo (Imlay, 2013) y reduciendo la producción de NADH en favor de la producción de NADPH (Rui *et al.*, 2010). Sin embargo, los resultados que se obtuvieron analizando las funciones génicas de los genes con CTGGGTAAA en su secuencia no son todos acordes a esta respuesta. Específicamente se observaron muchos genes asociados a la vía glicolítica, al ciclo de Krebs y a la cadena transportadora de electrones, las cuales deberían inactivarse en una respuesta normal a estrés oxidativo, debido a su generación acoplada de NADH e incluso de ROS en el caso de la cadena transportadora de electrones. Sin embargo, debido al rápido agotamiento del H<sub>2</sub>O<sub>2</sub> del medio extracelular en condiciones normales (Hébrard *et al.*, 2009), no se puede descartar la posibilidad de que la respuesta asociada a la activación de genes con CTGGGTAAA a las 4,25 horas sea una respuesta compensatoria en vez de una respuesta de defensa. Es decir, que el fin de la respuesta no sea aumentar la supervivencia, sino preparar a la bacteria para volver a replicarse luego de haber eliminado el H<sub>2</sub>O<sub>2</sub>. Es por esto que los resultados del análisis de enriquecimiento pueden no ser un reflejo muy preciso de los efectos reales del estrés oxidativo sobre la expresión de genes. Sin embargo, aún vale la pena medir los efectos observados sobre las otras funciones enriquecidas como la producción de aminoácidos aromáticos o de la vitamina B1 con el fin de determinar el efecto real del patrón bajo condiciones de estrés tanto sobre la supervivencia como de la adaptabilidad posterior al estrés.

## **7.- CONCLUSIÓN**

A partir de lo observado en esta memoria, se puede sugerir un mecanismo en el cual la presencia de los patrones de codones CTGGGTAAA y AAACCGCTG en genes de *S. Enteritidis* permite la traducción más eficiente de sus mensajeros, lo cual genera una respuesta más rápida frente al estrés oxidativo inducido por H<sub>2</sub>O<sub>2</sub>. También sugiere en el caso de CTGGGTAAA que un cambio en la velocidad de elongación en el proceso de traducción inducido por estrés podría ser capaz de generar un aumento medible en los niveles de proteínas. La bacteria podría utilizar entonces este tipo de respuesta traduccional para adaptarse a otras condiciones ambientales y la estrategia de búsqueda que hemos desarrollado podría permitir el estudio de estos procesos.

## 8.- BIBLIOGRAFÍA

Achard ME, Chen KW, Sweet MJ, Watts RE, Schroder K, Schembri MA y McEwan AG. An antioxidant role for catecholate siderophores in *Salmonella*. The Biochemical Journal 454(3): 543-549 (2013)

Aitken CE, Petrov A y Puglisi JD. Single ribosome dynamics and the mechanism of translation. Annual Review of Biophysics 39: 491-513 (2010)

Bali V y Bebok Z. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. The International Journal of Biochemistry and Cell Biology 64: 58-74 (2015)

Betancor L, Yin L, Martínez A, Fookes M, Sasias S, Schelotto F, Thomson N, Maskell D, y Chabalgoity JA. Genomic comparison of the closely related *Salmonella enterica* serovar enteritidis and Dublin. The Open Microbiology Journal 6: 5-13 (2012)

Birnboim HC y Doly J. A rapid alkaline lysis procedure for screening recombinant plasmid DNA. Nucleic Acids Research 7(6): 1513-1522 (1979)

Bossi L y Roth JR. The influence of codon context on genetic code translation. Nature 286(5769): 123-127 (1980)

Carlioz A y Touati D. Isolation of superoxide dismutase mutants in *Escherichia coli*: Is superoxide dismutase necessary for aerobic life? The EMBO Journal 5(3): 623-630 (1986)

Chamary JV y Hurst LD. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biology 6(9): R75 (2005)

Chan CT, Pang YL, Deng W, Babu IR, Dyavaiah M, Begley TJ y Dedon PC. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nature Communications 3: 937 (2012)

Ditmar KA, Sørensen MA, Elf J, Ehremberg M y Pan T. Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Reports 6(2): 151-157 (2005)

dos Reis, Wernisch L y Savva R. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. Nucleic Acids Research 31(23): 6976-6985 (2003)

Elgamal S, Artsimovitch I y Ibba M. Maintenance of transcription-translation coupling by elongation factor P. mBio 7(5): e01373-16 (2016)

Elgamal S, Katz A, Hersch SJ, Newson D, White P, Navarre WW e Ibba M. EF-P dependant pauses integrate proximal and distal signals during translation. PloS Genetics 10: e1004553 (2014).

Farr SB, D'Ari R y Touati D. Oxygen-dependent mutagenesis in *Escherichia coli* lacking superoxide dismutase. Proceedings of the National Academy of Sciences of the United States of America 83(21): 8268-8272 (1986)

- Farrugia G y Balzan R. Oxidative stress and programmed cell death in yeast. *Frontiers in Oncology* 2(64): 1-21 (2012)
- Gambino M y Cappitelli F. Mini-review: Biofilm responses to oxidative stress. *Biofouling* 32(2): 167-178 (2016)
- Gu M e Imlay JA. The *soxRS* response of *Escherichia coli* is directly activated by redox-cycling drugs rather than by superoxide. *Molecular Microbiology* 79(5): 1136-1150 (2011)
- Gutman GA y Hatfield GW. Nonrandom utilization of codon pairs in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 86(10): 3699-3703 (1989)
- Hawkins CL y Davies MJ. Generation and propagation of radical reactions on proteins. *Biochimica et Biophysica Acta* 1504(2-3): 196-219 (2001)
- Hébrard M, Viala JP, Méresse S, Barras F y Aussel L. Redundant peroxide scavengers contribute to *Salmonella* virulence and oxidative stress resistance. *Journal of Bacteriology* 191(14): 4605-4614 (2009)
- Henriques MX, Catalão MJ, Figueiredo J, Gomes JP y Filipe SR. Construction of improved tools for protein localization studies in *Streptococcus pneumoniae*. *PloS ONE* 8: e55049 (2013)
- Hersch SJ, Wang M, Zou SB, Moon K-M, Foster LJ, Ibba M y Navarre WW. Divergent protein motifs direct elongation factor P-mediated translational regulation in *Salmonella enterica* and *Escherichia coli*. *MBio* 4: e00180-13 (2013).
- Hidalgo E, Ding H y Dimple B. Redox signal transduction: mutations shifting [2Fe-2S] centers of the SoxR sensor-regulator to the oxidized form. *Cell* 88(1): 121-129 (1997)
- Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* 2(1): 13-34 (1985)
- Imlay JA. Pathways of oxidative damage. *Annual Review of Microbiology* 57: 395-418 (2003)
- Imlay JA. Cellular defenses against superoxide and hydrogen peroxide. *Annual Review of Biochemistry* 77: 755-776 (2008)
- Imlay JA. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nature Reviews Microbiology* 11(7): 443-454 (2013)
- Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics* 15(3): 205-213 (2014)
- Irwin B, Heck JD y Hatfield GW. Codon pair utilization biases influence translational elongation step times. *The Journal of Biological Chemistry* 270(39): 22801-22806 (1995)

- Katz A y Orellana O. Protein synthesis and the stress response. En: Cell-free protein synthesis. Biyani M (Ed.). Londres: InTech. DOI: 10.5772/50311. Disponible en: <https://www.intechopen.com/books/cell-free-protein-synthesis/protein-synthesis-and-the-stress-response> (2012)
- Kim K, Yang E, Vu G, Gong H, Su J, Liu F y Lu S. Mass spectrometry-based quantitative proteomic analysis of *Salmonella enterica* serovar Enteritidis protein expression upon exposure to hydrogen peroxide. BMC Microbiology 10: 166 (2010)
- Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV y Gottesman MM. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 315(5811): 525-528 (2007)
- Rui B, Shen T, Zhou H, Liu J, Chen J, Pan X, Liu H, Wu J, Zheng H y Shi Y. A systematic investigation of *Escherichia coli* central carbon metabolism in response to superoxide stress. BMC Systems Biology 4: 122 (2010)
- Sharp PM y Li WH. The Codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research 15(3): 1281-1295 (1987)
- Subramaniam AR, Pan T y Cluzel P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. Proceedings of the National Academy of Sciences of the United States of America 110(6): 2419-2424 (2013)
- Subramaniam AR, Deloughery A, Bradshaw N, Chen Y, O'Shea E, Losick R y Chai Y. A serine sensor for multicellularity in a bacterium. eLife 2: e01501 (2013)
- Thorgersen MP y Downs DM. Oxidative stress and disruption of labile iron generate specific auxotrophic requirements in *Salmonella enterica*. Microbiology 155(Pt 1): 295-304 (2009)
- Tian B, Xu Z, Sun Z, Lin J y Hua Y. Evaluation of the antioxidant effects of carotenoids from *Deinococcus radiodurans* through targeted mutagenesis, chemiluminescence, and DNA damage analyses. Biochimica et Biophysica Acta 1770(6): 902-911 (2007)
- Woese CR, Hinengardner RT y Engelberg J. Universality in the genetic code. Science 144(3621): 1030-1031 (1964)
- Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS y Liu Y. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Molecular Cell 59(5): 744-754 (2015)
- Zhan K, Narasimhan J y Wek RC. Differential activation of eIF2 kinases in response to cellular stresses in *Schizosaccharomyces pombe*. Genetics 168(4): 1867-1875 (2004)
- Zheng M, Aslund F y Storz G. Activation of the OxyR transcription factor by reversible disulfide bond formation. Science 279(5357): 1718-2171 (1998)
- Zhong J, Xiao C, Gu W, Du G, Sun X, He Q-Y y Zhang G. Transfer RNAs mediate the rapid adaptation of *Escherichia coli* to oxidative stress. PLOS Genetics 11(6): e1005302 (2015)

## 9.- ANEXOS

### 9.1.- contador123.pl (código)

```
#!/bin/perl
use warnings;
use strict;
use Bio::SeqIO;
my $to = time;
my $out = "Conteo codones total2.txt";
my $out2 = "Conteo patrones dobles total2.txt";
my $out3 = "Conteo patrones triples total2.txt";
my (@gennames,@protid,@single,@double,@triple,@match);
open IN, "<","P125109genbank.fasta" or die $!;
open IN2, "<","patronesdobles.txt" or die $!;
open IN3, "<","patronestriples2.txt" or die $!;
open IN4, "<","codones.txt" or die $!;
open OUT, ">$out" or die $!;
open OUT2, ">$out2" or die $!;
open OUT3, ">$out3" or die $!;
while (<IN>){
    if ($_ =~ /[gene=(\S+)/]){
        push @gennames, $1;
    }
    if ($_ =~ /protein_id=(\w+\d+\.\d+)/){
        push @protid, $1;
    }
}
while (<IN2>){
    $_ =~ /(\w+)/ and push @double, $1;
}
while (<IN3>){
    $_ =~ /(\w+)/ and push @triple, $1;
}
while (<IN4>){
    $_ =~ /(\w+)/ and push @single, $1;
}
my $seqio_obj = Bio::SeqIO->new(-file => "P125109genbank.fasta", -format => "fasta");
my (@genes,@maxs);
while ( my $seq_obj = $seqio_obj-> next_seq ) {
    push @genes, $seq_obj-> seq;
}
my ($max,$max2,$max3,$index);
my %genhash = ();
my %genhash2 = ();
my %genhash3 = ();
my %gent = ();
my %single =();
my %double = ();
for (0...$#genes) {
    $genhash{$_} = [];
    $genhash2{$_} = [];
    $genhash3{$_} = [];
```

```

$index = $_;
$max = (length $genes[$index])/3;
$max2 = $max-1;
$max3 = $max2-1;
push @maxs, $max;
for (0...$max) {
    my $codon = substr($genes[$index],3*$_,3);
    push @{$genhash{$index}}, $codon;
}
pop @{$genhash{$index}};
for (0...$#single){
    push @{$genhash{$index}}, $single[$_];
    $single{$_} = 0;
}
for (0...$max2) {
    my $codon = substr($genes[$index],3*$_,6);
    push @{$genhash2{$index}}, $codon;
}
pop @{$genhash2{$index}};
for (0...$#double){
    push @{$genhash2{$index}}, $double[$_];
    $double{$_} = 0;
}
for (0...$max3) {
    my $codon = substr($genes[$index],3*$_,9);
    push @{$genhash3{$index}}, $codon;
}
pop @{$genhash3{$index}};
for (0...$#triple){
    $gent{$_} = 0;
}
}
print OUT "Código\tgen";
print OUT2 "Código\tgen";
print OUT3 "Código\tgen";
for (@single) {
    print OUT "\t",$_;
}
for (@double) {
    print OUT2 "\t",$_;
}
for (@triple) {
    print OUT3 "\t",$_;
}
print OUT "\n";
print OUT2 "\n";
print OUT3 "\n";
for (0...$#genes) {
    print OUT $protid[$_], "\t",$gennames[$_];
    print OUT2 $protid[$_], "\t",$gennames[$_];
    print OUT3 $protid[$_], "\t",$gennames[$_];
    my ($stemp, $count) = ("@{$genhash{$_}}", 0);
    ($count = ~ s/($_)//g) and printf OUT "\t%d",($count - 1) for sort @{$genhash{$_}};
    my ($stemp2, $count2) = ("@{$genhash2{$_}}", 0);

```

```

    ($count2 = $temp2 =~ s/($_)//g) and printf OUT2 "\t%d",($count2 - 1) for sort
    @{$genhash2{$_}};
    my %copy = %gent;
    my @sort = sort @{$genhash3{$_}};
    my ($temp3, $count3) = ("@{$genhash3{$_}}", 0);
    for (@sort) {
        if ($count3 = $temp3 =~ s/($_)//g) {
            $copy{$_} = $count3;
        }
    }
    foreach my $key (sort(keys %copy)) {
        print OUT3 "\t", $copy{$key};
    }
    print OUT "\n";
    print OUT2 "\n";
    print OUT3 "\n";
}
my $out4 = "Conteo codones normalizados2.txt";
my $out5 = "Conteo patrones dobles normalizados2.txt";
my $out6 = "Conteo patrones triples normalizados2.txt";
open OUT4, ">$out4" or die $!;
open OUT5, ">$out5" or die $!;
open OUT6, ">$out6" or die $!;
print OUT4 "Código\tgen";
print OUT5 "Código\tgen";
print OUT6 "Código\tgen";
for (@single) {
    print OUT4 "\t",$_;
}
for (@double) {
    print OUT5 "\t",$_;
}
for (@triple) {
    print OUT6 "\t",$_;
}
print OUT4 "\n";
print OUT5 "\n";
print OUT6 "\n";
for (0...$#genes) {
    print OUT4 $protid[$_], "\t", $gennames[$_];
    print OUT5 $protid[$_], "\t", $gennames[$_];
    print OUT6 $protid[$_], "\t", $gennames[$_];
    my $norm = $maxs[$_];
    my ($temp, $count) = ("@{$genhash{$_}}", 0);
    ($count = $temp =~ s/($_)//g) and printf OUT4 "\t%.6f",(($count - 1)/$norm) for sort
    @{$genhash{$_}};
    my ($temp2, $count2) = ("@{$genhash2{$_}}", 0);
    ($count2 = $temp2 =~ s/($_)//g) and printf OUT5 "\t%.6f",(($count2 - 1)/($norm - 1)) for sort
    @{$genhash2{$_}};
    my %copy = %gent;
    my @sort = sort @{$genhash3{$_}};
    my ($temp3, $count3) = ("@{$genhash3{$_}}", 0);
    for (@sort) {
        if ($count3 = $temp3 =~ s/($_)//g) {

```

```

        $copy{$_} = $count3;
    }
}
foreach my $key (sort(keys %copy)) {
    printf OUT6 "\t%.6f", $copy{$key}/($norm - 2);
}
print OUT4 "\n";
print OUT5 "\n";
print OUT6 "\n";
}
my $ttot = time - $to;
print STDOUT "tiempo total = $ttot seg o ". ($ttot/60). " min\n";

```

## 9.2.- regresion.pl (código)

```

#!/bin/perl
use warnings;
use strict;
use Statistics::Regression;
my $to = time;
my $in1 = $ARGV[0];
my $out = "parametros para $in1.txt";
my $out2 = "parametros para $in1 sin cero.txt";
my (@genes,@cambios,@valores, $peso);
open IN, "<","cambiospaper.txt" or die $!;
open IN2, "<","Conteo $in1.txt" or die $!;
open OUT, ">$out" or die $!;
open OUT2, ">$out2" or die $!;
select OUT;

while (<IN>){
    chomp;
    my @linea = split /\t/, $_;
    push @genes, $linea[1];
    if ($_ =~ /\t(\S+) ± (\d+)/){
        push @valores, $1;
        push @cambios, $2;
    }
    else{
        push @valores, 0;
        push @cambios, 0;
    }
}
print "Codón\tPendiente\trsq\n";
print OUT2 "Codón\tPendiente\trsq\tadjrsq\tObservaciones\tIntercepto\n";
while (<IN2>){
    next if $_ =~ /Código/;
    next if $_ =~ /^gen/;
    my @linea = split /\t/, $_;
    print $linea[0], "\t";
}

```

```

print OUT2 $linea[0], "\t";
shift @linea;
my $reg = Statistics::Regression->new("salmonella",["Intercept", "Slope"]);
my $regSC = Statistics::Regression->new("salmonella2",["Intercept", "Slope"]);
for (0...$#linea){
    if ($cambios[$_] != 0){
        $peso = 1/($cambios[$_]**2);
    }
    else{
        $peso = 1;
    }
    $reg->include($valores[$_],[1.0,$linea[$_]],$peso);
    $regSC->include($valores[$_],[1.0,$linea[$_]],$peso) unless $linea[$_] == 0;
}
my @theta = $reg->theta;
my $rsq = $reg->rsq;
my @thetaSC = (0,0);
my $regn = $regSC->n;
my $rsqSC = 0;
my $adjrsqSC = 0;
if ($regn > 2){
    my $sst = $regSC->sst;
    if ($sst != 0){
        $rsqSC = $regSC->rsq;
        @thetaSC = $regSC->theta;
        $adjrsqSC = $regSC->adjrsq;
    }
}
print $theta[1], "\t", $rsq, "\ty = ", $theta[1], "x + ", $theta[0], "\n";
print OUT2 $thetaSC[1], "\t", $rsqSC, "\t", $adjrsqSC, "\t", $regn, "\t", $thetaSC[0], "\n";
}
my $ttot = time - $to;
print STDOUT "tiempo total = $ttot seg o ", ($ttot/60), " min\n";

```

### 9.3.- psfGFP-mCherry

El diseño del plásmido pBAD30SFIT (psfGFP-mCherry) está basado en el plásmido pBAD30XS (Elgamal *et al.*, 2014). Contiene un módulo de fusión transcripcional compuesto por la proteína fluorescente verde de plegamiento rápido sfGFP y la proteína fluorescente roja mCherry modificada a itag-mCherry (Henriques *et al.*, 2013). Este módulo incluye a su vez un sitio de clonamiento XhoI-SpeI en el cuarto codón de sfGFP. Para construir pBAD30SFIT, pBAD30 se digirió con EcoR1 y KpnI, lo cual permitió introducir los sitios de restricción BglII y PciI a través de un segmento de DNA de doble hebra compuesto por los partidores EKfwd y EKrev (Anexo 9.3.1) ligados entre sí. El gen de itag-mCherry se amplificó de un plásmido pBAD30HA (Elgamal *et al.*, 2016)

utilizando los partidores de PCR itagFwd e itagRev (Anexo 9.3.1). Este gen se introdujo entre los sitios PciI y KpnI. El gen de sfGFP con el sitio de clonamiento XhoI-SpeI en el cuarto codón se amplificó por PCR del plásmido pXG10sf-LacZ186 (Hersch *et al.*, 2013) utilizando los partidores sfFwd y sfRev (Anexo 9.3.1) y luego se insertó entre los sitios EcoR1 y BglII.

### **9.3.1 - Partidores para la construcción de pBAD30SFIT:**

EKFwd AATTCAGGACACAGATCTCCTAACACATGTTAGATGATTCCGGGGATCCGGGTAC

EKRev CCGGATCCCCGGAATCATCTAACATGTGTTAGGAGATCTGTGTCCTG

itagFwd CCTAACACATGTGAAAGGAGGA

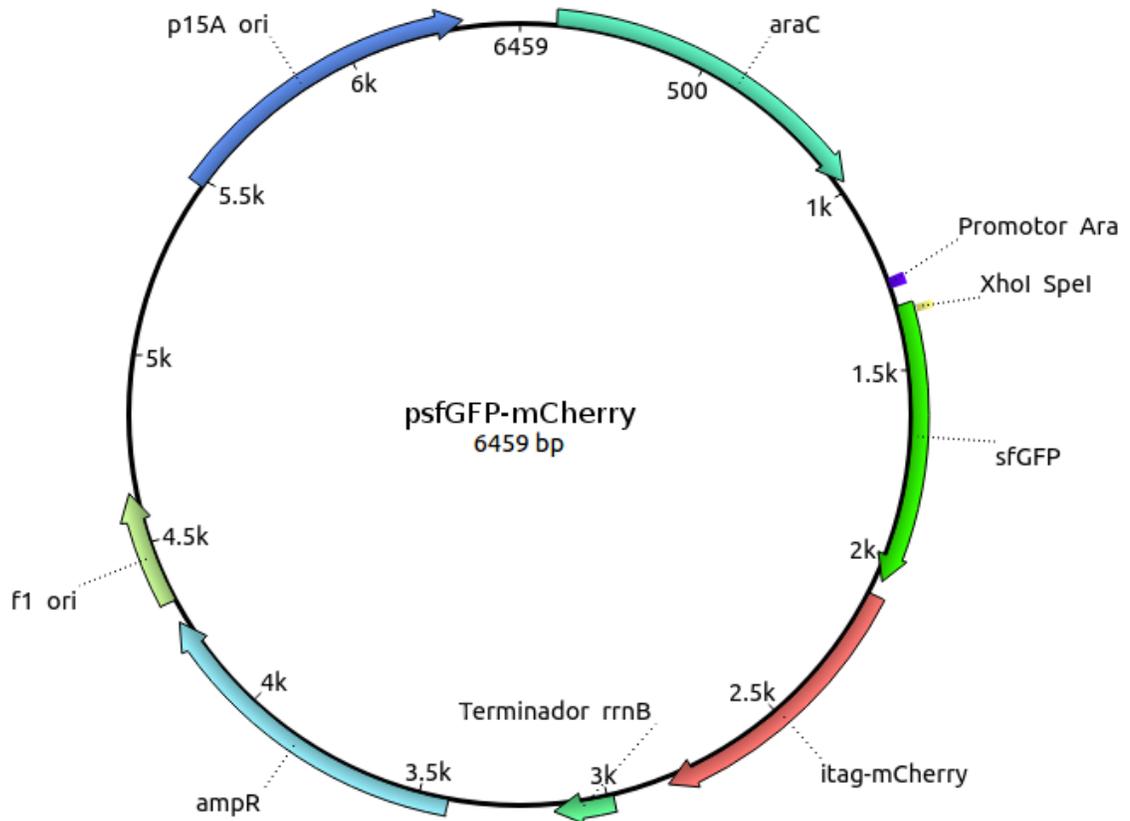
itagRev ATCCCCGGGTACCCGGATCCCC

sfFwd

GCTAGCGAATTCAGGAGGAATTTACCATGAGCAAACTCGAGGGCAACACTAGTGGAGAAGAACT  
TTTCACTGGAG

sfRev GTTAGGAGATCTGTGTTATTTGTAGAGCTCATCCATGCCATGTGTAATCCCAGCAGCAG

### 9.3.2 – Sitios relevantes de *psfGFP-mCherry*



*araC*: 96-974 (hebra complementaria)

Promotor Ara: 1248-1276

*sfgfp* : 1326-2060

Sitio XhoI-SpeI: 1335-1352

*itag-mCherry* : 2100-2837 (mCherry parte en 2154 con triplete AUG)

Terminador *rrnB*: 2985-3142

*ampR*: 3421-2481

f1 ori: 4340-4639

p15A ori: 5484-6311

### 9.3.3 - Secuencia

>psGFP-mCherry [1 6459]

```
1   ATCGATGCAT AATGTGCCTG TCAAATGGAC GAAGCAGGGA TTCTGCAAAC CCTATGCTAC
61  TCCGTCAAGC CGTCAATTGT CTGATTCGTT ACCAATTATG ACAACTTGAC GGCTACATCA
121 TTCACTTTTT CTTACAACC GGCACGGAAC TCGCTCGGGC TGGCCCCGGT GCATTTTTTA
181 AATACCCGCG AGAAATAGAG TTGATCGTCA AAACCAACAT TGCGACCGAC GGTGGCGATA
241 GGCATCCGGG TGGTGTCTAA AAGCAGCTTC GCCTGGCTGA TACGTTGGTC CTCGCGCCAG
301 CTTAAGACGC TAATCCCTAA CTGCTGGCGG AAAAGATGTG ACAGACGCGA CGGCACAAG
361 CAAACATGCT GTGCGACGCT GGCGATATCA AAATTGCTGT CTGCCAGGTG ATCGCTGATG
421 TACTGACAAG CCTCGCGTAC CCGATTATCC ATCGGTGGAT GGAGCGACTC GTTAATCGCT
481 TCCATCGGCC GCAGTAACAA TTGCTCAAGC AGATTTATCG CCAGCAGCTC CGAATAGCGC
541 CCTTCCCTT GCCCGGCGTT AATGATTTGC CCAAACAGGT CGTGAAATG CCGCTGGTGC
601 GCTTCATCCG GCGGAAAGAA CCCCCTATTG GCAAATATTG ACGGCCAGTT AAGCCATTCA
661 TGCCAGTAGG CGCGCGGACG AAAGTAAACC CACTGGTGAT ACCATTGCGG AGCCTCCGGA
721 TGACGACCGT AGTGATGAAT CTCTCCTGGC GGGAACAGCA AAATATCACC CGGTCCGCAA
781 ACAAATTCCT GTCCCTGATT TTTACCACC CCCTGACCGC GAATGGTGAG ATTGAGAATA
841 TAACCTTTCA TTCCCAGCGG TCGGTCGATA AAAAAATCGA GATAACCGTT GGCCTCAATC
901 GCGGTTAAAC CCGCCACCAG ATGGGCATTA AACGAGTATC CCGGCAGCAG GGGATCATTT
961 TGCGCTTCAG CCATACTTTT CATACTCCCG CCATTCAGAG AAGAAACCAA TTGTCCATAT
1021 TGCATCAGAC ATTGCCCGTCA CTGCGTCTTT TACTGGCTCT TCTCGCTAAC CAAACCGGTA
1081 ACCCCGCTTA TTAAGAGCAT TCTGTAACAA AGCGGGACCA AAGCCATGAC AAAAACCGCT
1141 AACAAAAGTG TCTATAATCA CGGCAGAAAA GTCCACATTG ATTATTTGCA CGCGCTCACA
1201 CTTTGTCTAT CCATAGCATT TTTATCCATA AGATTAGCGG ATCCTACCTG ACGCTTTTTA
1261 TCGCAACTCT CTACTGTTTC TCCATACCCG TTTTTTTGGG CTAGCGAATT CAGGAGGAAT
1321 TTACCATGAG CAAACTCGAG GGCAACACTA GTGGAGAAGA ACTTTTCACT GGAGTTGTCC
1381 CAATTCCTGT TGAATTAGAT GGTGATGTTA ATGGGCACAA ATTTTCTGTC CGTGGAGAGG
1441 GTGAAGGTGA TGCTACAAAC GGAAAGCTTA CCCTTAAATT TATTTGCACT ACTGAAAAC
1501 TACCTGTTC ATGGCCAACA CTTGTCACTA CTCTCACTTA TGGTGTCAA TGCTTTTCCC
1561 GTTATCCGGA TCATATGAAA CGGCATGACT TTTTCAAGAG TGCCATGCCC GAAGGTTATG
1621 TACAGGAACG CACTATATCT TTCAAAGATG ACGGGACTTA CAAGACCGGT GCTGAAGTCA
1681 AGTTTGAAGG TGATACCCCT GTTAATCGTA TCGAGTTAAA AGGTATTGAT TTTAAAGAAG
1741 ATGGAAACAT TCTCGGACAC AAACCTGATG ACAACTTTAA CTCACACAAT GTATACATCA
1801 CGGCAGACAA ACAAAGAAAT GGAATCAAAG CTAAC'TTCAA AATTTCGCAC AACGTTGAAG
1861 ATGGATCCGT TCAACTAGCA GACCATTATC AACAAAATAC TCCAATTGGC GATGGCCCTG
1921 TCCTTTTACC AGACAACCAT TACCTGTCGA CACAATCTGT CCTTTCGAAA GATCCCAACG
1981 AAAAGCGTGA CCACATGGTC CTTCTTGAGT TTGTAAGTGC TGCTGGGATT ACACATGGCA
2041 TGGATGAGCT CTACAAATA CACAGATCTC CTAACACATG TGAAAGGAGG AATTTACCAA
2101 TGCCGACATT AGAAATAGCA CAAAAAAAAG TGAGCAAGGG CGAAGAGAT AACATGGCTA
2161 TCATTAAGA GTTCATGCGC TTCAAAGTTC ACATGGAGGG TTCTGTAAAC GGTCACGAGT
2221 TCGAGATCGA AGGCGAAGGC GAGGGCCGTC CGTATGAAGG CACCAGACC GCCAAACTGA
2281 AAGTGACTAA AGGCGGCCCG CTGCCTTTTG CGTGGGACAT CCTGAGCCC CAATTTATGT
2341 ACGGTTCTAA AGCTTATGTT AAACCCGATA CGGATATCCC GGACTATCTG AAGCTGTCTT
2401 TTCCGGAAGG TTTCAAGTGG GAACGCGTAA TGAATTTTGA AGATGGTGGT GTCGTGACCG
2461 TCACTCAGGA CTCTCCCTG CAGGATGGCG AGTTCATCTA TAAAGTTAAA CTGCGTGGTA
2521 CTAATTTTCC ATCTGATGGC CCGGTGATGC AGAAGAAGAC GATGGGTGG GAGGCGTCTA
2581 GCGAACGCAT GTATCCGGA GATGGTGC GC TGAAGGCGA AATTAACAG CGCCTGAAAC
2641 TGAAAGATGG CGGCCATTAT GACGCTGAAG TGAAAACCAC GTACAAAGCC AAGAAACCTG
2701 TGCAGCTGCC TGGCGCGTAC AATGTGAATA TTAAACTGGA CATCACCTCT CATAATGAAG
2761 ATTATACGAT CGTAGAGCAA TATGAGCGCG CGGAGGGTCG TCATTCATACC GGTGGCATGG
2821 ATGAGCTCTA CAAATAATAG ATGATTCGGG GGATCCGGGT ACCCGGGGAT CCTCTAGAGT
2881 CGACCTGCAG GCATGCAAGC TTGGCTGTTT TGGCGGATGA GAGAAGATT TCAGCTGAT
2941 ACAGATTAAA TCAGAACGCA GAAGCGGTCT GATAAAACAG AATTTGCCTG CCGCGATGAG
3001 CGCGGTGGTC CCACCTGACC CCATGCCGAA CTCAGAAGTG AAACGCGCTA CCGCCGATGG
3061 TAGTGTGGGG TCTCCCCATG CGAGAGTAGG GAACTGCCAG GCATCAAATA AAACGAAAGG
3121 CTCAGTCGAA AGACTGGGCC TTTTCGTTTTA TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA
3181 GTAGGACAAA TCCGCCGGGA GCGGATTTGA ACGTTGCGAA GCAACGCCCC GGAGGGTGGC
```

3241 GGGCAGGACG CCCGCCATAA ACTGCCAGGC ATCAAATTA GCAGAAGGCC ATCCTGACGG  
3301 ATGGCCTTTT TGCGTTTCTA CAAACTCTTT TGTTTATTTT TCTAAATACA TTCAAATATG  
3361 TATCCGCTCA TGAGACAATA ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT  
3421 ATGAGTATTC AACATTTCCG TGTCGCCTTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT  
3481 GTTTTTGCTC ACCCAGAAAC GCTGGTGAAG GTAAAAGATG CTGAAGATCA GTTGGGTGCA  
3541 CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAAG TCCTTGAGAG TTTTCGCCCC  
3601 GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC  
3661 CGTGTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG  
3721 GTTGTAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA  
3781 TGCAGTGCCT CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC  
3841 GGAGGACCGA AGGAGCTAAC CGTTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT  
3901 GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA CACCACGATG  
3961 CCTGCAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTACT TACTCTAGCT  
4021 TCCGCGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC  
4081 TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAACTCG GAGCCGCTGA CCGTGGGTCT  
4141 CGCGGTATCA TTGCAGCACT GGGGCCAGAT GGTAAGCCCT CCCGTATCGT AGTTATCTAC  
4201 ACGACGGGGA GTCAGGCAAC TATGGATGAA CGAAATAGAC AGATCGCTGA GATAGGTGCC  
4261 TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT TTAGATTGAT  
4321 TTACGCGCCC TGTAGCGGCG CATTAAAGCG GCGGGTGTG GTGGTTACGC GCAGCGTGAC  
4381 CGCTACACTT GCCAGCGCCC TAGCGCCGCG TCCTTTGCTT TTCTTCCTTT CCTTCTCGC  
4441 CACGTTCCGC GGCTTTCCCC GTCAAGCTCT AAATCGGGGG CTCCCTTTAG GGTTCCGATT  
4501 TAGTGCTTTA CGGCACCTCG ACCCAAAAAA ACTTGATTTG GGTGATGGTT CACGTAGTGG  
4561 GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG GAGTCCACGT TCTTTAATAG  
4621 TGGACTTTG TTCCAAACTT GAACAACACT CAACCTATC TCGGGCTATT CTTTGTGATT  
4681 ATAAGGGATT TTGCCGATTT CGGCCATTG GTTAAAAAAT GAGCTGATTT AACAAAAATT  
4741 TAACGCGAAT TTTAACAAAA TATTAACGTT TACAATTTAA AAGGATCTAG GTGAAGATCC  
4801 TTTTTGATAA TCTCATGACC AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCAG  
4861 ACCCCGTAGA AAAGATCAAA GGATCTTCTT GAGATCCTTT TTTTCTGCGC GTAATCTGCT  
4921 GCTTGCAAAC AAAAAACCA CCGCTACCAG CGGTGGTTTG TTTGCCGAT CAAGAGCTAC  
4981 CAACTCTTTT TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA GATACCAAAT ACTGTCCTTC  
5041 TAGTGTAGCC GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT ACATACCTCG  
5101 CTCTGCTAAT CCTGTTACCA GTCAGGCATT TGAGAAGCAC ACGGTCACAC TGCTTCCGGT  
5161 AGTCAATAAA CCGGTAACC AGCAATAGAC ATAAGCGGCT ATTTAACGAC CCTGCCCTGA  
5221 ACCGACGACC GGGTCGAATT TGCTTTGCAA TTTCTGCCAT TCATCCGCTT ATTATCACTT  
5281 ATTCAGGCGT AGCACCAGGC GTTTAAGGCG ACCAATAACT GCCTTAAAAA AATTACGCCC  
5341 CGCCCTGCCA CTCATCGCAG TACTGTTGTA ATTCATTAAG CATTCTGCCG ACATGGAAGC  
5401 CATCACAGAC GGCATGATGA ACCTGAATCG CCAGCGGCAT CAGCACCTTG TCGCCTTGCG  
5461 TATAATATTT GCCGCTAGCG GAGTGTATAC TGGCTTACTA TGTTGGCACT GATGAGGGTG  
5521 TCAGTGAAGT GCTTCATGTG GCAGGAGAAA AAAGGCTGCA CCGGTGCGTC AGCAGAATAT  
5581 GTGATACAGG ATATATTCCG CTTCCCTCGCT CACTGACTCG CTACGCTCGG TCGTTCGACT  
5641 GCGGCGAGCG GAAATGGCTT ACGAACGGGG CGGAGATTTT CTGGAAGATG CCAGGAAGAT  
5701 ACTTAACAGG GAAGTGAGAG GGCCGCGGCA AAGCCGTTTT TCCATAGGCT CCGCCCCCT  
5761 GACAAGCATC ACGAAATCTG ACGCTCAAA CAGTGGTGGC GAAACCCGAC AGGACTATAA  
5821 AGATACCAGG CGTTTCCCCC TGGCGGCTCC CTCGTGCGCT CTCTGTTC TGCTTTCCG  
5881 TTTACCGGTG TCATTCCGCT GTTATGGCCG CGTTTGTCTC ATTCCACGCC TGACACTCAG  
5941 TTCCGGGTAG GCAGTTCGCT CCAAGCTGGA CTGTATGCAC GAACCCCCCG TTCAGTCCGA  
6001 CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGAAAGAC ATGCAAAAGC  
6061 ACCACTGGCA GCAGCCACTG GTAATTGATT TAGAGGAGTT AGTCTTGAAG TCATGCGCCG  
6121 GTTAAGGCTA AACTGAAAG ACAAGTTTTG GTGACTGCGC TCCTCCAAGC CAGTTACCTC  
6181 GGTTCAAAGA GTTGGTAGCT CAGAGAACCT TCGAAAAACC GCCCTGCAAG GCGGTTTTTT  
6241 CGTTTTTCTA GCAAGAGATT ACGCGCAGAC CAAAACGATC TCAAGAAGAT CATCTTATTA  
6301 ATCAGATAAA ATATTTGCTC ATGAGCCCGA AGTGGCGAGC CCGATCTTCC CCATCGGTGA  
6361 TGTCGGCGAT ATAGGCGCCA GCAACCGCAC CTGTGGCGCC GGTGATGCCG GCCACGATGC  
6421 GTCCGGCGTA GAGGATCTGC TCATGTTTGA CAGCTTATC