



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Programa de Maestría y Doctorado en Ciencias Bioquímicas

Instituto de Biotecnología

Análisis transcripcional de cepas de *E. coli* mutantes en operones ribosomales y RNA Polimerasa con mayor velocidad de crecimiento y menor adaptabilidad en condiciones de medio mínimo

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestro en Ciencias

PRESENTA:

IBT. CÉSAR ALEJANDRO MARTÍNEZ ORTÍZ

TUTOR PRINCIPAL

Dr. José Utrilla Carreri
Centro de Ciencias Genómicas, UNAM

MIEMBROS DEL COMITÉ TUTOR

Dr. Rafael Peña-Miller
Centro de Ciencias Genómicas, UNAM

Dr. Fidel Alejandro Sánchez Flores
Instituto de Biotecnología, UNAM

Cuernavaca, Morelos.

Mayo, 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

El presente trabajo fue realizado en el programa de biología de sistemas y biología sintética del centro de ciencias genómicas de la la universidad nacional autónoma de México, bajo la tutoría del Dr. José Utrilla Carreri. Durante el mismo se contó con la beca del consejo nacional de ciencia y tecnología (CONACyT) 825745. También se contó con el financiamiento del programa de apoyo a proyectos de investigación e innovación tecnológica (PAPIIT) de la UNAM, a través del proyecto IN213420 y del proyecto Newton advanced fellowship NA160328.

Índice

Agradecimientos.....	5
Abstract.....	6
Resumen.....	7
Introducción.....	8
Antecedentes.....	12
Hipótesis.....	19
Objetivo general.....	19
Objetivos particulares.....	19
Materiales y métodos.....	20
Caracterización de tasas de crecimiento.....	20
Cultivo y muestreo para la extracción de material de RNA en medio M9 con Glucosa al 0.4%.....	20
Lisis celular,extracción y preparación de RNA para la construcción de bibliotecas de RNA-Seq.....	21
Construcción de bibliotecas y secuenciación de lecturas.....	21
Procesamiento bioinformático para la obtención de genes expresados diferencialmente (GEDs).....	22
Estimación del cambio en distribución del proteoma.....	23
Análisis de interacción de GEDs con StringDb.....	26
Análisis de cambios en iModulones relevantes para la expresión de proteínas heterólogas.....	26
Resultados.....	28
Caracterización de tasas de crecimiento.....	28

Procesamiento bioinformático para la obtención de genes expresados diferencialmente (GEDs).....	28
Estimación del cambio en distribución del proteoma.....	34
Análisis de interacción de genes expresados diferencialmente con StringDb.....	53
Análisis de cambios en i-modulones relevantes para la expresión de proteínas heterólogas.....	57
Discusión.....	60
Conclusiones.....	66
Perspectivas.....	67
Referencias bibliográficas.....	68
Anexos.....	75

Agradecimientos

El presente trabajo fue desarrollado en el programa de biología sintética y biología de sistemas del centro de ciencias genómicas de la universidad nacional autónoma de México bajo la dirección del Dr. José Utrilla Carreri. Agradezco al Dr. José Utrilla Carreri por el apoyo, paciencia y asesoría que recibí de él como mi tutor principal, así como de los integrantes de mi comité tutor, el Dr. Fidel Alejandro Sánchez Flores del instituto de biotecnología de la universidad nacional autónoma de México y el Dr. Rafael Peña-Miller también del centro de ciencias genómicas de la universidad nacional autónoma de México. También agradezco por su apoyo y cooperación a mis compañeros de grupo de investigación, el MSc. David Hidalgo Vázquez, el MSc. Miguel Ángel Bello, el MSc. Gustavo Lastiri Plancardo, la Lic. Elisa Marquez y la Postdoc Leidy Bedoya. Agradezco al Dr. Ricardo Alfredo Grande Cano y la Ing. Gloria Vázquez Castro por su apoyo y colaboración para la construcción de bibliotecas de RNA-Seq que fueron parte esencial de este trabajo, así como a todos los integrantes de la unidad de secuenciación masiva del instituto de biotecnología de la universidad nacional autónoma de México. Agradezco de igual manera a Víctor del Moral y José Alfredo Hernández de la unidad de apoyo de tecnologías de información del centro de ciencias genómicas (UATI-CCG) por sus actividades de soporte y asesoría que hicieron posible realizar los análisis de datos bioinformáticos necesarios para este trabajo. Agradezco también a los trabajadores del sindicato de la universidad nacional autónoma de México que mantuvieron el espacio de trabajo que mis compañeros y yo compartimos siempre en condiciones óptimas.

Agradezco también a mis padres Alicia Jacqueline Ortiz García y César Martínez Morales, a mis hermanas Diana Lorena Martínez Ortíz y Mariana Gisela Martínez Ortíz y a mi compañera Diana Alvarado Jiménez por su apoyo incondicional durante toda la realización de este proyecto.

Por último, agradezco por el financiamiento de este proyecto de maestría al CONACyT y el PAPIIT.

Abstract

The metabolic engineering of bacterial strains is a very active area of research. A common engineering approach often involves targeting the regulation of routes to induce overexpression of an heterologous product of interest. Nevertheless, the insertion of a route and its overexpression comes at the expense of using cellular resources shared by functions of the whole system. This introduced burden has been observed to produce a wide variety of stress-related problems that negatively impact yield and protein quality, making the development of engineered strains a cumbersome process needing various steps of tuning. In this work, we further research the basis for a complementing approach focused on the reallocation of cell resources to provide a possible increase of resistance for metabolic burden and a possible increase in translational efficiency. We do so by studying the trade-off effect in resource allocation between growth and hedging functions in a ribosomal operons-deleted mutant strain of *Escherichia coli* (SQ53) as well as in a point-mutated strain at its beta RNA Polymerase subunit (rpoBE672K). Both mutants had an observed increase in growth-rate growing in minimal media M9 supplemented with glucose. Given that transcriptional data of both rpoBE672K and its control on the studied conditions were already available on NCBI, we chose only to process SQ53 and control samples in our lab. After culturing cells, we harvested them at an early stage of exponential growth phase, extracted and purified their RNA material, built RNA-Seq stranded libraries that were sequenced, quality checked, processed and analyzed to identify differentially expressed genes. This information was further combined with proteomics data, already available from a previously published report, to calculate the estimated change in proteome balance. The transcriptional data of both sets were processed using the same pipeline and the results were subsequently compared. This comparison was made in order to obtain a measurement of the overlap degree on the effects in resource distribution of the differentially expressed genes in both mutants, suggesting a shared mechanism of resource-reallocation via a differential use of its RNA polymerase. Given this finding and recent research on the topic, we came to the conclusion that the RNAP is an attractive engineering target for designing bacterial strains with higher yields.

Resumen

La ingeniería metabólica es un área de investigación con mucha actividad. Su objetivo suele ser la producción de un compuesto de interés en un sistema biológico. Este proceso introduce una carga extra sobre los recursos compartidos del sistema, generando condiciones de estrés que impactan sobre el rendimiento de expresión y la calidad del producto. Esto hace que el desarrollo sea laborioso y costoso, necesitando varios pasos de ajuste. En este proyecto, investigamos la base de un acercamiento complementario enfocado en la reorganización de recursos celulares para proveer un posible incremento de tolerancia a la carga metabólica introducida, acompañada de un posible incremento en la eficiencia traduccional. Lo hicimos a través de estudiar el cambio en la repartición de recursos en una cepa de *Escherichia coli* con operones ribosomales eliminados (SQ53) así como en una cepa con una mutación puntual en el gen que codifica la subunidad beta de la RNA polimerasa (rpoBE672K). Ambas cepas mutantes mostraron incrementos en su tasa de crecimiento en medio mínimo M9 con glucosa, comparado a la cepa silvestre. Dado que en el sitio del NCBI ya existían datos de lecturas del transcriptoma de la cepa rpoBE672K y su control en las condiciones estudiadas, solo elegimos procesar muestras de la cepa SQ53 y su control en nuestro laboratorio. Para determinar lo anterior, se procedió a llevar a cabo experimentos de RNA-seq, por lo que se tomó muestras de cultivos en una etapa temprana de la fase exponencial y se extrajo el RNA total para construir bibliotecas de cDNA. Las bibliotecas fueron secuenciadas, filtradas por calidad, procesadas y analizadas para determinar los genes expresados diferencialmente (GEDs). Esta información fue utilizada para estimar el cambio en el balance proteómico, incorporando datos de proteómica que previamente fueron obtenidos en las mismas condiciones de cultivo. Los perfiles transcripcionales de las cepas estudiadas fueron procesados a la par y con los mismos algoritmos. Sus resultados fueron comparados para investigar la existencia de grupos compartidos de GEDs, sugiriendo la existencia de un mecanismo compartido de repartición de recursos a través de un uso diferencial de la RNA polimerasa. Dado este hallazgo y recientes investigación en el tema, llegamos a la conclusión de que la RNA polimerasa es un blanco atractivo de ingeniería para el diseño de cepas bacterianas con rendimientos de producción más altos.

Introducción

Desde hace 40 años, la expresión de proteínas heterólogas en *Escherichia coli* continua siendo una importante área de investigación y desarrollo para la industria farmacéutica. En 2018, la producción de aproximadamente el 30% de proteínas terapéuticas comercializadas se hacía con diversas mutantes de *E. coli*. Este microorganismo es también un hospedero atractivo para la fabricación de plásmidos con aplicaciones de terapia génica (Castiñeiras et al., 2018).

Por varios años se ha trabajado en incrementar el tamaño del acervo sobre el funcionamiento metabólico de *E. coli* y se han registrado múltiples experiencias sobre su uso. Aún así, desarrollar proyectos de expresión de proteínas usando este hospedero sigue siendo una tarea altamente compleja y tardada. Frecuentemente se requieren de una combinación de varios pasos de ajuste como: optimización de codones, condiciones de crecimiento (medio nutritivo, temperatura y pH), selección de plásmido utilizado, suplementación de cofactores, sobreexpresión de proteínas chaperonas y sobreexpresión de proteínas asistentes de la formación de puentes disulfuro. Estas, son algunas de las medidas que se recomiendan dependiendo de la proteína a expresar. Aún así, no proveen completa garantía de que su implementación tenga resultados positivos (Rosano & Ceccarelli, 2014).

Estos ajustes son necesarios por la carga metabólica adicional introducida en el sistema para obtener un producto heterólogo. Es común observar que esta introducción estimule el sistema de respuesta a estrés del hospedero. La activación de este sistema termina impactando en muchos casos sobre la cantidad y calidad del producto final (Castiñeiras et al, 2018).

Los proyectos de biología sintética y biología de sistemas buscan generar sistemas de expresión más robustos a la expresión de proteínas heterólogas. Algunos de estos esfuerzos han ido en la dirección de continuar la extensión de los principios de repartición de recursos celulares y sus orígenes a nivel de sistema (Hausser et al, 2019; Utrilla et al, 2016). Otros han investigado los principios de eficiencia traduccional (Valgepea et al, 2013); relaciones de coordinación entre la tasa de crecimiento y la regulación de la expresión génica (Scott et al, 2010); estrategias de liberación de recursos a través de la supresión de funciones no esenciales utilizando

algoritmos de predicción (Lastiri et al, 2020); y modelos computacionales de metabolismo, expresión y regulación a nivel genómico para la simulación y predicción del comportamiento de sistemas de expresión (Tan et al, 2020; O'Brien et al, 2015 ; Sastry et al, 2019; Ebrahim et al, 2016).

Estos modelos coinciden en que los recursos del medio de crecimiento son repartidos en grupos distintos del proteoma de acuerdo a su abundancia, calidad energética y factores del contexto ambiental, como la acidez y temperatura del medio. Estos principios de administración resultan útiles para promover la supervivencia de los organismos en condiciones fluctuantes. Existe un módulo propuesto que agrupa a los componentes del proteoma en 3 fracciones o sectores: El sector *Q*, proteínas que realizan funciones de mantenimiento de la homeostasis y se asume que sus niveles de abundancia son robustos a perturbaciones. Se asume también que es un sector teóricamente de tamaño fijo e independiente de la tasa de crecimiento. El sector *P*, conformado por proteínas con funciones de metabolismo, con tamaño y composición dependientes de la tasa de crecimiento. Por último, el sector *R*, que agrupa proteínas asociadas con la síntesis y conformación de ribosomas. Se asume que este último sector también es una representación de la abundancia de ribosomas (Scott et al., 2010).

Se ha observado que *E. Coli* creciendo en medios de cultivo con diferente composición puede alcanzar la misma tasa característica de crecimiento. Esto se atribuye a una repartición similar del uso de recursos celulares entre los 3 sectores del proteoma a pesar de las diferencias entre medios. Es decir, que la expresión génica, representada por la configuración de los sectores del proteoma, tiene una relación mayor con la tasa característica de crecimiento que con la composición específica del medio. En particular, el tamaño del sector *R* o ribosomal tiene una fuerte correlación positiva con la tasa de crecimiento, al estar vinculado directamente con la capacidad de síntesis de proteínas (Scott et al., 2010).

La abundancia de ribosomas activos, factores de elongación (o EF-Tu) y tRNAs cargados con aminoácidos, son parámetros que limitan la eficiencia traduccional y con ello la tasa de crecimiento (Mori et al., 2016). Existe la teoría de que la utilización óptima de nutrientes podría maximizarse mediante la redistribución de los

recursos del proteoma, garantizando que los sustratos para la maquinaria traduccional estén siempre saturados (Dennis et al., 2008).

Es común observar tasas bajas de crecimiento en condiciones de medio mínimo. En estas condiciones, la saturación de complejos ternarios (factores de elongación más tRNAs cargados con aminoácidos) y la fracción de ribosomas activos son el factor limitante de la tasa de crecimiento. En medio rico se vuelve posible alcanzar tasas más altas de crecimiento, ya que en abundancia de recursos, la tasa solo depende de la abundancia total de ribosomas. En estas condiciones, la mayoría de los ribosomas se encuentran activos y saturados con complejos ternarios. Se propone que la estrategia de regulación de la tasa de crecimiento es diferente y más estricta en condiciones de medio mínimo porque tiene como objetivo prolongar la supervivencia esperando una fluctuación positiva de nutrientes. En condiciones de medio rico el objetivo cambia a maximizar la tasa de crecimiento lo más posible, ya que la abundancia de recursos no resulta limitante. Al limitar la tasa de crecimiento por medio de reducir la fracción activa de ribosomas y no a través de reducir su concentración, la reversión de la limitación, en caso de que las condiciones mejoren, termina teniendo un menor costo de recursos y también reduce el tiempo necesario para hacerlo (Dai et al., 2016).

Que exista una fracción inactiva de ribosomas en condiciones de medio mínimo, plantea la posibilidad de alcanzar tasas de crecimiento mayor, interfiriendo con la estrategia regulatoria que limita su uso. Esta perturbación podría tener como resultado una distribución del proteoma con una capacidad traduccional mayor. Esto también podría causar una pérdida de capacidad de adaptación, ya que se disminuirían los recursos dedicados a las funciones de respuesta a estrés, que son de mayor necesidad en condiciones de medio mínimo.

En la ingeniería de cepas optimizadas para condiciones industriales, se simplifica la investigación al enfocarse en diseñar estrategias que resulten útiles bajo condiciones controladas (Gorochowski & Ellis, 2018). En este caso, perder capacidad de adaptación a condiciones fluctuantes sería una consecuencia tolerable. Describir un mecanismo de control para la mejora de crecimiento descrita, podría ser un principio transferible al diseño de mejores sistemas de expresión de

proteínas a nivel industrial. (Bosdriez et al, 2015 ; Valgepea et al, 2013 ; Utrilla et al, 2014).

El presente proyecto de tesis de maestría tiene como uno de sus objetivos investigar el cambio en la expresión génica en una cepa mutante de *E. coli*, denominada SQ53, con genotipo $\Delta rmGBAD$ (con operones ribosomales eliminados). Esta mutante tiene el fenotipo de interés descrito arriba de manera general: una tasa de crecimiento más alta que la de la cepa silvestre en medio mínimo a costa de perder capacidad de adaptación.

Se aprovechará también la disponibilidad de datos transcriptómicos del experimento de otra mutante de *E. coli* con un fenotipo similar al mencionado. Esta mutante lleva la sustitución de un aminoácido en la subunidad beta de la RNA polimerasa, es la cepa rpoBE672K. La expresión diferencial de ambas cepas será comparada para investigar la presencia de un mecanismo de origen en común y con ello aportar al conocimiento que se tiene sobre los principios de repartición de recursos en estas condiciones.

Antecedentes

Organismos adaptados a ambientes de nutrición fluctuante, tienden a tener un número alto de operones ribosomales en comparación a otros adaptados en nichos más estables (Gyorfy et al., 2015). El mecanismo de control global por acumulación de (p)ppGpp también está conservado en organismos adaptados a ambientes fluctuantes, mientras que en organismos que habitan en nichos protegidos está ausente por completo (Wagner, 2010).

Ambos mecanismos comparten una estrecha relación de influencia sobre el uso de la RNA polimerasa. Los operones ribosomales tienen promotores de muy alta afinidad con la RNA polimerasa y la acumulación de (p)ppGpp puede restringir fuertemente la afinidad de promotores con la RNA polimerasa, incluidos los ribosomales. Al tener un rol central sobre la composición del transcriptoma, se puede decir que la distribución del uso de la RNA polimerasa sobre el genoma tiene un impacto alto en la jerarquía de regulación de expresión génica. Los efectos de su distribución son luego amplificadas al pasar por el proceso de traducción, generando fenotipos diferentes acordes a las condiciones de crecimiento.

En bacterias en crecimiento rápido, hasta 85% del RNA en la célula puede tratarse de RNA ribosomal (rRNA) y hasta 75% de la transcripción puede estar ocurriendo sobre operones ribosomales (Gyorfy et al., 2015; Mauri et al., 2014). Es necesario contar con la maquinaria traduccional correspondiente para suplir la demanda de capacidad de síntesis de proteínas en altas tasas de crecimiento. En estas condiciones, el 60% de la biomasa total pueden ser proteínas y casi la mitad del peso seco pueden ser ribosomas donde ~85% está activamente traduciendo proteínas (Dai et al., 2020).

La existencia de múltiples operones ribosomales en el genoma de *E. coli* es resultado de una adaptación para incrementar rápidamente la dosis génica de operones ribosomales durante la replicación. Con una dosis génica mayor, incrementa también la oportunidad de su transcripción. Esto un recurso muy útil para aprovechar oportunamente condiciones de abundancia de nutrientes. Sin embargo, reducir el número de operones ribosomales no necesariamente resulta en una tasa menor de crecimiento. También se ha observado que cepas mutantes de

E. coli que carecen de dos operones ribosomales pueden crecer a tasas similares a la cepa silvestre en medio rico. Existen mutantes con aún menos operones ribosomales que muestran ventajas de crecimiento en medios pobres y estables (Condon et al, 1995; Gyorfy et al, 2015).

Las observaciones de estas mutantes sugieren que es posible optimizar el número de operones ribosomales en el genoma. Pueden obtenerse cepas mutantes con una posible ventaja en eficiencia traduccional y mayor tasa de crecimiento. Sin embargo, aún hay incertidumbre sobre la extensión y detalle del impacto sobre la expresión génica que tiene una modificación como ésta.

En mutantes con operones ribosomales inactivados, se ha cuantificado que la relación de RNA/proteína y la concentración de ribosomas totales no muestran cambios significativos con respecto a la silvestre. La reducción de los operones ribosomales activos en el genoma es compensada por incrementos en su tasa de transcripción, sin que la concentración de (p)ppGpp cambie (Condon et al, 1993). Este registro de ausencia de cambios en la concentración de (p)ppGpp es una observación importante. La concentración de (p)ppGpp está inversamente relacionada con la capacidad de transcripción de RNA ribosomal, por lo tanto con la síntesis de ribosomas y también con la tasa de iniciación de la traducción. Es decir, la compensación de la tasa de transcripción en estas mutantes no ocurre por una concentración menor de (p)ppGpp.

El mecanismo de regulación por (p)ppGpp, también llamado 'respuesta estricta' tiene un rol central sobre el control global de la expresión génica. El producto (p)ppGpp es sintetizado por dos enzimas, RelA y SpoT. La enzima SpoT en condiciones de abundancia de aminoácidos tiene actividad hidrolasa de (p)ppGpp, limitando su acumulación. La enzima RelA va unida al ribosoma. La síntesis de (p)ppGpp en situación de escasez de aminoácidos se estimula por un cambio alostérico de RelA en respuesta a la carga repetida de tRNAs vacíos en el ribosoma. Una vez que esto ocurre, RelA sintetiza (p)ppGpp a partir de GTP (Winther et al., 2018).

Según diversas evidencias de estudios estructurales, al unirse (p)ppGpp a la proteína DksA, esta puede unirse a la subunidad beta de la RNA polimerasa y cambiar la estabilidad del complejo abierto. Este cambio de la estabilidad disminuye

la afinidad del complejo con promotores ricos en GC, tales como el promotor P1 de los operones ribosomales. La concentración de (p)ppGpp también regula negativamente la replicación a través de la inhibición de la primasa. Regula negativamente la traducción al competir con GTP por el sitio activo del factor de inicio de la traducción IF2 (Wagner, 2010). En condiciones de abundancia de recursos, no se acumula (p)ppGpp y la transcripción de los operones ribosomales se vuelve dependiente de la concentración de RNA polimerasa disponible. Es así como la tasa de transcripción de RNA ribosomal (rRNA) aumenta en paralelo con la tasa de crecimiento (Bremer, 2008).

Durante la respuesta estricta también ocurre la regulación positiva de genes relacionados con la respuesta de adaptación a diversos estreses. Se estima que ocurre de forma pasiva, debido a la liberación súbita de RNA polimerasa. Estos genes regulados positivamente cumplen funciones relacionadas con la inducción de la fase estacionaria, formación de biopelículas, síntesis y transporte de aminoácidos y degradación de proteínas, respuesta a estrés oxidativo, estrés osmótico y estrés de acidez por la regulación positiva del factor sigma RpoS, entre otros. La activación de estas funciones resulta conveniente para favorecer la supervivencia en condiciones poco favorables (Wagner, 2010; Mauri et al., 2014).

Análisis estructurales de la RNAP que contiene la mutación en la subunidad beta, en la cepa rpoBE672K, mostraron un efecto sobre la estabilidad del complejo abierto de la polimerasa en presencia de (p)ppGpp, alterando su dinámica de interacción con factores sigma. A esta alteración se le atribuye un efecto pleiotrópico que cambia la repartición de recursos en el transcriptoma, promoviendo la expresión de genes orientados al crecimiento y disminuyendo la expresión de genes orientados a la cobertura o respuesta a estrés (Utrilla et al., 2016).

En estos antecedentes podemos encontrar la importancia de regular la actividad y distribución de la RNA polimerasa sobre el genoma. Así como de la importante influencia que tiene sobre su regulación la dinámica de unión de la subunidad beta de la RNA polimerasa con (p)ppGpp y DksA durante la respuesta estricta. Y también la estrecha relación de la respuesta estricta con la multiplicidad de los operones ribosomales.

Todos los operones ribosomales de *E. coli* tienen la misma estructura general (Figura 1). Estudios sobre el impacto de la reducción de operones ribosomales han partido previamente de la suposición de que todos ellos poseen promotores con fuerzas idénticas debido a su composición muy similar. Sin embargo, esto contrasta con observaciones reportadas de un distinto impacto sobre la actividad de la transcripción de los operones al inactivar la misma cantidad de estos pero en combinaciones distintas (Gyorfy et al., 2015).



Figura 1. Estructura general de un operón ribosomal en *E. coli*. La cepa WT de *E. coli* tiene siete operones ribosomales nombrados de la siguiente manera: *rrnA*, *rrnB*, *rrnC*, *rrnD*, *rrnE*, *rrnG* y *rrnH*. Todos ellos tienen la misma estructura general como se muestra en la figura (Condon et al., 1995).

Algunos grupos han investigado la existencia de una regulación heterogénea, debido a que los operones tienen micro heterogeneidades en sus secuencias. Estos grupos han encontrado que existen diferencias en los niveles de actividad de los 7 operones en ausencia de limitación por acumulación de (p)ppGpp y también tienen diferente sensibilidad a la atenuación en presencia de su acumulación. Atribuyen estas diferencias a la presencia de un número distinto de sitios de unión para el activador transcripcional Fis. Específicamente, se encontró que los operones *rrnA*, *rrnB*, *rrnC*, *rrnD* y *rrnH* contaban con entre 3 a 4 sitios de unión a Fis mientras que *rrnE* y *rrnG* tienen 8 y 6 sitios de unión a Fis, respectivamente (Kolmsee et al., 2011).

Se analizó en ese mismo estudio el rango de atenuación de los promotores usando fusiones transcripcionales con el gen que codifica la proteína fluorescente GFP. Se encontró que la actividad de los promotores de los operones *rrnC* y *rrnE* disminuye más lento que la actividad del resto de los promotores de los otros operones ribosomales. Sus niveles de expresión se mantienen detectables incluso después de 20 a 30 horas de su inoculación en medio fresco. Después de ese mismo periodo de tiempo, la actividad del resto de los operones es indetectable.

Para el promotor de *rrnE* se determinó la ausencia de regulación por el represor H-NS. Este promotor fue el que registró mayor actividad por Fis durante la fase

exponencial de crecimiento. (Kolmsee et al., 2011; Maeda et al., 2015). No existe evidencia experimental para argumentar sobre el efecto aditivo o sustractivo de distintas combinaciones de operones. Pero las observaciones descritas por estos antecedentes sugieren que la cepa SQ53, que conserva los operones *rrnC*, *rrnH* y *rrnE*, podría tener una combinación con distinta tolerancia a la atenuación por (p)ppGpp, que otras combinaciones distintas del mismo número de operones.

La cepa SQ53 tiene una tasa de crecimiento mayor en ~11% respecto a la cepa silvestre MG1655 (WT) en condiciones de medio mínimo M9 con glucosa a una concentración de 4 gr/L. Adicionalmente, también perdió capacidad de adaptabilidad al observarse un alargamiento sobre su cambio diáuxico (Hidalgo et al, 2021) y una menor tolerancia a la sobrevivencia en fase estacionaria, previo a una reinoculación a medio fresco, volviéndose más sensible a la reinoculación entre más tiempo transcurría. Estas observaciones sobre la disminución de tolerancia al tiempo en fase estacionaria en mutantes de *E. Coli* con menos operones ribosomales también fueron reportadas previamente (Condon et al., 1995).

La cepa SQ53 fue construida por otro grupo de investigación como parte de una biblioteca de cepas con operones ribosomales eliminados de manera secuencial. La motivación general detrás de la construcción de las bibliotecas fue el de facilitar la investigación sobre proyectos centrados en el funcionamiento de los ribosomas en *E. Coli* (Quan et al., 2015).

En un estudio previo, otra cepa de *E. coli* con una mutación puntual en la subunidad beta de la RNA polimerasa, la cepa *rpoBE672K*, mostró un fenotipo de crecimiento similar al de la SQ53. Esta cepa fue obtenida como parte de un experimento de evolución adaptativa donde una cepa silvestre de *E. coli* MG1655 fue utilizada como punto de partida. En términos generales, el experimento consistió en múltiples pasajes del cultivo hacia medio fresco durante su fase exponencial, sin permitir que el cultivo entrase en fase estacionaria por una cantidad aproximada de 2000 generaciones (LaCroix et al, 2015).

La mutante *rpoBE672K* mostró un incremento en su tasa de crecimiento de ~25% acompañado también de pérdidas en capacidad de adaptación. Perdió tolerancia a choque ácido, mostró menor resistencia a antibióticos y también menor motilidad (Utrilla et al., 2016). También es necesario mencionar que su ventaja de crecimiento

fue observada en medio mínimo M9 con glucosa y que tanto esta mutante como la SQ53 tuvieron velocidades de crecimiento menor en medio rico LB.

En conjunto, estos antecedentes sugieren que tanto la mutante SQ53 como la rpoBE672K remodelan su repartición de recursos por medios similares. Específicamente, a través de un uso diferencial de la RNA polimerasa, impactando sobre la expresión génica durante la fase de crecimiento exponencial.

La similitud en su fenotipo sugiere que ambas mutantes lo logran trasladando recursos hacia funciones de crecimiento, teniendo un impacto negativo sobre sus funciones de cobertura. Dicho efecto debería reflejarse en la sobrerregulación en común de grupos de genes orientados al crecimiento y la subregulación de grupos de genes orientados a funciones de cobertura. La cepa SQ53 y rpoBE657 no han sido estudiadas en conjunto con este objetivo previamente. La observación de similitud de sus fenotipos a pesar de sus diferencias en la naturaleza de sus mutaciones fue parte importante de la motivación de este proyecto.

Se seleccionó la técnica de RNA-Seq en conjunto con un algoritmo de estimación del proteoma diseñado por Lastiri et al, 2020 como medio para realizar los análisis necesarios para estudiar la validez de esta hipótesis. Esta solución involucra la aceptación de un margen de error en la estimación del proteoma, ya que no toma en cuenta cambios en la regulación postraduccional y parte de la suposición que los cambios a nivel transcripcional tendrán el mayor impacto sobre la composición final del proteoma. Originalmente, para resolver este problema del margen de error, se había propuesto un método más apropiado para cuantificar los fragmentos que efectivamente serán traducidos en proteínas: el protocolo de *Ribosome profiling*.

El principio de este protocolo fue desarrollado a partir de la idea de mejorar la resolución del análisis del transcriptoma por RNA-Seq. Logrando esto a través de digerir las cadenas de mRNA que son transcritas pero no están siendo activamente traducidas. Al mismo tiempo digerir el mRNA alrededor de ribosomas que sí están en traducción activa, separar rRNA y depletar tRNAs para obtener una biblioteca que al secuenciarse reflejara el traductoma de la muestra. Después de hacer correcciones a nivel del procesamiento bioinformático para tomar en cuenta eventos de aborto de la traducción, el conteo de lecturas alineadas por gen podrían tomarse

como una aproximación mucho más cercana que un RNA-Seq tradicional al proteoma traducido real (McGlincy & Ingolia, 2017).

Durante la realización de este proyecto de tesis se realizaron grandes avances en la estandarización de una versión del protocolo de *Ribosome profiling* con puntos críticos reducidos y controles de verificación integrados. El protocolo avanzado también es de costo reducido en comparación con el protocolo en el que fue basado originalmente (Latif, 2015). Contiene adaptaciones tomadas de un protocolo simplificado que fue hecho con el objetivo de reducir costos, acortar tiempos de ejecución y reducir pérdidas de la muestra, un aspecto crítico en la ejecución del protocolo previo a la construcción de biblioteca (Reid et al, 2015).

Simplificar protocolos de alto impacto como estos es también útil para mejorar la reproducibilidad y reducir la dificultad de transferencia entre laboratorios. Desafortunadamente, por retrasos incurridos por la restricción de entrada a los laboratorios por la pandemia de COVID-19, no fue posible culminar con la estandarización de este protocolo y alcanzar a incluir resultados del mismo en este proyecto de investigación. Sin embargo, los avances obtenidos en forma de la última versión del protocolo estandarizado se incluyen en la sección de anexos como evidencia del trabajo realizado (Anexo 4).

Hipótesis

La eliminación de los operones ribosomales *rnnGBAD* en la cepa SQ53 tiene consecuencias a nivel global de regulación de la transcripción, de manera similar al efecto provocado por la mutación *rpoBE672K*. Estos cambios tienen como consecuencia una remodelación del proteoma y un posible incremento en la eficiencia traduccional a costa de la disminución de su capacidad de adaptación.

Objetivo general

Comparar la distribución del transcriptoma entre las cepas SQ53 y silvestre MG1655 de *E. coli*. Investigar similitudes en el patrón de expresión diferencial y distribución del proteoma obtenido para el experimento con la cepa SQ53 contra los obtenidos para el experimento de la cepa *rpoBE672K*. Discutir con hallazgos reportados en la literatura, para proponer un vínculo en común entre las diferentes perturbaciones de ambas y la similitud en su fenotipo.

Objetivos particulares

- 1 Diseñar y conducir un experimento para obtener lecturas por duplicado de RNA-Seq, de la cepa SQ53 y WT de *E. coli* en las mismas condiciones de crecimiento en las que fueron obtenidas las lecturas de la cepa *rpoBE672K*. Obtener sus listas de genes expresados diferencialmente.
- 2 Estimar el cambio en la distribución del proteoma utilizando datos de proteómica previamente disponibles.
- 3 Con base en los grupos de genes expresados diferencialmente compartidos entre la mutante SQ53 y la *rpoBE672K*, proponer relaciones del origen de su mutación con sus fenotipo de crecimiento y razones para su similitud.

Materiales y métodos

Los protocolos detallados, en el formato utilizado para su ejecución, incluyendo recomendaciones adicionales pueden encontrarse en la sección de anexos. En esta sección se incluye la metodología descrita a detalle.

- **Caracterización de tasas de crecimiento**

Se tomaron muestras de 900 uL de cultivo cada 15 minutos, hasta alcanzar una DO de 0.4 (600nm). Utilizando el software *graphpad Prisma*, los datos se procesaron para obtener sus tasas de crecimiento en fase exponencial y pruebas para validar su diferencia. Las tasas de crecimiento de la mutante *rpoBE672K* y su control fueron obtenidas del artículo donde se publicaron (Utrilla et al., 2016). Para determinar la significancia estadística de los resultados se usó la prueba t de student con 3 réplicas por cepa sobre la media de las tasas de crecimiento evaluando con un valor de $p < 0.05$.

- **Cultivo y muestreo para la extracción de material de RNA en medio M9 con Glucosa al 0.4%**

Para ambos cultivos se sembró un preinóculo 16 horas antes en matraces de 125 ml utilizando 25 ml de medio M9 con glucosa al 0.4% en una agitación de 210 rpm a 37°C. Se midió la DO a 600nm de los cultivos en una dilución 1:10 previo a la inoculación de los mismos, para calcular el volumen de inoculación necesario para iniciar el cultivo de muestreo a una OD600 de 0.05 en un volumen de 70 ml del mismo medio en matraces de 250 ml.

La incubación de los cultivos se hizo en una tina de baño maría a una temperatura controlada de 37°C utilizando agitadores magnéticos funcionando a 1000 rpm para mantenerlos saturados de oxígeno. Para la toma de lecturas de DO a 600 nm de los muestreos de los cultivos, se utilizó un espectrofotómetro modelo *biospectrometer kinetic* (Eppendorf #6136000002). Siguiendo las recomendaciones de uso del fabricante, se utilizó la misma celda para tomar todas las mediciones, reduciendo así el error técnico entre las lecturas. Para la caracterización del crecimiento se siguió el método descrito en el punto anterior y una vez que la absorbancia alcanzo un valor de 0.3 y 0.4 (600nm) (intervalo calificado como el inicio de la fase exponencial), se cosecharon 3 ml y se mezclaron con 6 ml de RNA Protect

(QIAGEN #76506) utilizando un vórtex por 5 segundos. Se incubó esta mezcla por 5 minutos a temperatura ambiente (15-25°C) y se centrifugó por 10 minutos a 5000 rcf. Se decantó el sobrenadante y se almacenaron las pastillas celulares en un ultracongelador a -80°C.

- **Lisis celular, extracción y preparación de RNA para la construcción de bibliotecas de RNA-Seq**

Las pastillas celulares se resuspendieron en 100 µl de buffer TE pH 8.0 con lisozima (20 mg/ml) (Sigma-Aldrich #L6876) y se agregó a la mezcla, 15 µl de proteinasa K (800 unidades/ml) (New England Biolabs #P8107S). La mezcla se agitó por 10 segundos utilizando un vórtex y se incubó a 25°C por 10 minutos en agitación constante. Se agregaron 350 µl de buffer de lisis del kit de purificación Quick-RNA Miniprep (Zymo Research #R1054) adicionado con 3.5 µl de beta-mercaptoetanol (Sigma-Aldrich #516732) y se mezcló en vórtex por 3 segundos.

La mezcla se centrifugó a 16000 rcf por 1 minuto para usar el sobrenadante en el siguiente paso. Se purificó el RNA del sobrenadante, utilizando el kit de purificación Quick-RNA MiniPrep mencionado anteriormente, siguiendo las instrucciones del fabricante, incluyendo el paso de digestión de DNA, para finalmente resuspender el RNA en 50 µl de agua libre de RNAsas. La integridad del material extraído se verificó utilizando la técnica de separación por electroforesis en gel de agarosa al 1% utilizando buffer TAE.

- **Construcción de bibliotecas y secuenciación de lecturas**

La construcción de bibliotecas se realizó utilizando como material de inicio 10 µL de RNA a una concentración mínima de 500 ng/µl. La separación de RNA ribosomal se ejecutó con el kit Zymo-Seq RiboFree Total RNA Library (Zymo Research R#3000). Las bibliotecas se construyeron siguiendo el protocolo de TruSeq Stranded mRNA (Illumina #20020594) para secuenciación con la plataforma Illumina Next-Seq 500. Las muestras fueron secuenciadas con una profundidad requerida de 10 millones de lecturas pareadas.

Las bibliotecas de la mutante rpoBE672K (SRR1514300 y SRR1514301) y la cepa control MG1655 (SRR1787590 y SRR1787591) fueron descargadas desde el servidor ENA (European Nucleotide Archive). Según su información disponible en el servidor, las bibliotecas fueron construidas utilizando el método dUTP y fueron

secuenciadas con la plataforma Illumina Mi-Seq con una profundidad requerida de 5 millones de lecturas pareadas.

- **Procesamiento bioinformático para la obtención de genes expresados diferencialmente (GEDs)**

El procesamiento se realizó utilizando el paquete de SystemPipeR del lenguaje de programación R en la plataforma Jupyter lab. Este paquete incluye herramientas para la generación de flujos de trabajo que facilitan la generación y pase de archivos entre pasos de procesamiento, de forma semi-automática, evitando la generación de errores por nombramiento equivocado de archivos (Backman & Girke, 2016).

El corte de adaptadores y bases de baja calidad se hizo utilizando Trim-Galore, programa basado en el programa cutadapt (Martin, 2011), el reporte integrado de calidad de las muestras se hizo con la herramienta fastQC (Andrews, 2010). Para el alineamiento de lecturas se utilizó el genoma de referencia GCF_000005845.2_ASM584v2 y el alineador Bowtie 2.0 (Langmead, 2012).

La transformación de archivos de alineamiento para la cuantificación de lecturas se hizo con las herramientas de Samtools. La función de summarizeOverlaps() (Lawrence et al., 2013) se utilizó para el conteo de lecturas con las opciones: *“mode = “Union”, singleEnd = FALSE, ignore.strand = FALSE, preprocess.reads = invertStrand”* apropiadas para este tipo de biblioteca.

Este procesamiento se realizó por separado para cada conjunto de lecturas utilizando los mismos parámetros tanto para las lecturas del experimento de la mutante rpoBE672K y su control como para las del experimento de la mutante SQ53 y su control.

Se inspeccionó la dispersión de las réplicas utilizando una exploración por reducción de dimensionalidad con el algoritmo de análisis de componentes principales para tener una noción de la reproducibilidad entre réplicas y tomar una decisión informada sobre el criterio de corte de significancia estadística al obtener los genes expresados diferencialmente de ambos experimentos.

Para los resultados de la cepa la SQ53 se decidió sobre un criterio menos estricto de corte de significatividad que para los resultados de la cepa rpoBE672K por haber tenido una mayor variabilidad entre las réplicas. El criterio es menos estricto para la

cepa SQ53 para evitar descartar la lista de genes expresados diferencialmente a genes que no tuvieron niveles muy consistentes entre réplicas de la misma cepa, es decir, aceptamos el riesgo de incluir algunos falsos positivos para no perder genes que pudieran estar expresados diferencialmente como falsos negativos. Los niveles usados para las pruebas fueron $p < 0.1$ con una magnitud de cambio logarítmica de base 2 diferente de 0 para la SQ53 y $p < 0.05$ con una magnitud de cambio logarítmica de base 2 diferente de 1 para la rpoBE672K.

Se utilizó el paquete de R, DeSeq2 (Love et al, 2014) para la obtención de listas de genes expresados diferencialmente, esta herramienta realiza pruebas estadísticas por gen con un p valor ajustado para tomar en cuenta el error de pruebas múltiples además de que valora la consistencia entre réplicas para realizar la prueba de diferencia significativa. Se realizó una validación adicional de los genes encontrados utilizando el servidor IDEAmex (Jiménez et al, 2019) para confirmar que las listas también fuesen encontradas utilizando otros paquetes de análisis.

- **Estimación del cambio en distribución del proteoma**

El cálculo de cuentas obtenidas para un determinado gen puede verse sesgado por la profundidad de secuenciación de la biblioteca o por el tamaño del gen. Entre mayor sea la profundidad de la biblioteca, las cuentas de un gen largo tenderán a inflarse más en comparación con las de un gen más corto en una biblioteca con menor profundidad, porque será menor su probabilidad de ser detectadas. Es decir, sería más probable detectar expresión diferencial para genes largos que para genes más cortos, solo por que el método es sensible a su diferencia de tamaño.

Para normalizar por tamaño de biblioteca y tamaño del gen, se puede obtener el número de lecturas por kilobase por millón de lecturas totales. El término para referirse a esta normalización es FPKM para lecturas pareadas, este algoritmo considera fragmentos para la cuenta total incluso si sólo uno del par alineó al genoma de referencia, dado que esta lectura haya tenido una alta calificación de calidad. Se aprovecha esta normalización asumiendo que cada FPKM equivale a un transcrito de mRNA de la proteína en cuestión.

La estimación del peso del proteoma en femtogramos a partir de datos de RNAseq de la cepa mutante y la silvestre para cada experimento se realizó siguiendo el principio diseñado por nuestro grupo de investigación y publicado en Lastiri et al,

2020, que utiliza las cuentas en FPKMs promediadas entre réplicas calculadas por el algoritmo del programa Cufflinks (Trapnell et al. 2010).

Después, el método asume que el transcriptoma en FPKMs de la cepa silvestre, al traducirse da como resultado el proteoma medido por Schmidt et al, 2016 en las mismas condiciones de crecimiento. Siguiendo esta suposición, se obtienen entonces las tasas de traducción para cada gen en unidades de copias de proteína por transcrito (s_i).

$$s_i = C_i \div r_i$$

Donde C_i corresponde a las copias de proteína reportadas asociadas al gen en particular y r_i a su abundancia en FPKMs obtenida. Las tasas de traducción por gen pueden asumirse como constantes a lo largo de un amplio rango de condiciones de crecimiento. Aunque sí tienen algo de variación, se ha observado que para la mayoría de los genes en *E. coli*, están adaptadas para ser robustas con el propósito de evitar el desperdicio de recursos, excepto para algunas proteínas donde resulta conveniente tener un control menos estricto, como proteínas de bajo costo proteómico que cumplen funciones de cobertura (Hausser, et al., 2019).

Para las tasas de traducción de las proteínas de las que no se tiene información de copias por proteína en el proteoma de Schmidt et al, 2016, se asume una tasa de traducción promedio, calculada a partir del conjunto obtenido con la operación anterior.

Luego, utilizando estas tasas de traducción se obtiene el peso en femtogramos de proteína por gen para la mutante (P_i) y para las proteínas faltantes de la silvestre. El valor de la multiplicación de $r_i \times s_i$ es multiplicado por su peso molecular (MW) y dividido entre el número de Avogadro (NA) en las unidades correspondientes.

$$P_i = r_i \times s_i \times MW \div NA$$

El cambio ocasionado por la expresión diferencial de los genes en las mutantes es sumado al proteoma de la WT control correspondiente a los genes no expresados diferencialmente para terminar con un balance total aproximado del proteoma de la mutante.

Es importante mencionar que las proteínas ribosomales son filtradas para no ser tomadas en cuenta en este paso del análisis, ya que siguen un modelo de regulación predominantemente a nivel traduccional (Bremmer et al., 2008) y no se podría confiar en que sus niveles proteómicos cambiarán únicamente con cambios observados a nivel transcriptómico, para poder incluirlas en el análisis lo apropiado sería contar con datos obtenidos por la técnica de *Ribosome profiling*, también llamada RiboSeq. Esta técnica se basa en la secuenciación selectiva de fragmentos que fueron capturados por ribosomas para ser traducidos y por lo tanto representan una cuantificación más confiable del número de copia de sus proteínas.

Al inicio de este proyecto se realizaron avances de estandarización para implementar esta técnica, sin embargo, por complicaciones de la pandemia por COVID-19, estos no pudieron ser completados y se eligió continuar el proyecto con datos de RNA-Seq en su lugar. El protocolo de estandarización avanzado de la técnica de RiboSeq se incluye en anexos (Anexo 4) ya que también fue parte de los productos de este proyecto de tesis.

Se hizo un análisis de pareto para señalar los principales genes donde ocurre el ahorro e inversión de recursos y se hicieron también análisis de inversión y ahorro a nivel del transcriptoma utilizando datos de longitud de los genes para transformar las unidades normalizadas de FPKMs a kilobases de mRNA y poder hacer una comparación en unidades absolutas.

Con el propósito de explorar el impacto sobre el cambio del tamaño del proteoma causado por el cambio en la expresión diferencial; la tasa de traducción específica de los genes; y el peso molecular de las proteínas, se realizaron pruebas de correlación lineal entre estas variables y así poder discutir sobre el peso de la contribución de estos tres factores.

Finalmente, se hizo un análisis de la redistribución del proteoma organizado por sectores utilizando una reducción por grupos de genes ortólogos. Las tablas utilizadas para realizar esta reducción están señaladas en Anexos (Anexo 2).

Todos los pasos de procesamiento de las listas de genes expresados diferencialmente, generación de figuras y pruebas estadísticas después de los pasos con R, fueron hechos con los paquetes de Python: pandas, scipy, matplotlib,

y seaborn. Las funciones generadas fueron empaquetadas en un módulo nombrado rnapropy para su uso en otros proyectos con requerimientos similares. El módulo y un ejemplo utilizando datos del proyecto puede ser consultado en el repositorio de github: <https://github.com/AlexMart1/MasterThesis>

- **Análisis de interacción de GEDs con StringDb**

El servidor en línea de StringDb permite hacer análisis de evaluación de distintos tipos de interacción entre las proteínas en una lista dada. En su versión más reciente permite hacer un análisis compuesto que es capaz de evaluar una lista de genes para determinar si existen relaciones de proximidad entre estos en el genoma del organismo, relaciones de co-ocurrencia a través de distintos genomas e interacciones bioquímicas revisadas experimentalmente, además de reportes de co-expresión.

También es posible evaluar si existen genes con enriquecimientos significativos en grupos de función biológica y rutas metabólicas. Dentro del análisis también incluye un enriquecimiento por minería de texto donde detecta si genes de la lista han sido mencionados juntos dentro de artículos de investigación, implicando que han sido estudiados juntos dentro de algún otro fenómeno biológico de interés (Szklarczyk et al, 2019).

Esto facilita la investigación bibliográfica y permite obtener predicciones con base a lo reportado sobre la interconexión de los genes encontrados a distintos niveles más rápidamente. Aprovechamos esta herramienta para explorar la ocurrencia de relaciones de interacción entre las proteínas de las listas de genes subregulados y sobrerregulados en cada experimento. Y también para explorar si las listas de genes sub y sobre expresados que coinciden entre los experimentos, conservan grupos de proteínas con relaciones de interacción y con ello aportar a la discusión sobre la existencia de un mecanismo de origen en común del fenotipo observado en las mutantes.

- **Análisis de cambios en iModulones relevantes para la expresión de proteínas heterólogas**

Mediante un análisis utilizando separación de señales por el algoritmo de composición de señales independientes sobre un conjunto de 147 lecturas de RNA-Seq de *E. coli* en distintas condiciones experimentales, un grupo de investigación

encontró conjuntos de genes regulados de manera independiente, los cuales no necesariamente se encontraban bajo el control del mismo factor de transcripción pero sí responden a señales extracelulares o intracelulares de manera coordinada. A estos conjuntos de genes los llamaron 'iModulones', los cuales fueron propuestos como una forma alternativa de analizar la red regulatoria de *E. coli* a nivel transcripcional (Sastry et al., 2019).

Utilizando la lista de iModulones de mayor relevancia para la expresión de proteínas heterólogas, correspondientes a la respuesta a diversos estreses, síntesis de aminoácidos y maquinaria traduccional (Tan et al., 2020), comparamos los datos de los análisis de expresión diferencial de las mutantes rpoBE672K y SQ53, y se discuten sus implicaciones.

Finalmente, en este paso del análisis se incluirán a las proteínas ribosomales, ya que conforman en su mayoría el iModulon de Traducción. Este grupo de proteínas es de especial interés por representar una parte importante del efecto correlacionado de inversión a recursos de crecimiento sobre el ahorro de recursos de cobertura.

Resultados

- **Caracterización de tasas de crecimiento**

La tasa de crecimiento obtenida de la mutante SQ53 es diferente del control con significancia estadística de $p < 0.05$. La diferencia observada entre la mutante y la silvestre fue del ~11% (Figura 2).

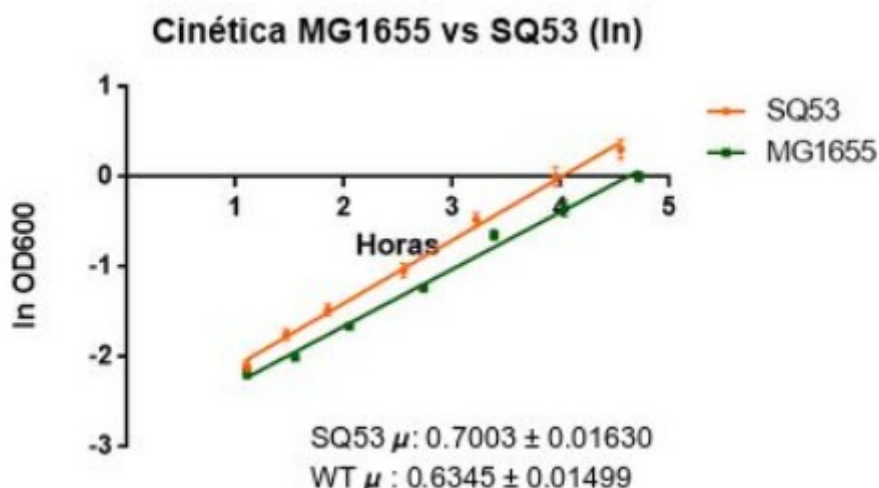


Figura 2. Curvas de crecimiento de las cepas SQ53 y silvestre, para RNA-Seq representadas en escala logarítmica. Los puntos representan el promedio de cada medición y se incluyen barras de error. Se muestra también la estimación de las tasas de crecimiento con un intervalo de 95% de confianza.

- **Procesamiento bioinformático para la obtención de genes expresados diferencialmente (GEDs)**

La construcción de las bibliotecas de la cepa silvestre (WT) y la cepa SQ53 se logró satisfactoriamente siguiendo los pasos mencionados en la metodología. La calidad del material de RNA utilizado para la construcción se analizó utilizando un ensayo de electroforesis en el equipo *Bioanalyzer 2100 Expert*. Los resultados de este ensayo dieron un RIN entre 7.9 y 9.3 para todas las cepas. En los anexos se incluye la figura que el equipo otorga como parte del resultado representando el análisis por electroforesis con mayor detalle (Anexo 1).

Después del corte de adaptadores de las lecturas con Trim-Galore, se encontraron diferencias en la consistencia de calidad de las lecturas a través de los ciclos de

secuenciación entre los experimentos de WT vs rpoBE672K y WT vs SQ53, más notables entre las lecturas de la rpoBE672K y SQ53 (Figura 3).

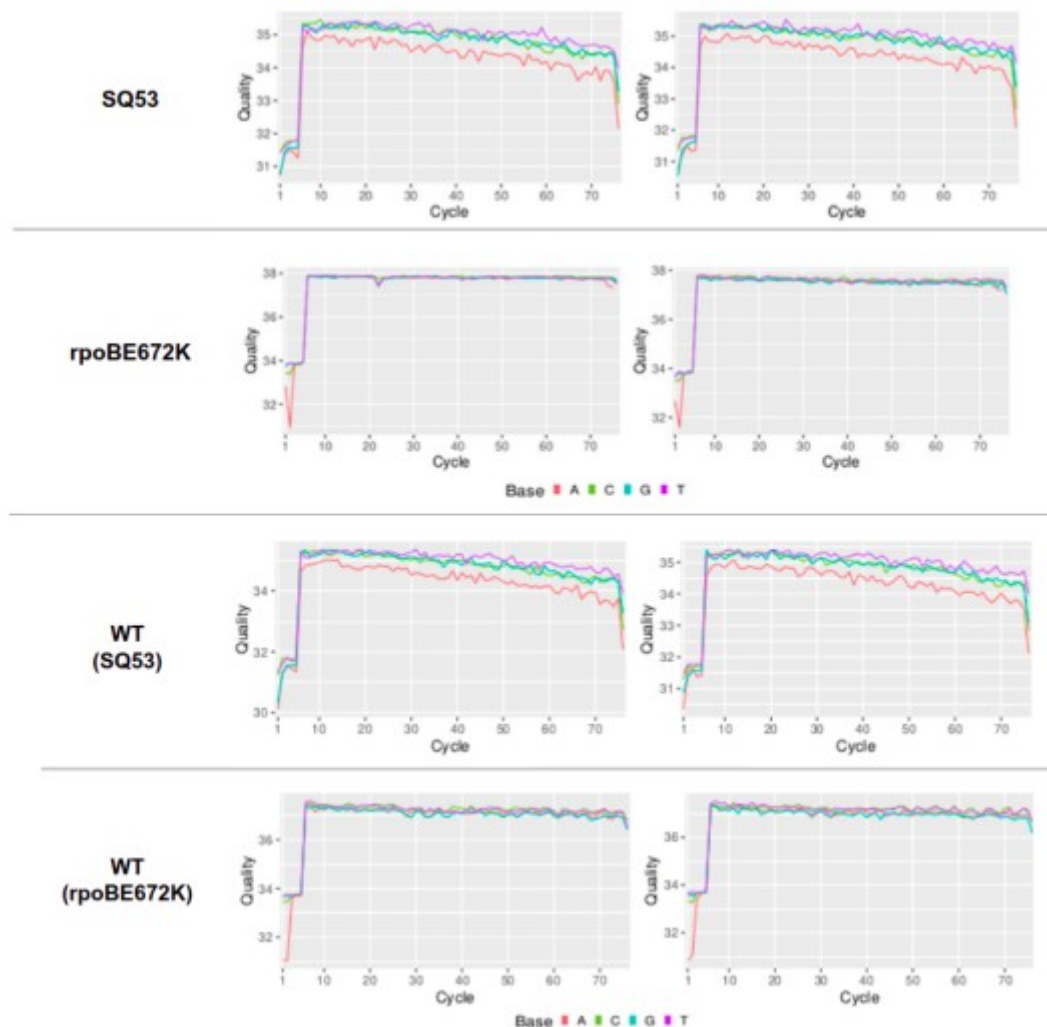


Figura 3. Análisis de calidad de la secuenciación por ciclos para las réplicas de las cepas SQ53 y rpoBE672K. Se muestra la calificación de calidad en el eje Y correspondiente a la escala logarítmica Phred y en el eje X el ciclo de secuenciación correspondiente a la calificación. Se muestran en distintos colores las series correspondientes a cada una de las distintas bases (A,C,G y T).

Estas diferencias pudieron deberse a la suma de varios factores. Los reactivos usados durante el proceso de incubación de los cultivos y extracción de RNA, diferencias debido a la ejecución por distintos operadores, las distintas plataformas de secuenciación utilizadas y la profundidad de secuenciación usada para los ensayos son todos factores que pudieron haber introducido ruido en cada experimento. De entre estos factores, se ha observado que entre mayor es la profundidad, la sensibilidad al ruido aumenta (Tarazona et al, 2011).

Esto no fue interpretado como una razón de descarte de las réplicas del experimento de la cepa SQ53. Más bien, se identificó la necesidad de seleccionar un corte de significancia estadística y magnitud de cambio que fuese tolerante al ruido y balancear en la medida de lo posible la cantidad de posibles falsos negativos. Para el experimento de la cepa rpoBE672K, sí se pudo ser más estricto con ambos parámetros de corte, aprovechando la consistencia entre sus réplicas.

Todos los archivos de lecturas para todas las muestras, se alinearon con el genoma de referencia por encima del 90%, las del experimento SQ53 entre el 95-96% y las del experimento rpoBE672K entre el 98-99% (Tabla 1). Este resultado nos muestra la baja o poca presencia de material contaminante en las muestras. No se realizaron análisis para investigar la presencia de fuentes de contaminación.

	Lecturas totales (Millones)	Lecturas alineadas (Millones)	Porcentaje de Alineamiento
WT - 1 (SQ53 exp)	29.34	28.15	95.96 %
WT - 2 (SQ53 exp)	27.16	26.22	96.50 %
SQ53 - 1	28.84	27.81	96.4 %
SQ53 - 2	32.43	31.29	96.47 %
WT - 1 (rpoBE672K exp)	12.61	12.54	99.45 %
WT - 1 (rpoBE672K exp)	8.6	8.55	99.43 %
rpoBE672K - 1	4.19	4.15	98.94 %
rpoBE672K - 2	7.63	7.55	98.96 %

Tabla 1. Tabla de estadísticas de alineamiento para todas las réplicas experimentales y de control.

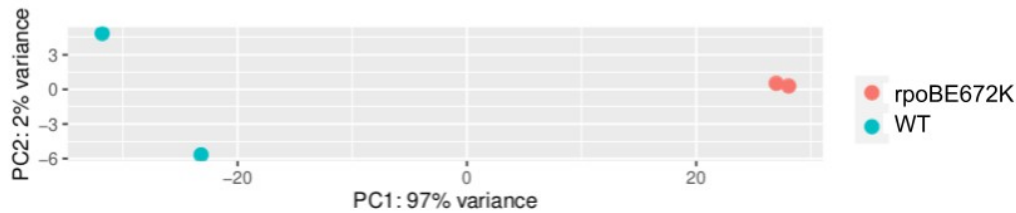


Figura 5. Análisis de componentes principales para las lecturas experimentales de la cepa rpoBE672K y control. Análisis realizado utilizando el algoritmo de PCA para revisar la dispersión entre los datos de las réplicas de las lecturas de las cepas WT y rpoBE672K de Utrilla et al, 2016.

Los resultados del análisis por componentes principales del experimento correspondiente a la cepa rpoBE672K (Figura 5) son congruentes con lo esperado por su baja profundidad de secuenciación con menor sensibilidad al ruido. Las réplicas de cada grupo se separan con muy poca variación entre ellas y su componente principal de separación parece corresponder con la fuente de variación estudiada, la mutación. (Utrilla et al, 2016)

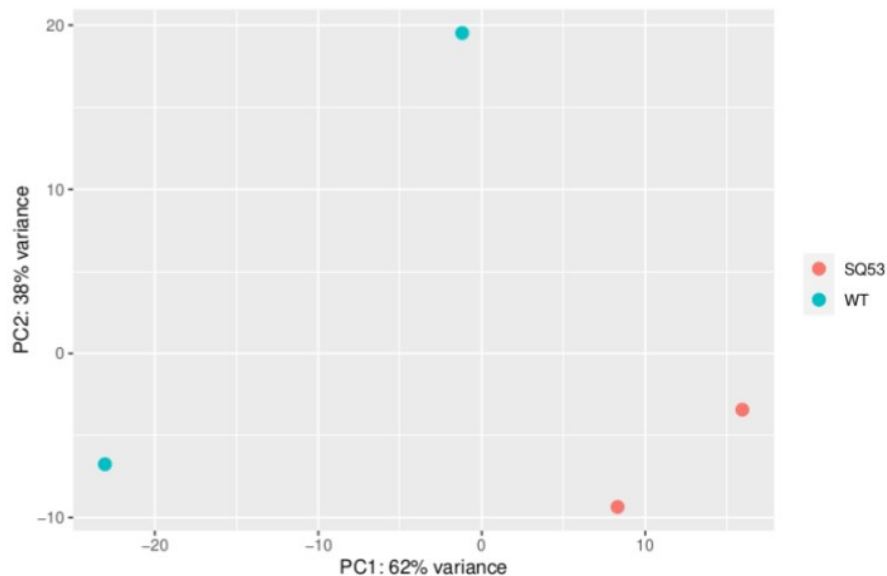


Figura 6. Análisis de componentes principales (PCA) para las lecturas experimentales de la cepa SQ53 y control.

Las réplicas del experimento de la cepa SQ53 tienen una mayor variabilidad entre ellas a lo largo del total de genes evaluados, sobre todo entre las réplicas de la cepa WT. Pensamos que la variabilidad entre réplicas es mayor para este experimento que para el de la cepa rpoBE672K vs WT porque la profundidad de secuenciación utilizada fue mayor y con ello aumentó la presencia de ruido en las lecturas. Lo

mejor habría sido incrementar el número de réplicas para este experimento pero ello quedó fuera del alcance de este proyecto.

El algoritmo del paquete de análisis DESeq2 obtiene la lista de genes expresados diferencialmente descalificando aquellos genes cuyas cuentas entre réplicas tienen baja consistencia. Como resultado, se puede tener una mayor confianza de que los genes calificados como expresados diferencialmente no son falsos positivos, a costa del riesgo de obtener falsos negativos descalificados por falta de reproducibilidad.

Utilizando el método estadístico DESeq2, se encontraron 327 genes expresados diferencialmente en la cepa rpoBE672K con un LogFC 2 y pvalue < 0.05 y 114 para la cepa SQ53 con un LogFC 2 y pvalue < 0.1, con respecto a la cepa silvestre. Al pasar esta lista de genes a través del filtro de remoción de tRNAs, rRNAs, pseudogenes y proteínas ribosomales, el número de genes diferencialmente expresados se redujo a 272 en la cepa rpoBE672K y 104 en la cepa SQ53.

La lista de genes expresados diferencialmente encontrados por el análisis anterior también fueron detectados con otros softwares de análisis. Se utilizó el servidor IDEAmex citado en la sección anterior que ayuda a realizar análisis con otros métodos de forma rápida e integrada.

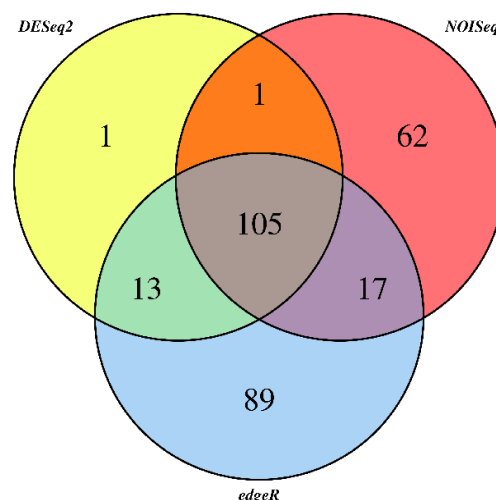


Figura 7. Diagrama de Venn que muestra los genes expresados diferencialmente en la cepa SQ53 con respecto a la silvestre. Se muestra el número de genes diferencialmente expresados en usando los siguientes métodos edgeR, NOISeq y DESeq2.

Después de hacer una revisión de sobreposición con los resultados obtenidos para este estudio con la pipeline antes mencionada, se encontró una coincidencia con 91 de los 105 genes señalados en la triple intersección encontrada por IDEAmex. Esto

valida con mayor confiabilidad los resultados obtenidos utilizando solo DESeq2 para este estudio en particular. Los genes no incluidos en la lista estudiada que sí aparecen en la lista de IDEAmex son en su mayoría tRNAs y rRNAs, genes que por las consideraciones del estudio habrían sido filtrados más adelante.

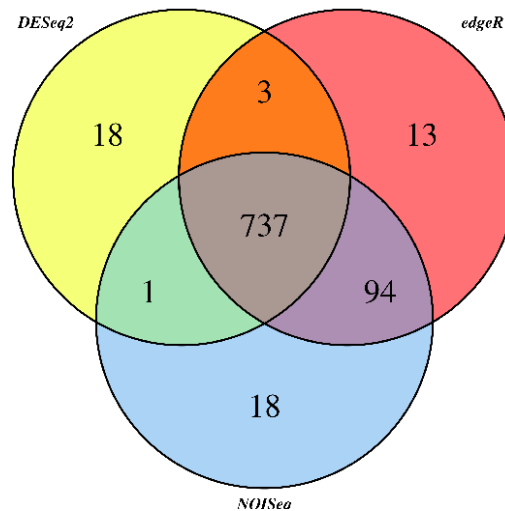


Figura 8. Diagrama de Venn de grupos de genes expresados diferencialmente para el experimento rpoBE672K. Identificación de grupos y diagrama generado utilizando la plataforma IDEAmex con los programas edgeR, NOISeq y DeSeq2.

Para el experimento de expresión diferencial de la rpoBE672K fueron detectados más genes expresados diferencialmente que los determinados por el procesamiento utilizado por nosotros. Es probable que la causa de la diferencia sea la mayor consistencia entre réplicas y otras diferencias en opciones utilizadas para ejecutar el análisis. El método fue repetido asegurándonos que los datos de entrada utilizados y los valores de los parámetros de corte fuesen los mismos y a pesar de ello la diferencia aquí reportada persistió. No continuamos investigando la raíz detrás de ello y lo atribuimos a una diferencia en la aplicación del algoritmo que desconocimos. Analizando las coincidencias entre las listas se encontró que 195 de los 272 genes detectados por el método utilizado aparecen en el conjunto de la triple intersección dado por IDEAmex.

De los genes no incluidos en la lista de los 272 la mayoría son tRNAs y rRNAs, el resto de la lista no incluidos conservaban una relación funcional o de proximidad genómica con los incluidos en la lista de los 272. Por motivos de consistencia en el procesamiento se conservó la lista de menor tamaño de 272 genes para los siguientes pasos en el procesamiento.

Esta diferencia en abundancia de genes detectados por todos los métodos disponibles entre ambos experimentos, sugiere que tal vez existan más diferencias entre la cepa SQ53 y la silvestre que fueron descartadas por penalización de la inconsistencia entre réplicas y no necesariamente porque su expresión no sea diferente en realidad. Los resultados del análisis por categorías funcionales de los grupos de genes diferencialmente expresados se encuentra unas secciones más adelante. Estos brindan una vista a un nivel más alto sobre las funciones que están siendo sobrerreguladas o subreguladas y se comenta el impacto que esto tiene sobre el fenotipo.

- **Estimación del cambio en distribución del proteoma**

Los resultados de la estimación del proteoma asumen que las tasas de síntesis específicas de proteína para las cepas mutantes son las mismas que las ajustadas para la cepa silvestre, según lo descrito en la metodología. Por las suposiciones hechas, se acepta que existe un mayor margen de error que con una cuantificación directa ya que otras diferencias en las etapas posteriores a la transcripción podrían alterar los niveles finales de abundancia de proteínas. No fue incluido dentro del alcance de este proyecto una validación experimental de las estimaciones realizadas a partir del cálculo descrito en la metodología. Sin embargo, se argumenta que el cálculo tiene utilidad para aproximar el posible proteoma dado el transcriptoma de cada cepa. Los resultados presentados por las figuras 9 y 14 se analizan desde esa perspectiva.

Esta sección también incluye otros análisis adicionales al proceso de estimación, que son útiles en la discusión. También se presentan los cambios asociados por grupos de genes asignados a distintos sectores del proteoma al final de la sección.

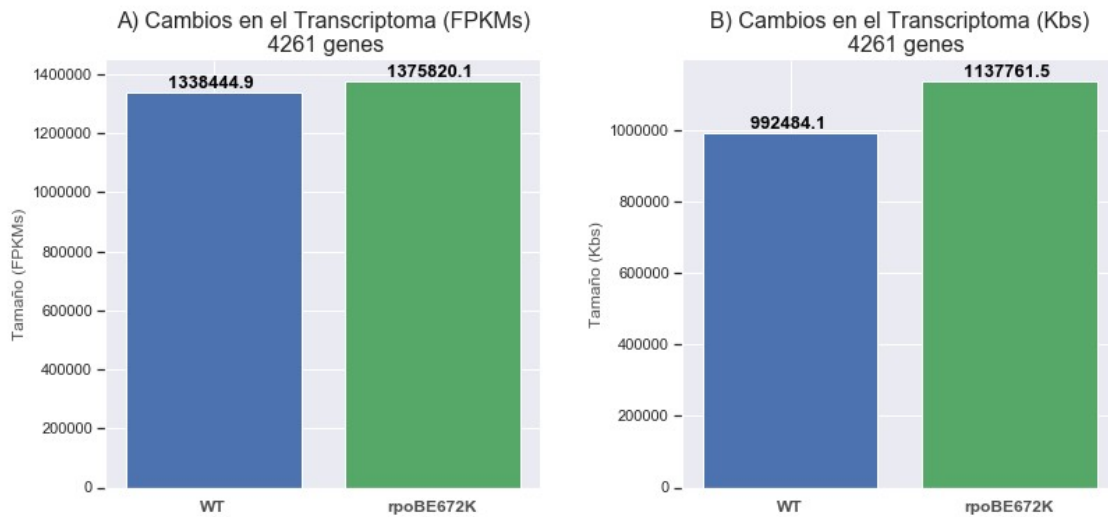


Figura 9. Comparación del tamaño del transcriptoma total para las lecturas del experimento WT vs. rpoBE672K. A) Total acumulado en FPKMs (unidades normalizadas por profundidad de biblioteca y longitud de cada gen) para los experimentos de la cepa WT y la cepa rpoBE672K. B) Total acumulado en Kbs (kilobases, unidades absolutas) para los experimentos de la cepa WT y la cepa rpoBE672K.

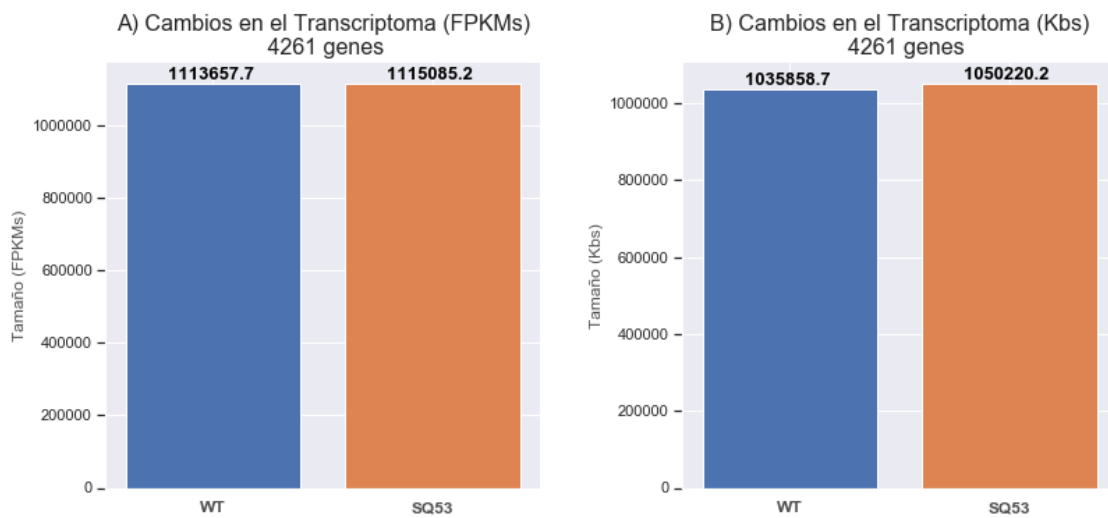


Figura 10. Comparación del tamaño del transcriptoma total para las lecturas del experimento WT vs. SQ53. A) Total acumulado en FPKMs (unidades normalizadas por profundidad de biblioteca y longitud de cada gen) para los experimentos de la cepa WT y la cepa SQ53. B) Total acumulado en Kbs (kilobases, unidades absolutas) para los experimentos de la cepa WT y la cepa SQ53.

La primera observación es que el transcriptoma es de un tamaño mayor en ambas cepas mutantes (figura 9 y figura 10), la diferencia es más notoria para la rpoBE672K que para la SQ53. Al pasar de unidades normalizadas por profundidad de biblioteca y longitud del gen (FPKMs) a unidades absolutas (Kilobases),

podemos notar que la diferencia es un poco más fácil de apreciar. Partiendo de esta observación, pasamos a hacer esta misma comparación pero separando por grupos de genes sobrerregulados y subregulados, esperando aprender más sobre los detalles de la diferencia.

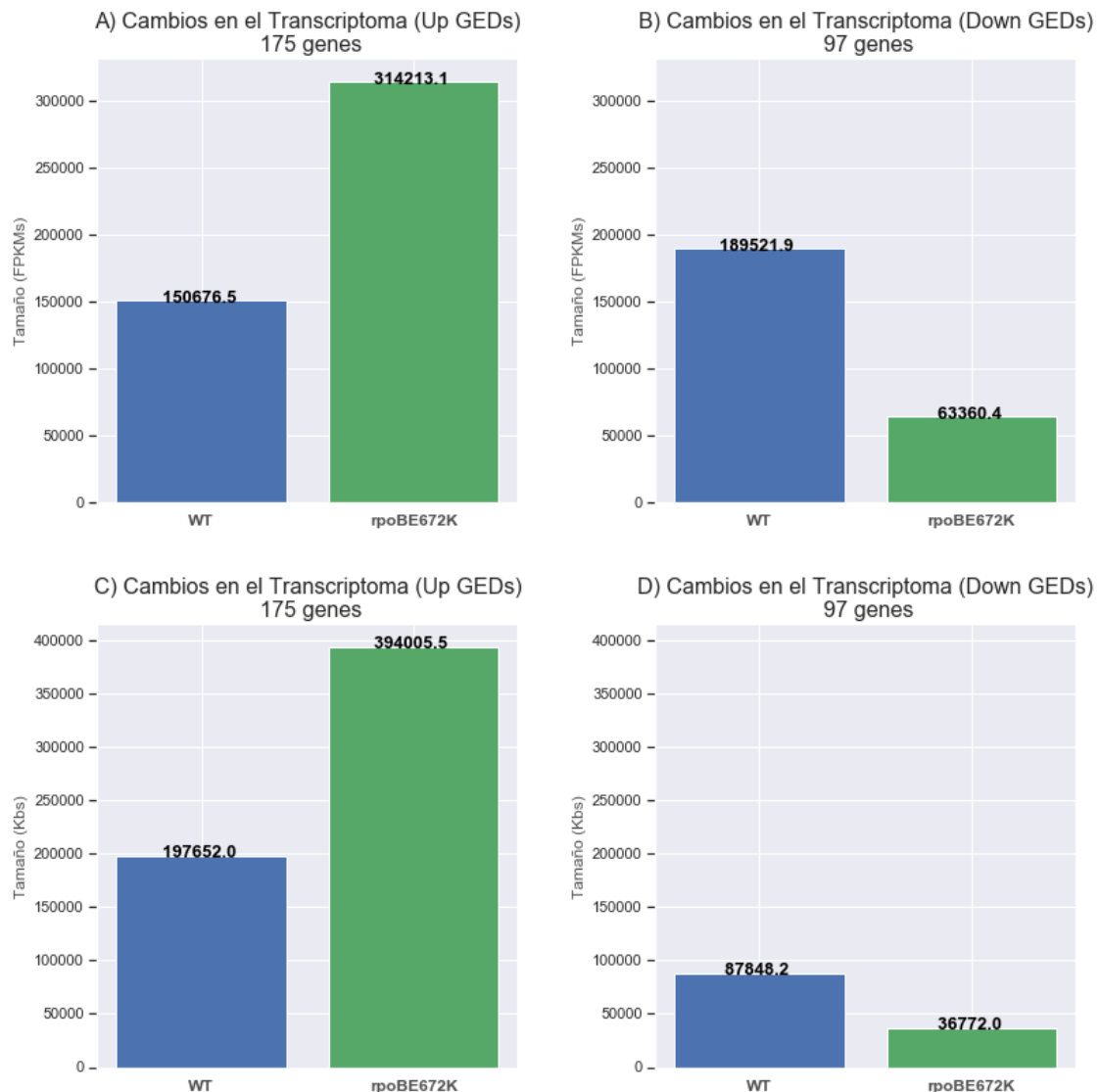


Figura 11. Comparación del tamaño del transcriptoma reorganizado para las lecturas del experimento WT vs. rpoBE672K por grupos de genes subregulados y sobrerregulados. A) y B) Tamaño total de los grupos ordenados por conjuntos sobrerregulados y subregulados en FPKMs. C) y D) Tamaño total de los grupos ordenados por conjuntos sobrerregulados y subregulados en kilobases (Kbs)

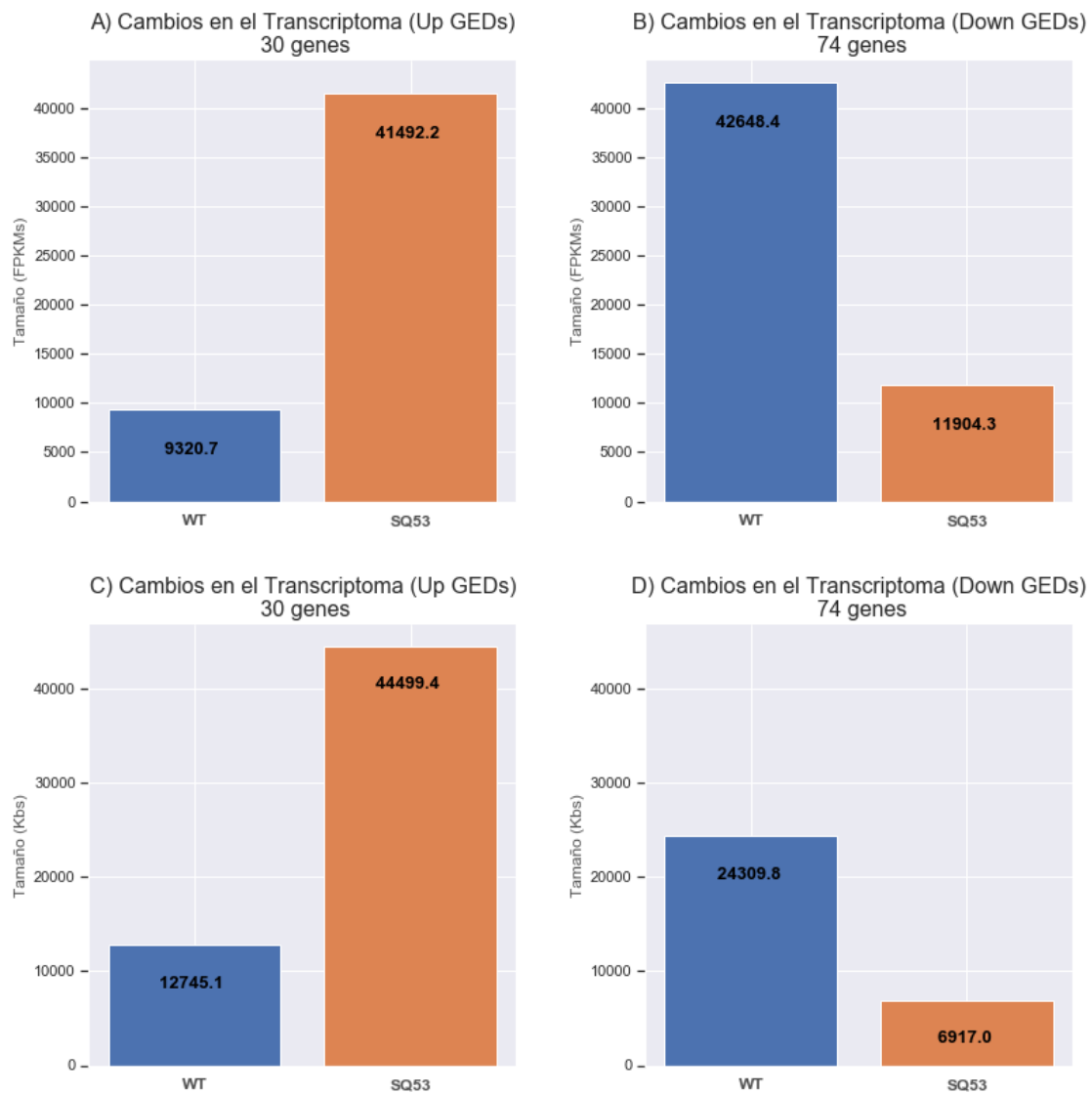


Figura 12. Comparación del tamaño del transcriptoma reorganizado para las lecturas del experimento WT vs. SQ53 por grupos de genes sobrerregulados y subregulados. A) y B) Tamaño total de los grupos ordenados por conjuntos sobrerregulados y subregulados en FPKMs. C) y D) Tamaño total de los grupos ordenados por conjuntos sobrerregulados y subregulados en kilobases (Kbs)

A pesar de las diferencias en escala y el número distinto de genes en cada grupo en ambos experimentos, se puede observar un patrón interesante en esta comparación del transcriptoma dividido por grupos de genes sobrerregulados y subregulados. Centrando nuestra atención en las gráficas C) y D) de las figuras 11 y 12 donde la escala está en unidades absolutas de kilobases, podemos notar que la longitud de los genes correspondientes a los subregulados parecen ser más cortas que para los genes sobrerregulados.

Hacemos esta observación porque a pesar de que en el experimento SQ53 son más del doble los genes subregulados que los genes sobrerregulados, su cambio absoluto es menor que el cambio absoluto de los sobrerregulados. Aunque en el experimento rpoBE672K es mayor la cantidad de genes sobrerregulados que los subregulados, la diferencia del cambio absoluto es más del doble de la que podríamos esperar si la distribución de la longitud de los genes en cada grupo fuese igual. Dada esta observación, continuamos a revisar si existía entonces diferencia significativa entre las medianas de la longitud de los genes entre estos grupos.

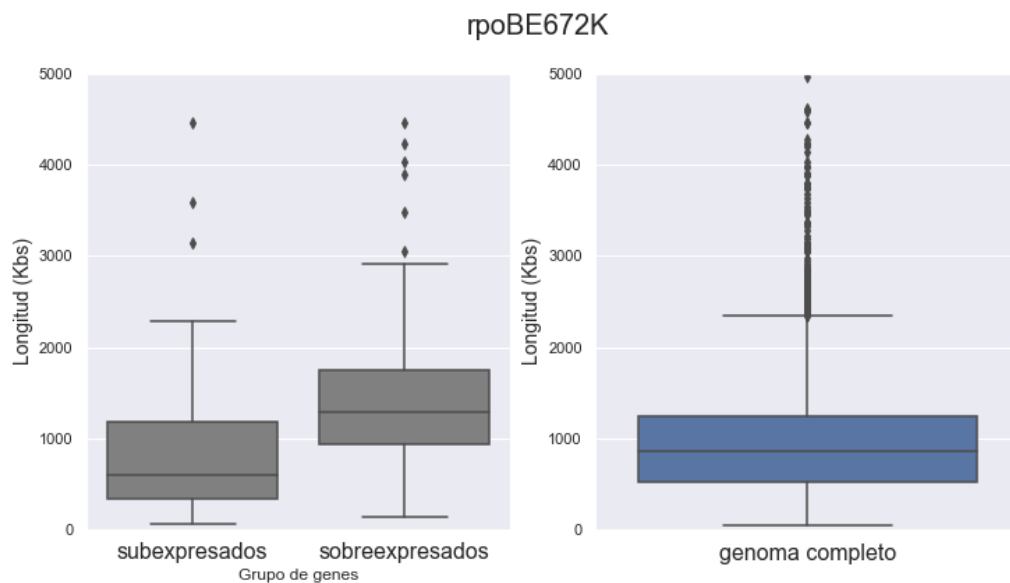


Figura 13. Diagrama de cajas de la longitud en kilobases (Kbs) de los grupos de genes expresados diferencialmente en rpoBE672K y genoma completo. La prueba estadística de Mood para evaluar igualdad de medianas entre los grupos de genes subregulados y sobrerregulados da un valor de $p=2.26 \times 10^{-8}$. La prueba estadística de Mood es una prueba no-paramétrica para comparar medianas de dos muestras

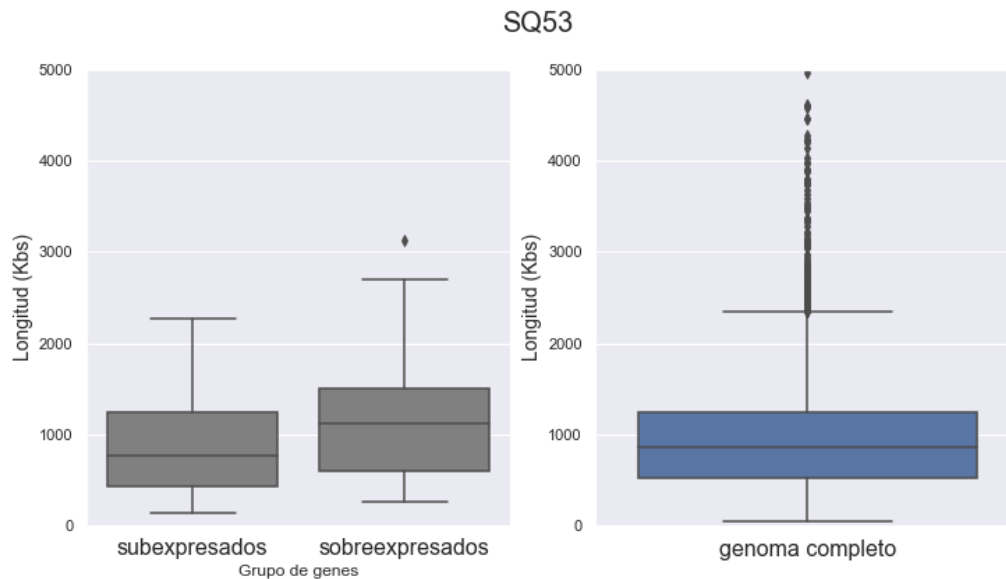


Figura 14. Diagrama de cajas de la longitud en kilobases (Kbs) de los grupos de genes expresados diferencialmente en SQ53 y genoma completo. La prueba estadística de Mood para evaluar igualdad de medianas entre los grupos de genes subregulados y sobregulados da un valor de $p=0.0972$. La prueba estadística de Mood es una prueba no-paramétrica para comparar medianas de dos muestras

Podemos observar en las figuras 13 y 14 que la distribución y mediana de la longitud de los genes parece estar sesgada alrededor de un valor más grande para los genes sobregulados que para los subregulados. Es decir, por lo general, los genes sobregulados suelen ser más largos que los genes subregulados. Especulamos que si se repitiese el experimento, terminaríamos con una lista más grande y completa de genes expresados diferencialmente para la SQ53 y veríamos una diferencia más marcada en longitud de genes entre genes subregulados y sobregulados, como en la rpoBE672K.

También podemos hacer la comparación con la mediana y distribución completa de la longitud de los genes en *E. coli* para analizar, si además las distribuciones de los genes expresados diferencialmente también son diferentes de la distribución general. Como podemos apreciar, los genes sobregulados en ambas mutantes parecen tener una longitud mediana diferente y mayor a la general y los genes subregulados tienen una longitud mediana diferente y menor a la general.

En trabajos anteriores se ha observado que la longitud de los genes, junto con su abundancia estable, son los factores de mayor peso en el costo energético de un gen a nivel de la transcripción, independientemente de su secuencia específica de nucleótidos (Lynch & Marinov, 2015). Esto sugiere que la transcripción de los genes

sobrerregulados podría tener un costo energético mayor al ahorrado por la subexpresión de otros genes en ambas mutantes.

Dadas estas condiciones y que no existan diferencias en sus tasas de traducción que compensen las diferencias en su abundancia de transcripción, podríamos esperar que a nivel del proteoma también exista un aumento del tamaño del proteoma en ambas mutantes.

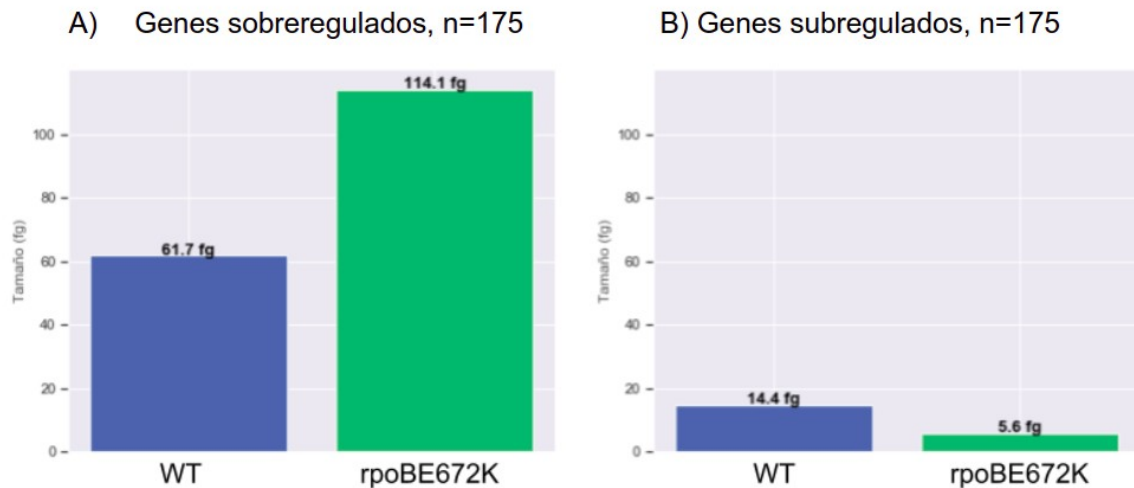


Figura 15. Comparación de cambios en el tamaño del proteoma estimado en femtogramos (fg) para el experimento WT vs. rpoBE672K. A) Cambio del proteoma atribuible a los genes sobreregulados. B) Cambio del proteoma atribuible a los genes subregulados. Ambas gráficas comparten la escala del eje Y.

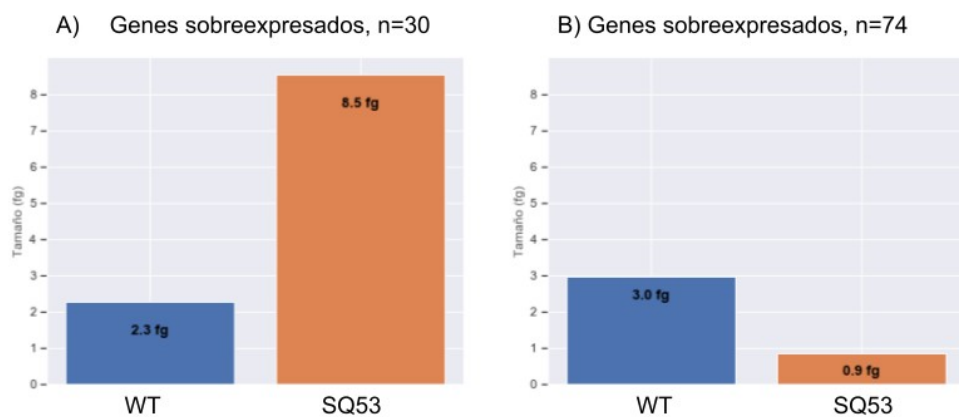


Figura 16. Comparación de cambios en el tamaño del proteoma estimado en femtogramos (fg) para el experimento WT vs. SQ53. A) Cambio del proteoma atribuible a los genes sobreregulados. B) Cambio del Proteoma atribuible a los genes subregulados.

A pesar de que entre los experimentos de las mutantes, la cantidad de genes que conforman los grupos de genes expresados diferencialmente es muy distinta, podemos ver que el patrón observado en el transcriptoma es el mismo que el

observado a nivel del proteoma. Los genes sobrerregulados parecen tener tasas de traducción lo suficientemente altas y diferentes a las de los genes subregulados como para que no se compense la diferencia en abundancia de transcritos. Es necesario reiterar que este análisis no incluye datos sobre las diferencias que podrían venir de regulaciones post-traduccionales, se asume que a nivel global no tendrían un impacto importante sobre el patrón observado en las figuras 15 y 16.

El costo energético de la traducción de un gen, en unidades de fosfatos hidrolizados, está principalmente determinado por el tamaño de la proteína, la composición específica de su secuencia de aminoácidos y su abundancia en estado estable (Lynch & Marinov, 2015).

A pesar de que este estudio no cubre las diferencias a nivel de la composición específica de la secuencia de aminoácidos entre los grupos de las proteínas sobreexpresadas y subexpresadas, la diferencia en su abundancia y tamaño sugieren que la inversión energética demandada a nivel de la traducción es superior al ahorro energético posiblemente incurrido en ambas mutantes.

Investigando la relevancia por separado en diferencias de abundancia de transcripción, tasas de traducción y peso molecular contra su correspondiente contribución al cambio en femtogramos del balance del proteoma total, podemos ver cuál es la fuerza de contribución de estas tres variables. Además de analizar si también están correlacionadas la abundancia de transcripción y las tasas de traducción, es decir, si los genes que tienen altas abundancias de transcripción, son también los que tienen tasas más altas de traducción.

Inversión vs Ahorro - rpoBE672K

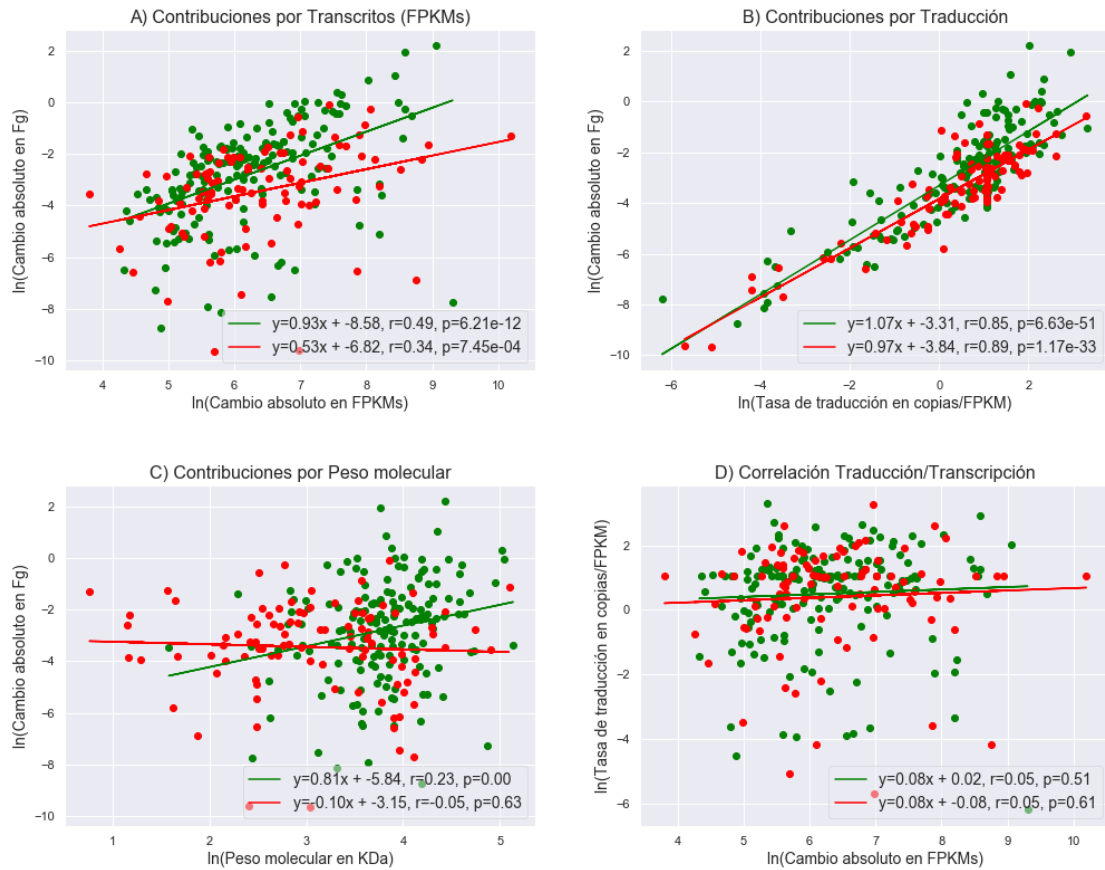


Figura 17. Comparación de inversión vs. ahorro en rpoBE672K. Genes sobrerregulados en color verde para representar la inversión de recursos y genes subregulados en color rojo representan ahorro de recursos. En cada gráfica se muestran también los resultados del análisis de correlación lineal correspondiente a cada grupo en su respectivo color. Logaritmo natural del cambio absoluto en femtogramos por proteína en función de A) el logaritmo natural del cambio en abundancia de transcritos, B) el logaritmo natural de la tasa de traducción (copias/FPKM) y C) el logaritmo natural del peso molecular en kilodaltones. D) Análisis de correlación entre el logaritmo natural de las tasas de traducción (copias/FPKM) y el logaritmo natural del cambio absoluto en FPKMs.

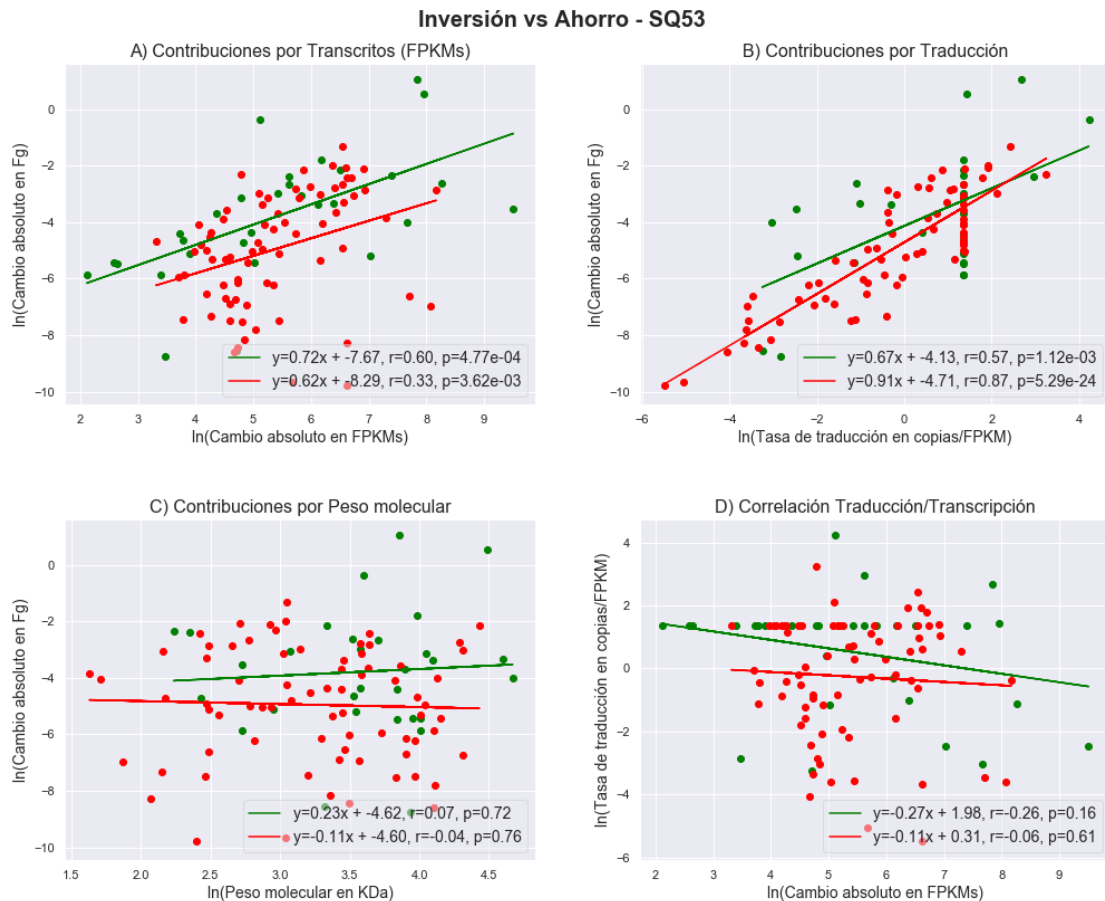


Figura 18. Comparación de inversión vs. ahorro en SQ53. Genes sobrerregulados en color verde para representar la inversión de recursos y genes subregulados en color rojo representan ahorro de recursos. En cada gráfica se muestran también los resultados del análisis de correlación lineal correspondiente a cada grupo en su respectivo color. Logaritmo natural del cambio absoluto en femtogramos por proteína en función de A) el logaritmo natural del cambio en abundancia de transcritos, B) el logaritmo natural de la tasa de traducción (copias/FPKM) y C) el logaritmo natural del peso molecular en kilodaltones. D) Análisis de correlación entre el logaritmo natural de las tasas de traducción (copias/FPKM) y el logaritmo natural del cambio absoluto en FPKMs.

Lo que podemos observar en las figuras 17A, 17B, 18A y 18B es que tanto la abundancia de transcripción como la traducción están correlacionadas con el cambio de tamaño del proteoma. Pero en la subfigura D se puede observar que no están relacionadas entre ellas, es decir que un gen de alta transcripción, no necesariamente tiene una alta tasa de traducción.

Incluso hay varios genes subregulados encontrados en muy baja abundancia que tienen tasas de traducción más altas que otros genes sobrerregulados. Es la suma de la combinación a través de ambas etapas de la síntesis de proteínas lo que genera el efecto completo del aumento y reducción del proteoma.

También en las figuras 17C y 18C, podemos concluir que el peso molecular de cada proteína tiene poca correlación con el cambio en el tamaño del proteoma. Este conjunto de observaciones sugiere que el origen en común en los fenotipos de las mutantes podría deberse a un origen coordinado desde un uso distinto de la RNA polimerasa. Esta diferencia en el uso impacta sobre la transcripción de grupos de genes orientados al crecimiento y sobre la distribución de la maquinaria traduccional. Este uso diferencial podría estar provocando cambios en la transcripción de varios genes en grupos co-regulados en lugar de la sobreexpresión o subexpresión de unos pocos genes regulados independientemente.

A nivel de la traducción y como grupo, los genes subregulados parecen tener tasas ligeramente más bajas que los genes sobrerregulados, según las figuras 17B y 18B. Esto puede tener relación con la presión evolutiva sobre las funciones que desempeñan. Los genes sobrerregulados en ambas mutantes desempeñan funciones que es posible asociar a ventajas de crecimiento y que en condiciones de abundancia de recursos, resultaría favorable expresar a tasas altas de traducción. Los genes subregulados deben poder expresarse bajo condiciones de estrés o escasez de recursos para desempeñar sus funciones de cobertura. En estos contextos, expresar estos genes a tasas menores de traducción sería más conveniente al tener un más bajo costo energético que cubrir (Hausser et al., 2019).

Al hacer la prueba estadística de Mood de comparación de medianas, llegamos a la conclusión de que no podemos establecer una diferencia significativa entre las tasas de traducción de los genes subregulados y sobrerregulados ($p = 0.80013$ para la rpoBE672K, $p = 0.514$ para la SQ53). Esto sugiere que el cambio en la abundancia de transcripción tiene un mayor impacto sobre el cambio del tamaño del proteoma que una diferencia en las tasas de traducción de los grupos de genes expresados diferencialmente.

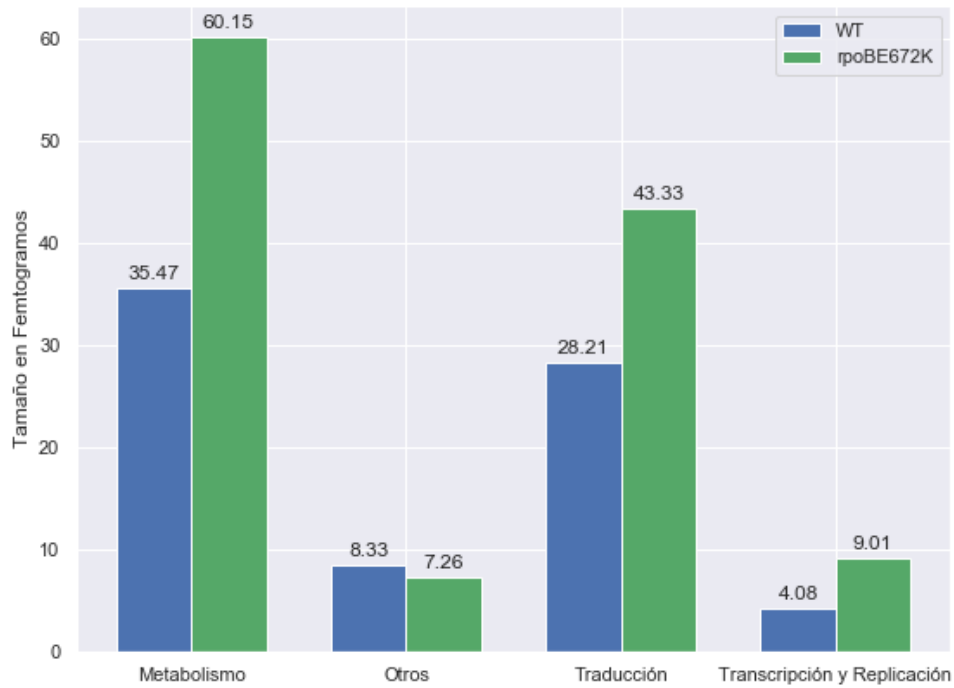


Figura 19. Distribución de cambios en el proteoma divididos en sectores para el experimento de rpoBE672K. Se indican cambio en el metabolismo, traducción, transcripción y replicación, y otros.

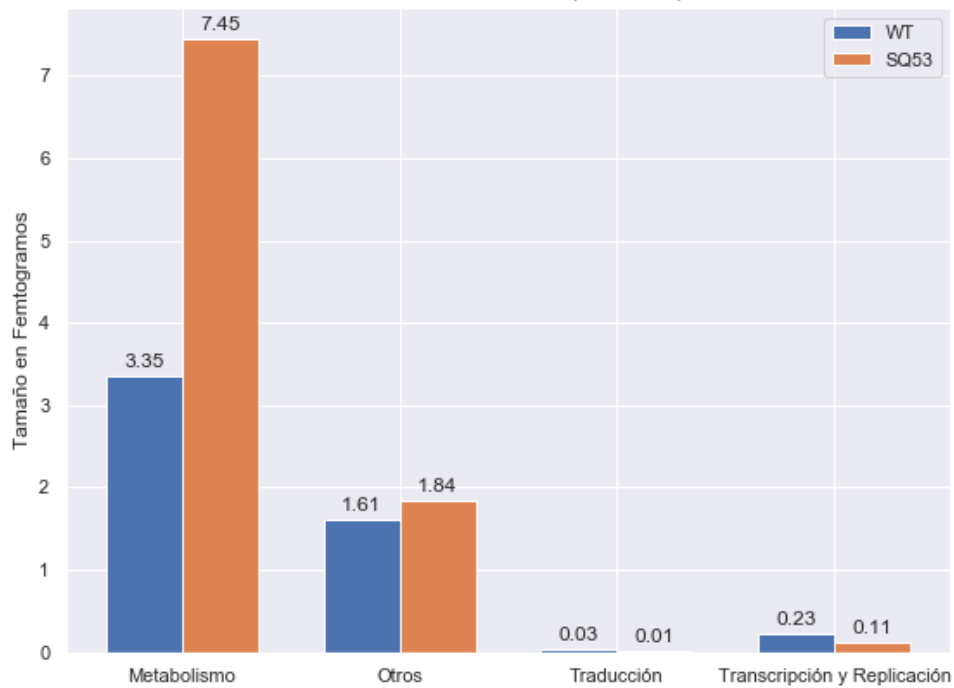


Figura 20. Distribución de cambios en el proteoma divididos en sectores para el experimento de SQ53. Se indican cambio en el metabolismo, traducción, transcripción y replicación, y otros.

En la reorganización del sector del metabolismo, en ambas mutantes hubo un incremento de alrededor del doble de recursos asignados. Considerando que en las condiciones de crecimiento en medio mínimo, este es un sector que en la cepa silvestre ya estaría recibiendo recursos con prioridad para suplir la necesidad de síntesis *de novo* de aminoácidos, nucleótidos, vitaminas etc. (Basan et al., 2015; Zhu et al., 2019), podríamos suponer que esto implica una ventaja para mejorar la capacidad de síntesis de proteínas. Esta ventaja podría estar ayudando a mantener saturado de sustratos a la maquinaria traduccional y sería un factor clave para mantener el aumento en la tasa de crecimiento. Esta ventaja estaría siendo adquirida a costa de un mayor costo energético que una vez que ocurre el cambio diáuxico, resulta en una menor adaptabilidad.

A nivel del sector de la traducción podemos ver que la diferencia entre lo observado en ambas mutantes es muy grande, esto es debido a que la proteína que contribuye a la mayoría del cambio en rpoBE672K es una de las que se traducen en mayor abundancia en *E. coli*, incluso en la cepa silvestre. Esta proteína es el factor de elongación EF-Tu, codificado por los genes *tufA* y *tufB*. El gen *tufA* también está expresado por encima de la cepa silvestre en las cuentas normalizadas por DESeq2 de la cepa mutante SQ53 pero las réplicas de la cepa silvestre tuvieron mucha variación (Figura 21). El algoritmo del software DESeq2 lo descarta como gen expresado diferencialmente debido a esta variación entre réplicas. Sin embargo, podría tratarse de un resultado falso negativo y no fue realizado un ensayo de validación de los resultados del análisis del transcriptoma con otros experimentos. Sería útil realizar una revisión adicional por Northern blot, por ejemplo.

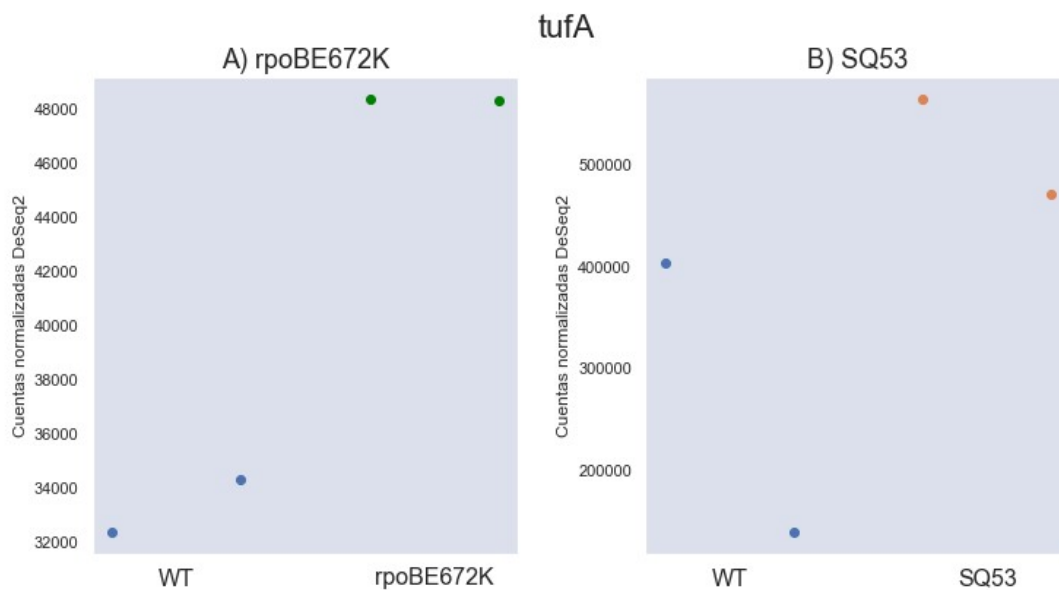


Figura 21. Cuentas normalizadas obtenidas con DeSeq2 de *tufA* (EF-Tu) entre réplicas en los experimento rpoBE672K y SQ53.

La ocurrencia de esta inconsistencia entre réplicas que descartó a *tufA* (EF-Tu) de la lista de genes expresados diferencialmente en SQ53 tuvo un beneficio inesperado para el análisis. Ya que hizo más fácil observar el decremento de aproximadamente 3 veces el tamaño del sector de la traducción. De 0.03 a 0.01 femtogramos para los genes expresados diferencialmente correspondientes a este sector (Figura 20).

Esta disminución corresponde a la subexpresión de *rmf*, un gen que codifica para la proteína RMF (*Ribosome Modulation Factor*). En la cepa silvestre, su acumulación en condiciones de crecimiento rápido solo ocurre hasta la transición a la fase estacionaria (Izutsu et al, 2001). Sin embargo, se ha observado que su síntesis puede ocurrir desde más temprano en el ciclo celular, su concentración es inversamente proporcional a la tasa de crecimiento en condiciones de medio mínimo y también puede observarse en la fase de crecimiento exponencial, no solo a la entrada a la fase estacionaria.

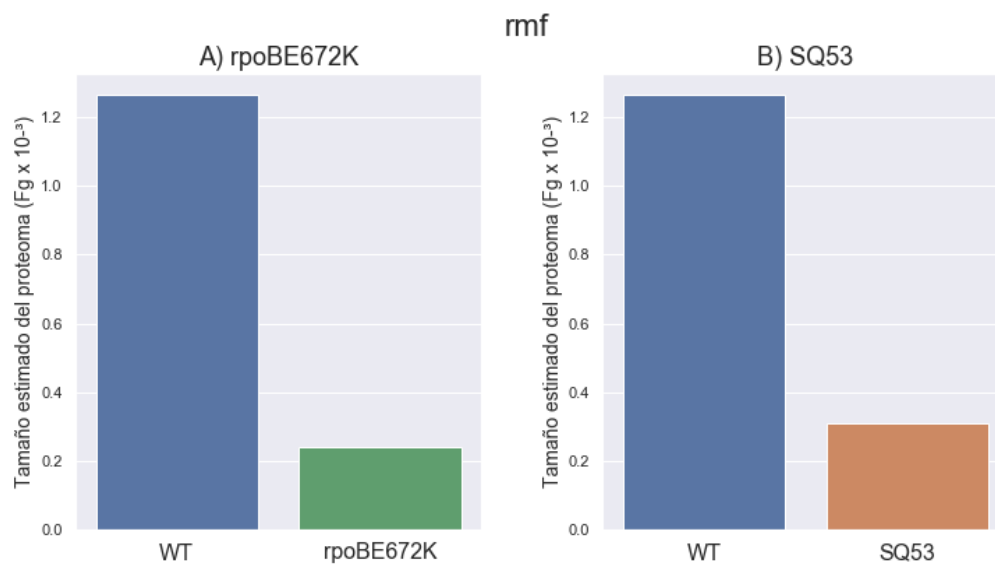


Figura 22. Comparación del proteoma estimado correspondiente a la expresión del gen *rmf* (RMF) en las mutantes *rpoBE672K* y *SQ53* con su cepa control.

La disrupción de la expresión de RMF resulta en la pérdida de formación de dímeros 100S de ribosomas y la subsecuente inviabilidad en la fase estacionaria (Yamagishi et al, 1993). Su mecanismo de regulación no está completamente definido, lo que se sabe es que *in vivo* puede expresarse en ausencia de RpoS y parece ser solo dependiente de la acumulación de (p)ppGpp (Izutsu et al, 2001). A pesar del aumento general en el sector de la traducción observado en *rpoBE672K*, en esta mutante, *rmf* también está subregulado y en proporciones muy similares.

Con el apoyo de diagramas de pareto, pasamos a hacer una revisión de los principales genes que representan el mayor incremento en la demanda de recursos y también cuáles fueron aquellos que representaron el mayor ahorro.

Diagrama de Pareto del Aumento del Proteoma - *rpoBE672K*
40 de 175 genes

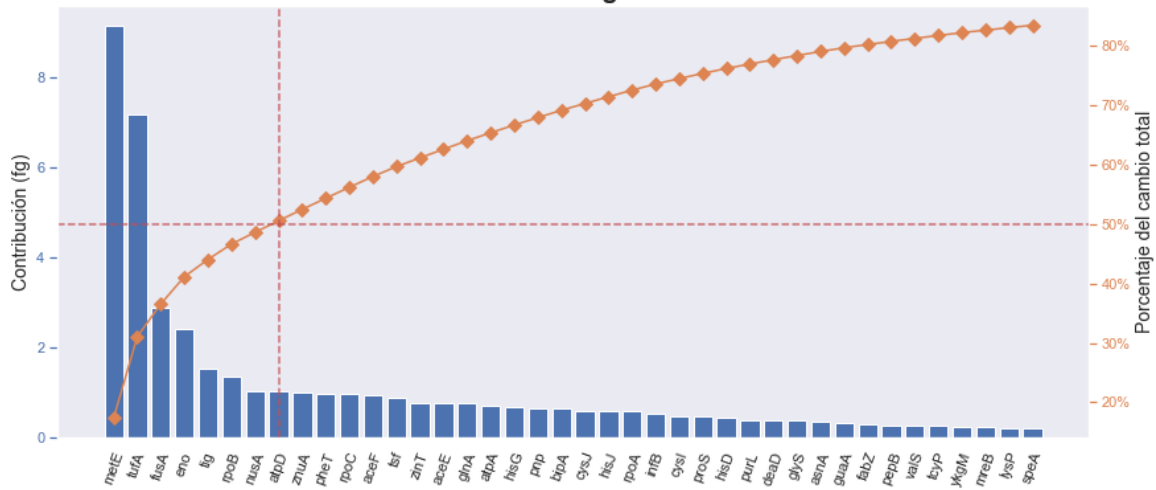


Figura 23. Diagrama de Pareto del aumento del proteoma en *rpoBE672K*. Se muestra con líneas punteadas en rojo la división correspondiente a los genes que aportan el 50% del aumento total. Estos genes, en orden de mayor a menor contribución, son: *metE*, *tufA*, *fusaA*, *eno*, *tig*, *rpoB*, *nusaA* y *atpD*.

El gen *metE* cumple una función muy importante en el metabolismo de *E. coli*. Codifica para una proteína muy conservada entre bacterias. Su función es catalizar la reacción final en la biosíntesis *de novo* de metionina, un aminoácido de iniciación de la traducción y precursor para otros aminoácidos (Deobald, 2020). También necesita de Zinc para su síntesis y actividad, lo que podría tener un efecto sinérgico con que *zntT* y *znuA* también estén en esta lista de genes sobrerregulados, ya que estos son transportadores de zinc (Gabbianellin et al., 2011). Se ha observado que la sobreexpresión de *metE* en mutantes de *E. coli* contribuye a mantener tasas mayores de crecimiento en presencia de acetato en el medio y también confiere otras ventajas, como la capacidad de crecer en una temperatura de hasta 45°C, la cual es una condición no-permisiva para *E. coli* (Mordukhova & Pan, 2013).

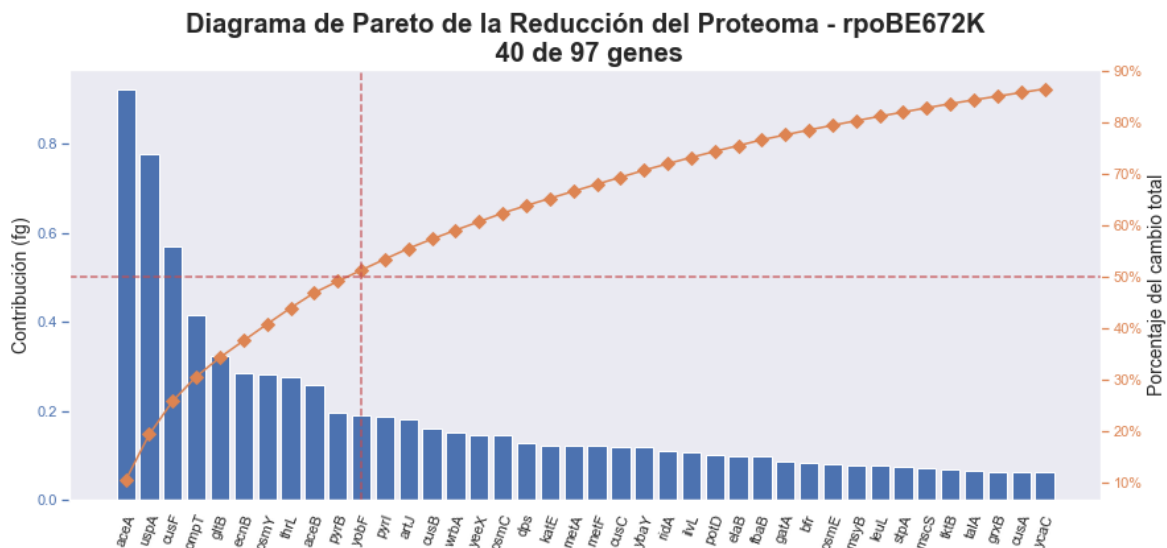


Figura 24. Diagrama de Pareto de la reducción del proteoma en rpoBE672K. Se muestra con líneas punteadas en rojo la división correspondiente a los genes aportan el 50% de la reducción total. Estos genes, en orden de mayor a menor contribución, son: *aceA*, *uspA*, *cusF*, *ompT*, *ecnB*, *osmY*, *thrL*, *aceB* y *pyrB*.

La proteína UspA forma parte de una familia que se activa en respuesta a la mayoría de los estreses celulares, la única condición que parece no activarla es el choque frío (Siegele, 2005). Se ha observado también que su síntesis se induce tan pronto la tasa de crecimiento sufre una caída por debajo de su máximo en un determinado medio, es esencial para sobrevivir diversos tipos de estrés por un periodo prolongado (Nystrom & Neidhardt, 1994). Dado este trasfondo, se esperaría que la proteína más bien estuviese sobrerregulada para contribuir en la respuesta al estrés. Que esté subregulada podría jugar un papel muy importante en su pérdida de capacidad de adaptación.

El gen *osmY* codifica para una proteína periplásmica que confiere resistencia a ambientes hiperosmóticos, como el que podría encontrarse en el sistema digestivo, uno de los ambientes donde se puede encontrar a *E. coli*. Es una función de prevención adaptativa (Yim & Villarejo, 1992).

El gen *aceB* codifica para la malato sintasa A y forma parte del operón de acetato que induce su expresión para facilitar el consumo de acetato en el medio (Maloy & Nunn, 1981), es coherente su reducción en esta mutante con su dificultad adquirida para adaptarse a las condiciones de acumulación de acetato. También de la misma ruta metabólica está presente el gen *aceA*, que codifica para la isocitrato liasa, enzima que cataliza el primer paso clave del ciclo del glioxilato.

La proteína Dps ha sido muy estudiada por su rol de regulación de otros genes en respuesta a estrés y es una proteína muy conservada. Se considera que tiene un rol crucial en la protección del DNA durante condiciones de escasez de nutrientes, estrés oxidativo y ácido (Calhoun & Kwon, 2011).

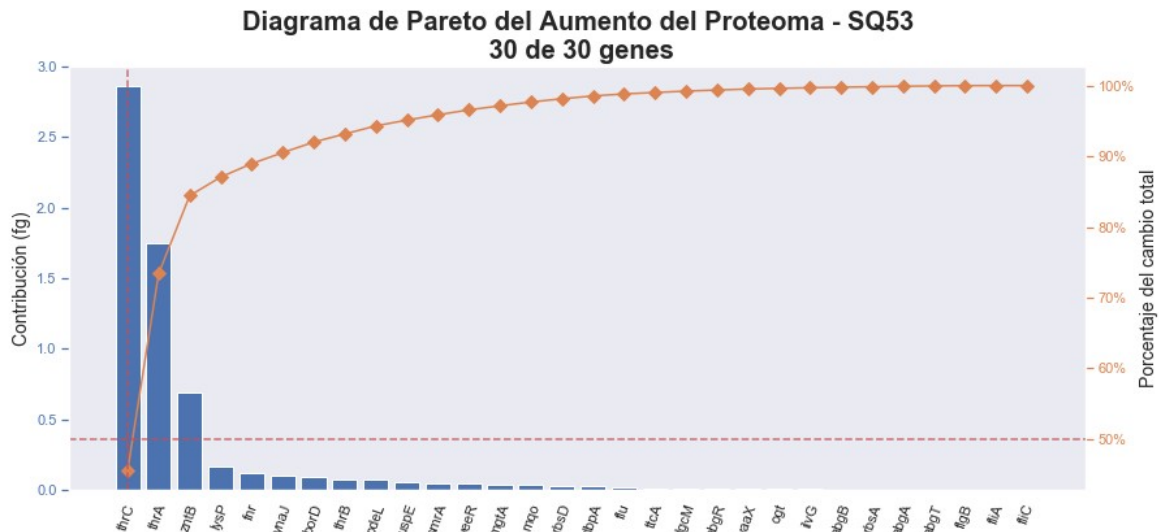


Figura 25. Diagrama de Pareto del aumento del proteoma en SQ53. Se muestra con líneas punteadas en rojo la división correspondiente a los genes aportan el 50% del aumento total. En este caso es un único gen: *thrC*. En conjunto con *thrA* y *zntB* estos aportan el 80% del aumento total.

Los dos principales contribuyentes en el aumento de la SQ53 son *thrC* y *thrA*, los cuales son genes que codifican para proteínas importantes para la biosíntesis de Treonina (Zhang et al, 2015). El tercer gen *zntB*, es un transportador de zinc que actualmente se considera que tiene funciones de importación de zinc, según bases estructurales (Gati et al., 2017). Su sobreexpresión podría estar siendo de utilidad para estimular la síntesis de metionina, sin embargo, una importación en exceso de zinc también puede tener efectos tóxicos.

Diagrama de Pareto de la Reducción del Proteoma - SQ53
30 de 74 genes

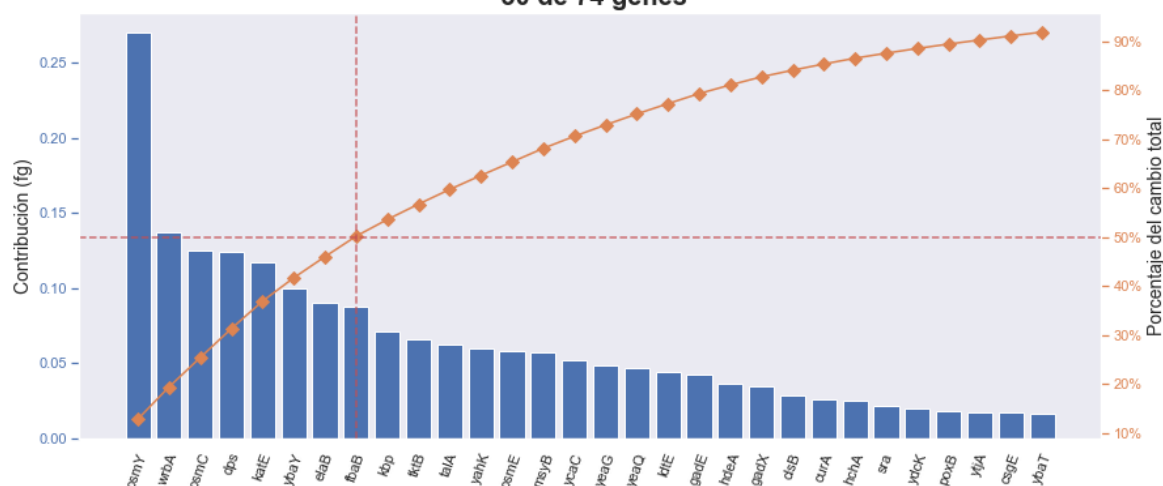


Figura 26. Diagrama de pareto de la reducción del proteoma en SQ53. Se muestra con líneas punteadas en rojo la división que señala cuáles genes aportan el 50% de la reducción total. Estos genes son, en orden de mayor a menor contribución: *osmY*, *wrbA*, *osmC*, *dps*, *katE*, *ybaY*, *elaB* y *fbaB*.

Los genes *osmY* y *osmC* de funciones relacionadas con anticipación de estrés hiperosmótico, aparecen también en las funciones subexpresadas para la SQ53. El gen *wrbA* codifica para la proteína WrbA y al igual que la proteína Dps, está altamente conservada. Desempeña funciones de respuesta a diversos tipos de estrés y de hecho se ha observado que está co-regulada con Dps (Natalello et al, 2007).

- **Análisis de interacción de genes expresados diferencialmente con StringDb**

Los siguientes mapas tienen la intención de mostrar las relaciones que existen de interacción proteína-proteína, co-expresión, co-ocurrencia a través de genomas relacionados y cercanía de distancia genómica. Se hicieron dos listas por cada experimento, una de genes subregulados y otra de genes sobrerregulados para cada experimento de las cepas mutantes estudiadas. En la descripción de la figura de cada mapa se indica el p-valor resultado de la prueba estadística de enriquecimiento hecha por la propia plataforma de StringDb (<https://string-db.org/>). Entre menor sea el *p*-valor, se puede tener una mayor confianza de que el número de interacciones entre las proteínas del grupo es mayor al que se esperaría de un conjunto aleatorio de genes del mismo tamaño, muestreado del mismo genoma. Es decir, un valor muy bajo de *p*-valor en los grupos de genes sobrerregulados y subregulados sustentaría que existe una alta probabilidad de que su expresión diferencial tenga un origen regulatorio en común. Entre más alto el p-valor, sería más probable que los genes expresados diferencialmente tengan distintos orígenes regulatorios. Las figuras 27 y 28, mostradas en la siguiente página, correspondientes a los genes subregulados y sobrerregulados de la mutante rpoBE672K y las figuras 29 y 30 de la mutante SQ53 tienen en todos los casos un p-valor muy inferior a 1.0e-16.

Los distintos colores de los nodos en los mapas ayudan a mostrar subgrupos con genes que tienen un alto grado de interacción. Estos subgrupos fueron generados con un algoritmo de agrupamiento utilizando un parámetro de número de grupos arbitrario. Los subgrupos encontrados son de utilidad para observar que es posible encontrar subgrupos que comparten relaciones de interacción más cercanas entre ellos que con el grupo en general. El análisis por i-modules de la siguiente sección es más apropiado para observar grupos de genes con funciones biológicas relacionadas.

Los genes resaltados en color rojo en la figura 30 corresponden a diversas respuestas a estrés, entre ellas estrés osmótico y estrés oxidativo. La mayoría de los genes en este grupo tienen una relación fuerte con Dps, un factor que tiene varias funciones descritas como regulador de genes de respuesta a estrés y también desempeña tareas de protección de la integridad del DNA en condiciones de estrés.

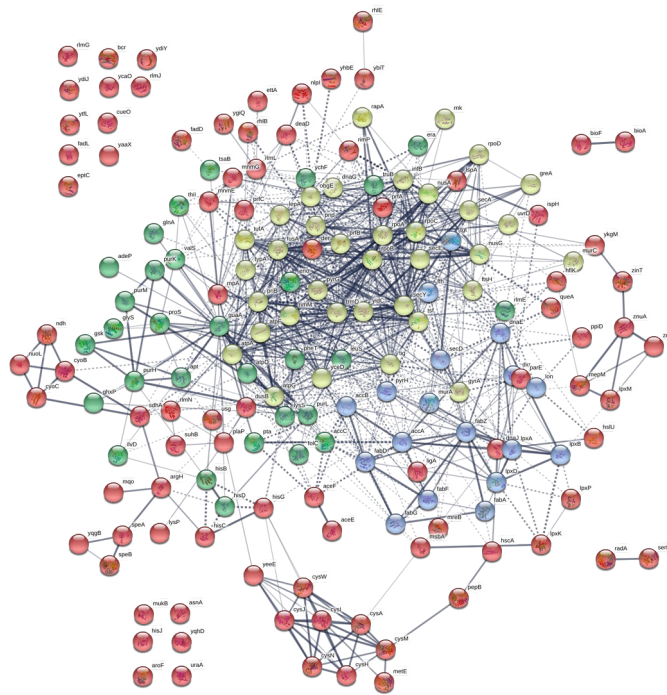


Figura 27. Mapa de confianza de interacción entre genes sobrerregulados en el experimento rpoBE672K. El resultado del análisis de enriquecimiento fue de un valor de $p < 1.0e-16$.

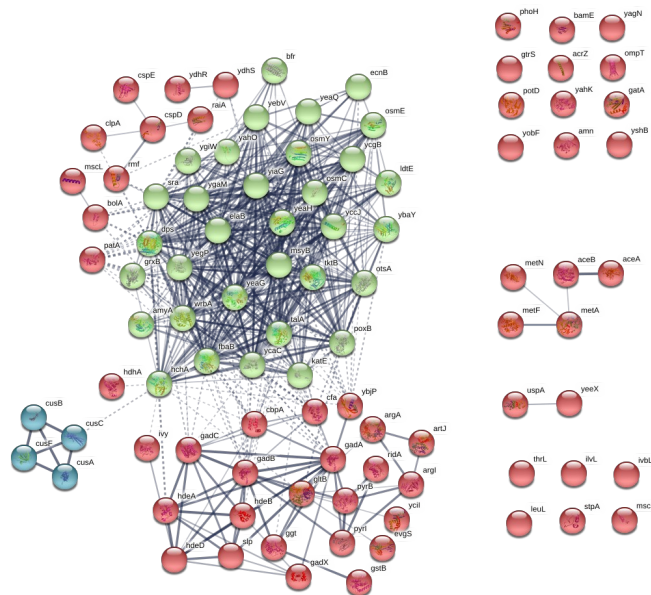


Figura 28. Mapa de confianza de interacción entre genes subregulados en el experimento rpoBE672K. El resultado del análisis de enriquecimiento fue de un valor de $p < 1.0e-16$.

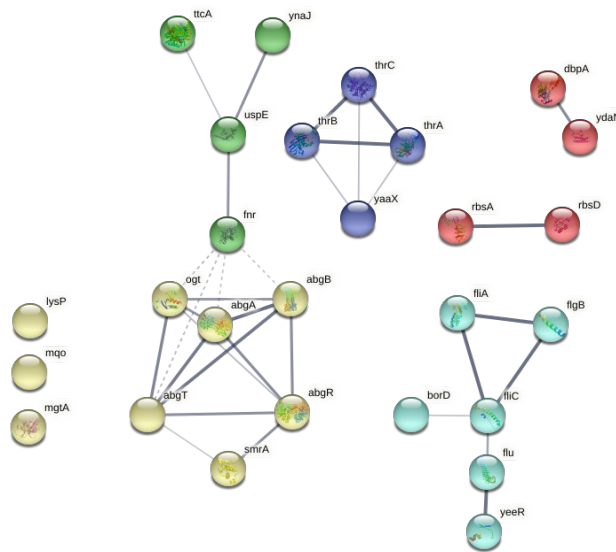


Figura 29. Mapa de confianza de interacción entre genes sobrerregulados en el experimento SQ53. El resultado del análisis de enriquecimiento fue de un valor de $p < 1.0e-16$.

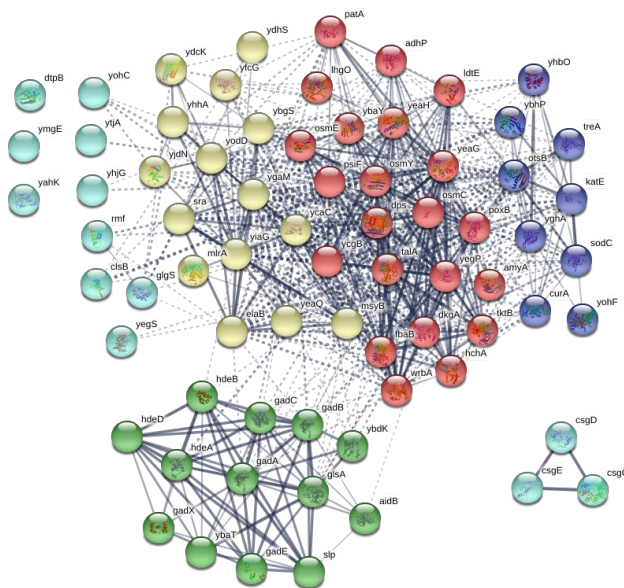


Figura 30. Mapa de confianza de interacción entre GEDs subregulados en el experimento SQ53. El resultado del análisis de enriquecimiento fue de $p < 1.0e-16$.

Los genes *rmf* y *sra* también aparecen en la figura 30 con interacciones relativamente débiles comparadas con las relaciones que existen en el grupo rojo que está más densamente relacionado. Ambos son genes que se activan como parte del mecanismo de inhibición de ribosomas en preparación a la fase estacionaria (Izutsu et al, 2001). El otro grupo de genes resaltado en color azul corresponde a genes relacionados con la respuesta a estrés ácido, son genes que

desempeñan funciones de captura de protones en el medio a través de integrarlos al metabolismo de aminoácidos.

Esta observación de agrupamiento de los genes en subgrupos funcionales co-regulados está mejor representada por el análisis por imodulones presentados en la siguiente sección.

Los resultados de la figura 31 parecen apoyar que el fenotipo de crecimiento acelerado no es resultado de la sobrerregulación de una selección específica de genes orientados al crecimiento sino más bien a la sobrerregulación de grupos de genes con funciones de orientación al crecimiento. Las diferencias entre los genes utilizados en cada mutante podrían estar relacionadas con la diferencia de su ventaja adquirida (rpoBE672K: ~25%, SQ53:~11%).

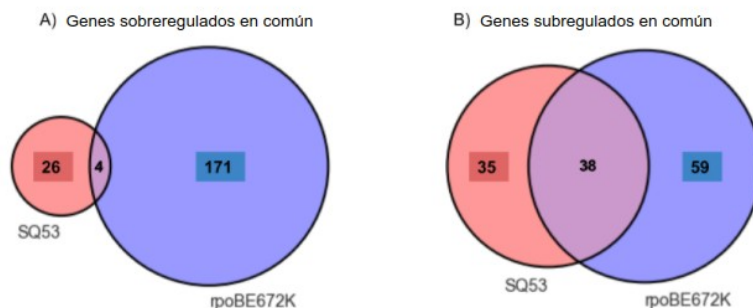


Figura 31. Diagrama de Venn de número de genes sobrerregulados y subregulados compartidos entre las cepas SQ53 y rpoBE672K.

Los genes sobrerregulados en común entre las mutantes son solo 4 y StringdB no encontró relaciones de interacción entre ellos. En la figura 32 se muestra que los genes subregulados en común entre las mutantes muy probablemente comparten funciones biológicas y un alto grado de interacción. Llama la atención que se nota claramente que pueden dividirse en dos grandes subgrupos densamente relacionados.

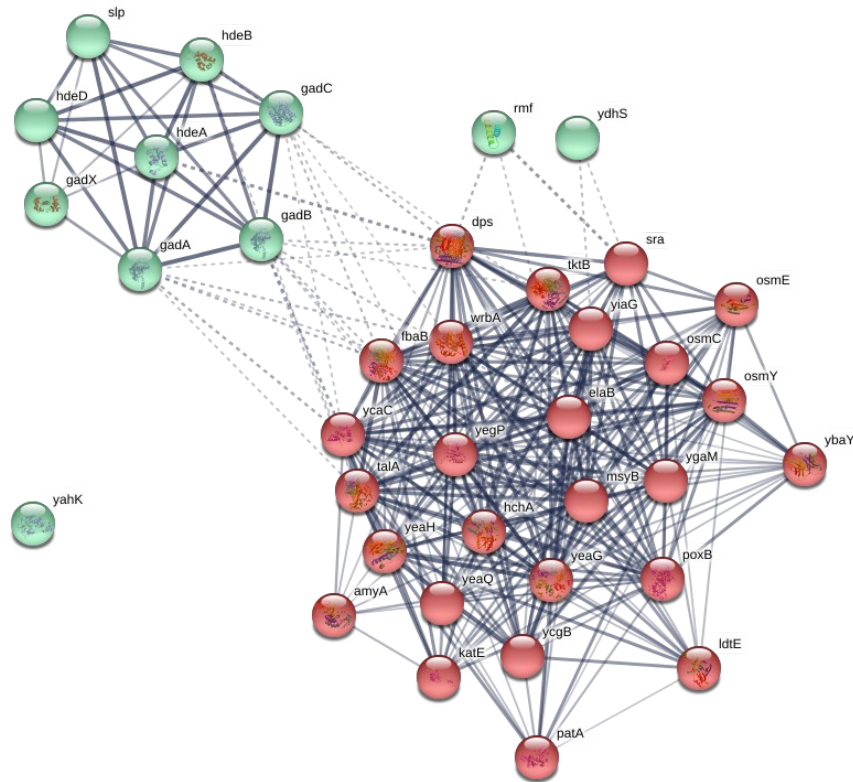


Figura 32. Mapa de redes de interacción entre genes subregulados en común entre el experimento SQ53 y el experimento rpoBE672K. El resultado del análisis de enriquecimiento fue de un valor de $p < 1.0e-16$. El color sirve para separar

El grupo coloreado en rojo contiene genes de respuesta a diversos estreses, como el gen *dps* que provee de protección a diversos tipos de estrés (oxidativo, por radiación UV y gamma, y toxicidad por iones metálicos) (Antipov et al, 2017), entre otros. En el grupo coloreado en verde la mayoría son genes de respuesta a estrés ácido, se encuentran los genes que conforman el sistema dependiente de glutamato GAD que está involucrado en el mantenimiento de la homeostasis del pH y son inducidos en condiciones de bajo pH. (Hommais et al, 2004)

- **Análisis de cambios en i-modulones relevantes para la expresión de proteínas heterólogas**

En la investigación publicada por Tan et al, 2020, estudiaron el patrón de expresión génica a través de estos i-modulones en cepas expresando proteínas heterólogas. Llegaron a la conclusión de que aquellas cepas que disminuyen su inversión en recursos al i-modulon de RpoS e invierten recursos en los otros modulones, siendo

el más importante de ellos, el de la traducción, lograban niveles de expresión más altos de las proteínas heterólogas.

Los i-modulones MetJ, ArgR, ArcA, Tyr-R y BCAA-2 agrupan genes con función de biosíntesis de aminoácidos. Los i-modulones CusR y CysB agrupan genes con función de mantenimiento de la homeostasis del estado redox intracelular, su sobreexpresión es necesaria al expresar proteínas heterólogas para contender con el mayor estrés redox que ocurre por la generación de intermediarios para la síntesis de proteínas. Los i-modulones His-tRNA y Traducción agrupan genes con función de síntesis de proteínas, asociados directamente a los componentes de la maquinaria traduccional. El i-modulon RpoH agrupa genes con función de plegamiento de proteínas, su sobreexpresión es provocada en situaciones de mayor demanda de síntesis de proteínas ya que también con ello incrementa la demanda de actividad de plegamiento. El i-modulon RpoS agrupa genes con diversas funciones de respuesta a estrés.

Si bien, menos genes de la lista completa fueron alineados a los i-modulones enlistados por Tan et al, 2020 para el experimento de la mutante SQ53 (figura 35) que para la mutante rpoBE672K (figura 34), la respuesta en ambas mutantes comparte el patrón de subexpresar genes del i-modulon RpoS y sobreexpresar otros i-modulones que serían necesarios para contender con una mayor demanda de síntesis de proteínas. Esto es coherente con su fenotipo de tasas de crecimiento más altas a la silvestre a costa de perder capacidad de respuesta a estrés.

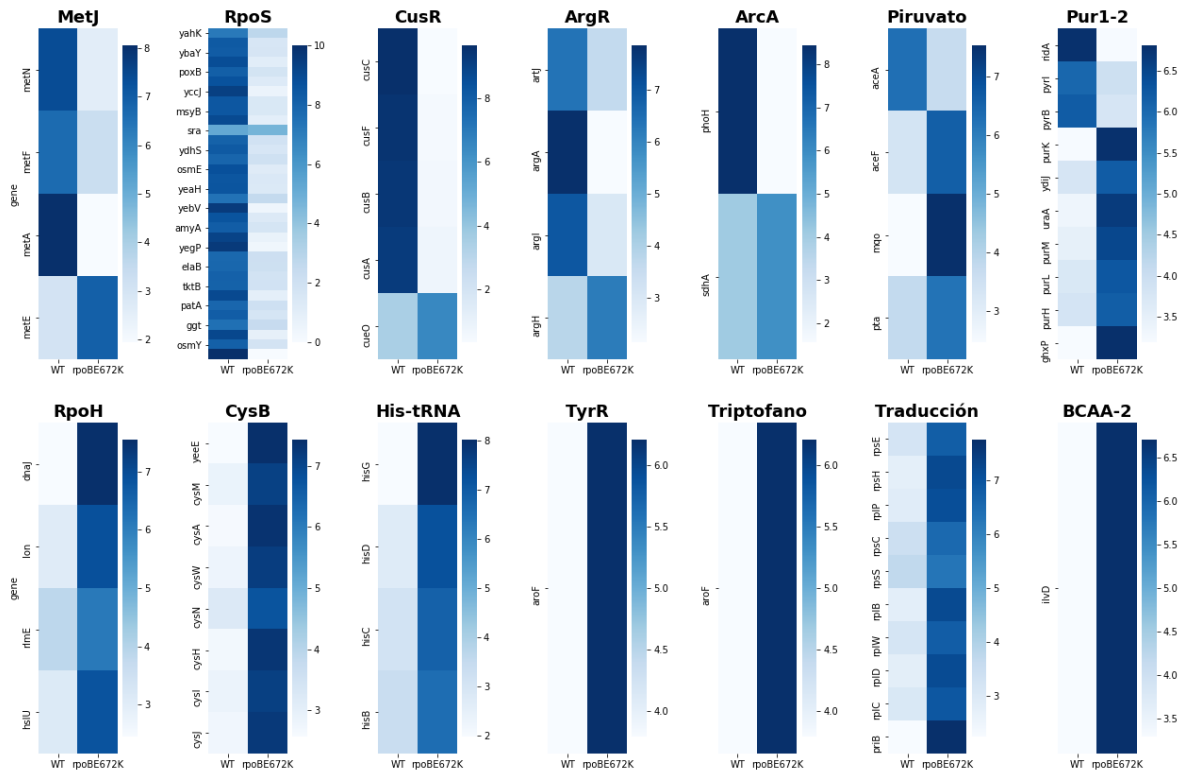


Fig 34. Cambios en i-modulones de relevancia para la expresión de proteínas heterólogas, WT vs. mutante *rpoBE672K*.

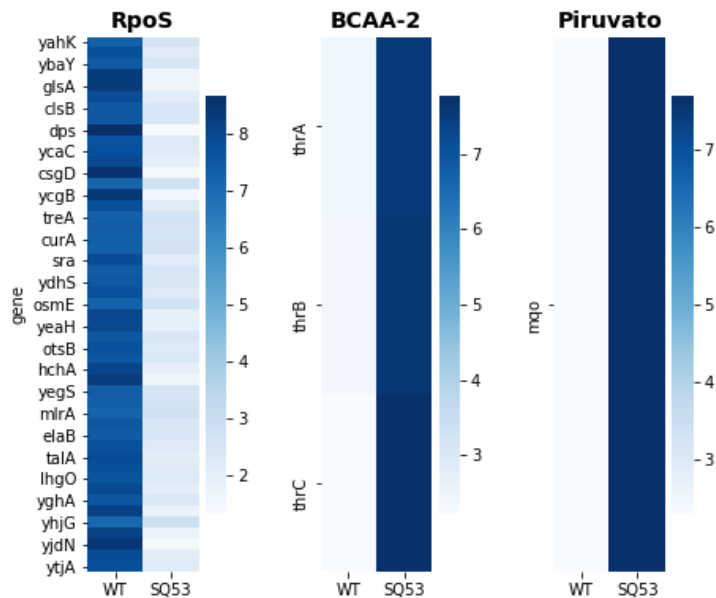


Fig 35. Cambios en i-modulones de relevancia para la expresión de proteínas heterólogas, WT vs. mutante *SQ53*.

Sin embargo, esto no necesariamente significa que tengan una mayor capacidad de producción de proteínas heterólogas. Experimentos utilizando cepas con operones

eliminados y midiendo la expresión de proteínas heterólogas en condiciones de crecimiento acelerado concluyeron que estos parámetros tenían una relación inversamente proporcional (Kim et al, 2021). Esto también fue observado por la medición de un menor presupuesto proteómico disponible en las mutantes por experimentos realizados en nuestro grupo de investigación (Hidalgo et al, 2021). Puede ser que el patrón de la expresión génica de las mutantes con crecimiento acelerado, originado en un secuestro más fuerte de la RNA polimerasa afecta no solo la disponibilidad de la RNA polimerasa para la expresión de proteínas de respuesta a estrés sino también su disponibilidad para la expresión de proteínas heterólogas.

Discusión

E. coli silvestre utiliza estrategias de reorganización de recursos y adaptación del metabolismo como medio para mejorar su adaptabilidad en condiciones donde no puede incorporar aminoácidos del ambiente y la prioridad es sintetizarlos *de novo* aprovechando el carbono disponible en el medio. El metabolismo de sobre-flujo, en conjunto con la activación de las rutas de síntesis de proteínas metabólicas por la acumulación de (p)ppGpp, desvían recursos hacia la síntesis de aminoácidos en condiciones de medio mínimo para resolver su falta de disponibilidad en el medio. (Basan et al., 2015; Zhu et al., 2019).

Al mismo tiempo, la respuesta estricta disminuye rápidamente la síntesis de rRNA a la vez que la acumulación de (p)ppGpp inhibe el inicio de la traducción a través de competir con GTP por la unión con el factor de inicio de la traducción IF2 (Wagner, 2010). Gradualmente, también se controla la fracción de ribosomas activos a través de la acumulación de RMF (*Ribosome Modulation Factor*), la cual es dependiente de la acumulación de (p)ppGpp (Izutsu et al, 2001). Otros factores también contribuyen a la inhibición del ensamblaje de ribosomas activos en condiciones de medio mínimo como HPF (Hibernation Promotion Factor), YfiA y RsfA (Dai et al., 2016).

La adaptabilidad de *E. coli* en medio mínimo mejora al utilizar estas medidas que restringen la tasa de crecimiento. Esta restricción representa una adaptación del metabolismo y maquinaria traduccional que prepara a la célula para entrar en un estado de bajo consumo energético, además de que se libera RNA polimerasa de

los operones ribosomales y otras funciones de crecimiento. En este estado se prolonga el tiempo en el que la población puede esperar un cambio repentino en la disponibilidad de recursos y también incrementa la adaptabilidad al haber más RNA polimerasa disponible para expresar funciones de cobertura.

Como se mencionaba en la introducción y antecedentes, se conoce que organismos que viven en ambientes fluctuantes conservan la capacidad de acumular (p)ppGpp y realizar cambios globales en la expresión génica de acuerdo a esta señal. La multiplicidad de operones ribosomales juega un papel importante dentro de ese esquema de regulación y está igualmente conservada en organismos adaptados a ambientes fluctuantes.

Puede ser que en la mutante SQ53 esté involucrado otro factor que contribuye a hacer más fino el control de la respuesta que en la WT, siendo también el factor que origina su ventaja en medio mínimo: la heterogeneidad en presión regulatoria de la respuesta estricta sobre los operones. Dos de los operones que permanecen en el genoma de la mutante, *rrnE* y *rrnH*, han sido caracterizados por separado como los operones con mayor actividad, tanto en presencia como en ausencia de acumulación de (p)ppGpp (Maeda et al., 2015).

Utilizando lecturas de RNA-Seq no podemos cuantificar la actividad de transcripción del rRNA porque su depleción es parte del proceso de construcción de la biblioteca para secuenciación. Sin embargo, podemos considerar que modelos matemáticos han previsto que durante la respuesta estricta, la liberación de RNA polimerasa de los operones ribosomales habilita la expresión oportuna de genes de respuesta a diversos estreses. Estos grupos funcionales están regulados por los factores de expresión RpoS, RpoN y RpoH principalmente (Mauri et al, 2014). Podría pensarse que reducir el número de operones ribosomales tendría el efecto de facilitar la liberación de la RNA polimerasa oportunamente y por lo tanto sobrerregular la expresión de genes de respuesta a estrés. Sin embargo, el efecto observado en ambas mutantes es el contrario (i-modulon RpoS en figuras 34 y 35), esto puede deberse a que a pesar de que los operones ribosomales en el genoma se han reducido, los operones que quedan son de los que tienen mayor actividad y por ello esté más impedido liberar RNA polimerasa a tiempo para expresar los genes de respuesta a estrés.

Resultados de experimentos de cuantificación de concentración de ribosomas marcados con GFP muestran que si existe variación en contenido ribosomal entre las mutantes de crecimiento rápido en comparación con la WT, sin embargo esta variación es ligera y podría no ser suficiente para justificar la diferencia en la tasa de crecimiento (Hidalgo et al, 2021). Por esta razón y los resultados de la comparación de los perfiles transcripcionales, proponemos que la ventaja adquirida por la cepa SQ53, sea en mayor parte debido a un uso diferencial de la RNA polimerasa ocasionada por la selección particular de operones ribosomales en su genoma.

Adicionalmente, podemos discutir que los niveles más bajos de expresión de *rmf* detectados en la mutante SQ53, también detectados en *rpoBE672K* (figura 22), podrían ser señal de una fracción mayor de ribosomas activos. Trabajos previos han demostrado que en medio mínimo la fracción de ribosomas activos constituyen el mayor determinante de la tasa de crecimiento, junto con la abundancia de complejos ternarios (Dai et al., 2016). La suma de contar con una concentración más grande de complejos ternarios cargados con amino ácidos, sugerida por la expresión aumentada de *tufA* en *rpoBE672K* más la expresión aumentada de rutas de síntesis de aminoácidos en SQ53, podría ser señal de que en las mutantes exista una fracción mayor de ribosomas activos con los sustratos suficientes para lograr una eficiencia traduccional mayor. Aunque tendría que hacerse un experimento enfocado específicamente en medir diferencias en la fracción de ribosomas activos en las mutantes, una consecuencia esperada de esto sería un incremento de la tasa de crecimiento sin necesariamente contar con una concentración mucho mayor de ribosomas.

En un medio donde existe abundancia de aminoácidos, no ocurre una acumulación de (p)ppGpp que active un desvío de carbono hacia la síntesis de aminoácidos ya que estos son incorporados directamente desde el medio. El metabolismo fermentativo o de sobreflujo tampoco es utilizado y más bien el respiratorio es utilizado en su lugar, ya que aunque tiene un costo proteómico más alto, este será más fácil de sostener al incorporar los aminoácidos del medio y a futuro representará una ventaja de mayor eficiencia energética (Basan et al., 2015).

En medio rico tampoco ocurre una limitación estricta de la síntesis de rRNA y tampoco del inicio de la traducción. En estas condiciones el factor limitante de la

velocidad de crecimiento se vuelve la capacidad de transcripción de rRNA y no la capacidad de síntesis *de novo* de aminoácidos junto con la fracción activa de ribosomas, como es el caso en medio mínimo. Esta es la explicación que proponemos por la cual la ventaja de crecimiento no ocurre en medio rico para la mutante SQ53 que en estas condiciones de abundancia de nutrientes estaría en desventaja de capacidad de transcripción de rRNA frente a la silvestre (Quan et al., 2015).

Para discutir el origen de la disminución de la adaptabilidad a las fluctuaciones en el ambiente. En este caso, hablando del cambio diauxico, hay que considerar las consecuencias del incremento de la tasa de crecimiento. Ya que estas condiciones imponen demandas sobre el metabolismo que necesitan una respuesta de expresión génica y distribución del proteoma que podrían no estar presentes en las mutantes.

La estimación del proteoma total de las mutantes, en ambos casos, es mayor al del proteoma total de la silvestre (figuras 15 y 16); sin embargo, experimentos de medición del presupuesto proteómico realizados para un proyecto relacionado dentro de nuestro grupo de investigación (Hidalgo et al, 2021), muestran que las mutantes no tienen un mayor presupuesto disponible.

El transcriptoma podría no estar siendo traducido en su totalidad, en otras palabras, la RNA polimerasa podría estar siendo sobre utilizada en los genes sobrerregulados sin que exista un presupuesto para traducirlos óptimamente. Al mismo tiempo, conforme progresa el crecimiento y se consumen los recursos del medio, comenzarán a acumularse condiciones de distintos estreses: la acumulación de la generación de intermediarios NADPH, la acumulación de acetato en el medio como consecuencia del consumo de glucosa en el metabolismo de sobreflujo, una mayor demanda de transporte de nutrientes y también la demanda de proteínas chaperonas incrementará. Sin una adecuada disponibilidad de la RNA polimerasa para expresar los genes de respuesta a estos estreses, la adaptabilidad se verá disminuida.

Genes que asisten en el mantenimiento de la homeostasis osmótica, el mantenimiento del balance redox para prevenir la acumulación de metales en estado tóxico, el amortiguamiento de pH intracelular, el abastecimiento de funciones

de plegamiento y genes que protegen la integridad del DNA en estas condiciones, se expresan como parte de la respuesta estricta y muchos de ellos son regulados por RpoS, RpoN o RpoH para contender con esta exigencia (Tan et al., 2020).

En ambas mutantes, el i-modulon de respuesta a estrés RpoS se ve impactado en la disminución de su expresión en ambos casos (figuras 34 y 35). Al mismo tiempo, en condiciones de escasez de aminoácidos, la respuesta estricta normalmente se encargaría de restringir la tasa de crecimiento a través de disminuir la eficiencia de síntesis de transcripción de rRNA, el inicio de la traducción y promover la dimerización de ribosomas activos, ayudando a aliviar el incremento de la demanda sobre proteínas de respuesta a estrés. Como se comentaba anteriormente, en las mutantes la eficiencia de síntesis de transcripción de rRNA podría ser mayor debido a sus mutaciones y también la inhibición de ribosomas activos podría tener como consecuencia un menor alivio del estrés.

Adicionalmente, en la búsqueda de diferencia en la mediana de las longitud de los grupos de genes expresados diferencialmente, encontramos que los genes sobrerregulados tenían una mediana mayor a la de los genes subregulados en la mutante rpoBE672K con un p-valor mucho menor a 0.05 (Figura 13); en la mutante SQ53 el p valor fue apenas menor a 0.1 (Figura 14) con lo que no se puede decir con la misma confianza que para esta mutante las medianas sean diferentes, pero sí se observa una tendencia a serlo. Ésta diferencia podría tener un peso importante sobre la capacidad de adaptabilidad pues genes más largos tienen un costo bioenergético más alto, tanto a nivel de la transcripción, como a nivel de la traducción (Lynch & Marinov, 2015). En un contexto de medio mínimo donde ahorrar recursos es la estrategia más exitosa de supervivencia a largo plazo, crear una demanda energética más alta sería contraproducente para la adaptación.

A pesar de que ambas cepas pierden adaptabilidad, el grado en que son afectadas es muy distinto, siendo más severo el efecto en la mutante SQ53 que en la rpoBE672K. Esto también se refleja en que el incremento en la tasa de crecimiento adquirido es mayor para la rpoBE672K (~25%) que para la SQ53 (~11%). Si bien ambas mutantes coinciden en varios genes subregulados, el número de coincidencias en genes sobrerregulados es mucho menor (figura 31). De esto podemos especular que tal vez la mayor ventaja de la cepa rpoBE672K sobre la

SQ53 se deba a una diferencia en la selección de genes sobrerregulados. Para responder esta pregunta sería necesario hacer análisis adicionales de comparación de costo bioenergético entre los grupos de genes expresados diferencialmente además de obtener una lista de genes generada con mayor confianza a partir de réplicas con mayor grado de reproducibilidad en la cepa SQ53.

Los resultados de análisis de interacción utilizando la herramienta de stringDb muestran que los grupos de genes subregulados en común entre las mutantes tienen un alto número de interacciones de distintos tipos, uno de estos tipos de interacciones es la cercanía en su ubicación en el genoma (figura 32). Se ha observado que genes cercanos en el genoma tienden a estar co-regulados (Pannier et al, 2017). Puede ser que el patrón de subexpresión en común entre las mutantes pueda atribuirse a esta observación, donde los genes subregulados, que se regulan positivamente como parte del mecanismo de adaptación para limitar el crecimiento, reciben menor atención de la RNA polimerasa debido a que esta se encuentra secuestrada en los operones ribosomales y los genes sobrerregulados. Este secuestro tendría razón para ocurrir en ambas mutantes por orígenes distintos pero con consecuencias similares. A pesar de la baja coincidencia específica, gen por gen, de los genes sobrerregulados (figura 31), son similares las categorías las que se asocian los genes sobrerregulados y subregulados en ambos casos, como se ha discutido en los párrafos anteriores. Esta similitud general coincide con ubicar el origen del mecanismo de reorganización de recursos en un uso diferencial de la RNA polimerasa en ambas mutantes, donde la RNA polimerasa no está siendo reclutada para sobreexpresar genes en específico sino grupos funcionales orientados al crecimiento y no subexpresando genes en específico sino grupos funcionales orientados a la respuesta a estrés o cobertura.

Conclusiones

Nuestros resultados indican que las mutantes SQ53 y rpoBE672K causan una reorganización del proteoma de *E. coli* originado en un uso diferencial de la RNA polimerasa, enfocando la utilización de recursos a funciones de crecimiento a costa de una menor adaptabilidad por la subexpresión de funciones de cobertura. El fenotipo de crecimiento acelerado impone una demanda de mantenimiento de la homeostasis celular y respuesta a estrés que no puede ser suplido por esta configuración del proteoma en las condiciones de crecimiento de medio mínimo.

A su vez, estas proteínas asociadas al crecimiento pertenecen a grupos de alto costo energético, añadiendo otro factor que podría estar contribuyendo a su menor adaptabilidad. Su sobreexpresión, vía un uso diferencial de la RNA polimerasa, limita fuertemente la expresión de funciones de cobertura, que serían necesarias para contender con el estrés adicional. Así mismo, en la WT la limitación de la expresión de estas funciones prepara la distribución de recursos durante fases de diauxia y controla la tasa de crecimiento durante la fase estacionaria en espera de un cambio en la disponibilidad de recursos en el ambiente. Sin la limitación de la expresión de funciones de crecimiento y la expresión oportuna de funciones de cobertura a través de un uso normal de la RNA polimerasa, las mutantes no logran adaptarse oportunamente al cambio en estas condiciones.

La mutación sobre la RNA polimerasa en la mutante rpoBE672K parece tener el efecto con mejores resultados sobre el crecimiento debido a diferencias específicas en los grupos de genes sobrerregulados respecto a la SQ53, ya que en los grupos de genes subregulados tienen un alto grado de similitud sobre varios genes de alta influencia sobre el fenotipo. Adicionalmente, podríamos especular que conserva una mayor flexibilidad de control de uso de RNA polimerasa sobre sus operones ribosomales en comparación con la SQ53, que tiene menor heterogeneidad en la actividad individual de sus operones ribosomales restantes. Podríamos concluir que la cepa rpoBE672K representa un mejor modelo para continuar el desarrollo de cepas adaptadas para una mayor capacidad de síntesis de proteínas a costa de una menor adaptabilidad a cambios en el medio y que la RNA polimerasa es un blanco útil para la ingeniería de cepas mejoradas.

Perspectivas

Realizar perfiles transcripcionales por RNA-Seq de la cepa rpoBE672K en nuestro laboratorio podría ser muy útil para afinar la detección de similitudes y diferencias de expresión diferencial entre las mutantes con el objetivo de profundizar sobre el origen del mayor grado de ventaja adquirida en velocidad de crecimiento por la rpoBE672K. Otra alternativa para lograr el mismo objetivo podría ser aprovechar la estrategia de normalización de cepas propuesta por Sastry et al., 2019 donde lograron normalizar la expresión de la cepa BW25113 para hacerla comparable con una mutante construida desde la cepa MG1655 de *E. coli*; en este caso la estrategia tendría el objetivo de separar el ruido por las diferencias incurridas por el método de secuenciación entre laboratorios que hace que ni las cepas mutantes ni las silvestres de distintos laboratorios sean comparables entre ellas y solo puedan compararse dentro de sus mismos grupos experimentales.

Otro experimento que sería conveniente repetir sería el RNA-Seq de la cepa SQ53 y WT en nuestro laboratorio. Podría aprovecharse la experiencia ganada con este proyecto para modificar los parámetros de secuenciación y hacerlo a una menor profundidad de biblioteca. De esta manera se probaría si la reducción del ruido capturado contribuye a la reproducibilidad entre réplicas. De los resultados de este experimento podría confirmarse la especulación en este proyecto sobre la existencia de falsos negativos importantes en la lista de genes expresados diferencialmente de la cepa SQ53, como es el caso de *tufA* más otros más que pudieron haber dado información útil para la discusión. Adicionalmente también podrían hacerse ensayos de cuantificación específica de proteínas seleccionadas de estas listas para validar con mayor certeza sus niveles de expresión.

Un experimento más de continuidad a este proyecto podría ser indagar si el patrón de genes sobregulados largos sobre genes subregulados más cortos está presente también en otras cepas con fenotipos similares en condiciones de crecimiento de medio mínimo. De este modo podríamos aportar un detalle más a los principios de repartición de recursos en estos sistemas si es que el patrón está conservado.

Referencias bibliográficas

- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Antipov, S. S., Tutukina, M. N., Preobrazhenskaya, E. V., Kondrashov, F. A., Patrushev, M. V., Toshchakov, S. V., Dominova, I., Shvyreva, U. S., Vrublevskaia, V. V., Morenkov, O. S., Sukharicheva, N. A., Panyukov, V. V., & Ozoline, O. N. (2017). The nucleoid protein Dps binds genomic DNA of *Escherichia coli* in a non-random manner. *PLoS one*, 12(8), e0182800. <https://doi.org/10.1371/journal.pone.0182800>
- Backman TWH, Girke T (2016). "systemPipeR: NGS workflow and report generation environment." *BMC Bioinformatics*, 17(1). doi: [10.1186/s12859-016-1241-0](https://doi.org/10.1186/s12859-016-1241-0).
- Basan, M., Hui, S., Okano, H. *et al.* (2015). Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528, 99–104. <https://doi.org/10.1038/nature15765>
- Bosdriez E., Molenaar D., Teusink B., Bruggerman F. (2015). How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS Journal*. 282: 2029-2044.
- Bremer, H., & Dennis, P. P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* , 3(1). doi:10.1128/ecosal.5.2.3
- Castiñeiras T. S., Williams S. G., Hitchcock A .G., Smith D.C. (2018). *E. coli* strain engineering for the production of advanced biopharmaceutical products. *FEMS Microbiology Letters*, 365(15). <https://doi.org/10.1093/femsle/fny162>
- Calhoun, L. N., & Kwon, Y. M. (2011). Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in *Escherichia coli*: a review. *Journal of applied microbiology*, 110(2), 375–386. <https://doi.org/10.1111/j.1365-2672.2010.04890.x>
- Condon, C., French, S., Squires, C., & Squires, C. L. (1993). Depletion of functional ribosomal RNA operons in *Escherichia coli* causes increased

expression of the remaining intact copies. *The EMBO journal*, 12(11), 4305–4315.

- Condon, C., Liveris, D., Squires, C., Schwartz, I., & Squires, C. L. (1995). rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rrn* inactivation. *Journal of bacteriology*, 177(14), 4152–4156. <https://doi.org/10.1128/jb.177.14.4152-4156.1995>
- Condon, C., Squires, C., & Squires, C. L. (1995). Control of rRNA transcription in *Escherichia coli*. *Microbiological reviews*, 59(4), 623–645.
- Dai, X., & Zhu, M. (2020). Coupling of Ribosome Synthesis and Translational Capacity with Cell Growth. *Trends in biochemical sciences*, 45(8), 681–692. <https://doi.org/10.1016/j.tibs.2020.04.010>
- Dai, X., Zhu, M., Warren, M., Balakrishnan, R., Patsalo, V., Okano, H., Williamson, J. R., Fredrick, K., Wang, Y.P., Hwa, T. (2016). Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nature microbiology*. 2 (December): 16231.
- Deobald, D., Hanna, R., Shahryari, S. et al.(2020). Identification and characterization of a bacterial core methionine synthase. *Sci Rep* 10, 2100. <https://doi.org/10.1038/s41598-020-58873-z>
- Ebrahim, A., Brunk E., Tan, J., O'Brien, E. J., Kim, D., Szubin, R., Lerman, J. A. et al. (2016). Multi-Omic Data Integration Enables Discovery of Hidden Biological Regularities. *Nature Communications* 7 (October): 13091.
- Gabbianelli, R., Scotti, R., Ammendola, S., Petrarca, P., Nicolini, L., & Battistoni, A. (2011). Role of ZnuABC and ZinT in *Escherichia coli* O157:H7 zinc acquisition and interaction with epithelial cells. *BMC microbiology*, 11, 36. <https://doi.org/10.1186/1471-2180-11-36>
- Gati, C., Stetsenko, A., Slotboom, D. J., Scheres, S., & Guskov, A. (2017). The structural basis of proton driven zinc transport by ZntB. *Nature communications*, 8(1), 1313. <https://doi.org/10.1038/s41467-017-01483-7>
- Gorochofski, T. E., & Ellis T. 2018. Designing Efficient Translation. *Nature Publishing Group* 36 (10). *Nature Publishing Group*: 934–35. [doi:10.1038/nbt.4257](https://doi.org/10.1038/nbt.4257). 33
- Gyorfy, Z., Draskovits, G., Vernyik, V., Blattner, F. F., Gaal, T., & Gyorgy P. (2015). Engineered Ribosomal RNA Operon Copy-Number Variants of *E. Coli*

Reveal the Evolutionary Trade-Offs Shaping rRNA Operon Number. *Nucleic Acids Research* 43 (3): 1783–94.

- Hausser, J., Mayo, A., & Alon, U. (2019). Central dogma rates and the trade-off between precision and economy. *Nature Communications*. <https://doi.org/10.1038/s41467-018-07391-8>
- Hommais, F., Krin, E., Coppée, J. Y., Lacroix, C., Yeramian, E., Danchin, A., & Bertin, P. (2004). GadE (YhiE): a novel activator involved in the response to acid environment in *Escherichia coli*. *Microbiology (Reading, England)*, 150(Pt 1), 61–72. <https://doi.org/10.1099/mic.0.26659-0>
- Hidalgo, D., Martinez, C., Jiménez, J., Palsson B., Utrilla, J. (2021). Ribosomal RNA gene deletion promote faster growth with a fitness trade-off. *Biorxiv*. <https://doi.org/10.1101/2021.08.01.454633>
- Izutsu, K., Wada, A., Wada, C. (2001). Expression of ribosome modulation factor (RMF) in *Escherichia coli* requires ppGpp. *Genes to cells*. 6(8). <https://doi.org/10.1046/j.1365-2443.2001.00457.x>
- Jiménez-Jacinto V, Sanchez-Flores A and Vega-Alvarado L (2019) Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis. *Front. Genet.* Vol 11, pg 279. <https://doi.org/10.3389/fgene.2019.00279>
- Jong, Hidde de, Johannes Geiselmann, and Delphine Ropers. (2017). Resource Reallocation in Bacteria by Reengineering the Gene Expression Machinery. *Trends in Microbiology*. Elsevier Ltd. doi:10.1016/j.tim.2016.12.009.
- Kim, J., Darlington, A. P. S., Bates, D. G., Jimenez, J. I. (2021). The interplay between growth rate and nutrient quality defines gene expression capacity. *bioRxiv*. doi: <https://doi.org/10.1101/2021.04.02.438188>
- Klumpp, Stefan, Zhongge Zhang, and Terence Hwa. (2009). Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* 139 (7): 1366–75. doi:10.1016/j.cell.2009.12.001.
- Kolmsee, T., Delic, D., Agyenim, T., Calles, C., & Wagner, R. (2011). Differential stringent control of *Escherichia coli* rRNA promoters: effects of ppGpp, DksA and the initiating nucleotides. *Microbiology (Reading, England)*, 157(Pt 10), 2871–2879. <https://doi.org/10.1099/mic.0.052357-0>

- LaCroix, R. A., Sandberg, T. E., O'Brien, E. J., Utrilla, J., Ebrahim, A., Guzman, G. I., Feist, A. M. (2014). "Use of Adaptive Laboratory Evolution To Discover Key Mutations Enabling Rapid Growth of Escherichia coli K-12 MG1655 on Glucose Minimal Medium". *Applied and Environmental Microbiology*, 81(1), 17–30. doi:10.1128/aem.02246-14
- Langmead B, Salzberg S. 2012. "Fast gapped-read alignment with Bowtie 2". *Nature Methods* . 9:357-359.
- Lastiri, G., Mercado-Hernández, J.S., Juhyun, K. Jiménez, J., Utrilla, J. (2020). A quantitative method for proteome reallocation using minimal regulatory interventions. *Nature Chemical Biology*. <https://doi.org/10.1038/s41589-020-0593-y>
- Latif, Haythem, Richard Szubin, Justin Tan, Elizabeth Brunk, Anna Lechner, Karsten Zengler, and Bernhard O. Palsson. (2015). "A Streamlined Ribosome Profiling Protocol for the Characterization of Microorganisms." *BioTechniques* 58 (6): 329–32.
- Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *PLoS Computational Biology*, 9. doi: [10.1371/journal.pcbi.1003118](https://doi.org/10.1371/journal.pcbi.1003118), <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15, 550. doi: 10.1186/s13059-014-0550-8.
- Lynch, M., & Marinov, G. K. (2015). The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences of the United States of America*, 112(51), 15690–15695. <https://doi.org/10.1073/pnas.1514974112>
- Maeda, M., Shimada, T., & Ishihama, A. (2015). Strength and Regulation of Seven rRNA Promoters in Escherichia coli. *PloS one*, 10(12), e0144697. <https://doi.org/10.1371/journal.pone.0144697>
- Maloy, S., Nunn, W. (1981). Genetic Regulation of the Glyoxylate Shunt in Escherichia coli K-12. *Journal of Bacteriology* 149(1) p. 173-180.

- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), pp. 10-12. doi:<https://doi.org/10.14806/ej.17.1.200>
- Mauri M, Klumpp S. (2014). A Model for Sigma Factor Competition in Bacterial Cells. *PLoS Comput Biol* 10(10): e1003845. <https://doi.org/10.1371/journal.pcbi.1003845>
- McGlincy, N. J., & Ingolia, N. T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods (San Diego, Calif.)*, 126, 112–129. <https://doi.org/10.1016/j.ymeth.2017.05.028>
- Mordukhova, E. A., & Pan, J. G. (2013). Evolved cobalamin-independent methionine synthase (MetE) improves the acetate and thermal tolerance of *Escherichia coli*. *Applied and environmental microbiology*, 79(24), 7905–7915. <https://doi.org/10.1128/AEM.01952-13>
- Mori, Matteo, Severin Schink, David W. Erickson, Ulrich Gerland, and Terence Hwa. (2017). “Quantifying the Benefit of a Proteome Reserve in Fluctuating Environments.” *Nature*
- Natalello A, Doglia SM, Carey J, Grandori R. (2007). Role of flavin mononucleotide in the thermostability and oligomerization of *Escherichia coli* stress-defense protein WrbA. *Biochemistry*;46(2):543-553. doi:10.1021/bi061769c
- Nyström, T., & Neidhardt, F. C. (1994). Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Molecular microbiology*, 11(3), 537–544. <https://doi.org/10.1111/j.1365-2958.1994.tb00334.x>
- O’Brien, Edward J., Jonathan M. Monk, and Bernhard O. Palsson. (2015). “Using Genome-Scale Models to Predict Biological Capabilities.” *Cell* 161 (5). Cell Press: 971–87. doi:10.1016/j.cell.2015.05.019.
- Pannier L, Merino E, Marchal K, Collado-Vides J. (2017). Effect of genomic distance on coexpression of coregulated genes in *E. coli*. *PLOS ONE* 12(4): e0174887. <https://doi.org/10.1371/journal.pone.0174887>
- Quan, S., Skovgaard, O., McLaughlin, R. E., Buurman, E. T., & Squires, C. L. (2015). Markerless *Escherichia coli* *rrn* Deletion Strains for Genetic

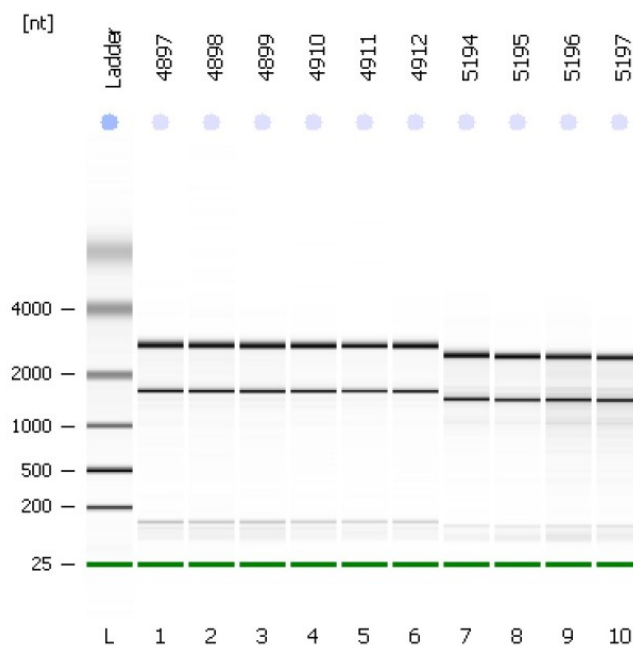
Determination of Ribosomal Binding Sites . G3| Genes|Genomes|Genetics , 5(12), 2555–2557. doi:10.1534/g3.115.022301

- Reid, D. W., Shenolikar, S., & Nicchitta, C. V. (2015). Simple and inexpensive ribosome profiling analysis of mRNA translation. *Methods (San Diego, Calif.)*, 91, 69–74. <https://doi.org/10.1016/j.ymeth.2015.07.003>
- Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology* . Frontiers Research Foundation.<https://doi.org/10.3389/fmicb.2014.00172>
- Ross, W., Sanchez-Vazquez, P., Chen, A. Y., Lee, J. H., Burgos, H. L., & Gourse, R. L. (2016). ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. *Molecular cell*, 62(6), 811–823. <https://doi.org/10.1016/j.molcel.2016.04.029>
- Sastry, A. V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D., Choudhary, K. S., Yang, L., King, Z. A., & Palsson, B. O. (2019). The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nature communications*, 10(1), 5536. <https://doi.org/10.1038/s41467-019-13483-w>
- Schmidt A, Kochanowski K, Vedelaar S, et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol*. 2016;34(1):104-110. doi:10.1038/nbt.3418
- Scott, Matthew, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, y Terence Hwa. (2010). “Interdependence of Cell Growth and Gene Expression: Origins and Consequences.” *Science* 330 (6007): 1099–1102. <https://doi.org/10.1126/science.1192588>.
- Siegele, D. (2005). Universal Stress Proteins in *Escherichia Coli*. *Journal of Bacteriology*, 187(18): 6253–6254. doi: 10.1128/JB.187.18.6253-6254.2005
- Szenk, M., Dill, K. A., & de Graff, A. (2017). Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. *Cell systems*, 5(2), 95–104. <https://doi.org/10.1016/j.cels.2017.06.005>
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

- Tan, J., Sastry, A. V., Fremming, K. S., Bjørn, S. P., Hoffmeyer, A., Seo, S., Voldborg, B. G., Palsson, B. O. (2020). Independent component analysis of E. coli's transcriptome reveals the cellular processes that respond to heterologous gene expression. *Metabolic Engineering*. <https://doi.org/10.1016/j.ymben.2020.07.002>.
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome research*, 21(12), 2213–2223. <https://doi.org/10.1101/gr.124321.111>
- Trapnell, C., Williams, B., Pertea, G. et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511–515 . <https://doi.org/10.1038/nbt.1621>
- Utrilla, J., O'Brien, E. J., Chen, K., McCloskey, D., Cheung, J., Wang, H., Palsson, B. O. (2016). “Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution”. *Cell Systems* , 2(4), 260–271. doi:10.1016/j.cels.2016.04.003
- Valgepea, Kaspar, Kaarel Adamberg, Andrus Seiman, and Raivo Vilu. 2013. “Escherichia Coli Achieves Faster Growth by Increasing Catalytic and Translation Rates of Proteins.” *Molecular bioSystems* 9 (9): 2344–58.
- Wagner R. (2010). ppGpp Signaling. *Bacterial Signaling*. <https://doi.org/10.1002/9783527629237.ch22>
- Winther, K. S., Roghanian, M., & Gerdes, K. (2018). Activation of the Stringent Response by Loading of RelA-tRNA Complexes at the Ribosomal A-Site. *Molecular cell*, 70(1), 95–105.e4. <https://doi.org/10.1016/j.molcel.2018.02.033>
- Yamagishi, M., Matsushima, H., Wada, A., Sakagami, M., Fujita, N., & Ishihama, A. (1993). Regulation of the Escherichia coli rmf gene encoding the ribosome modulation factor: growth phase- and growth rate-dependent control. *The EMBO journal*, 12(2), 625–630.
- Yim, H. Villarejo, M. (1992). osmY, a New Hyperosmotically Inducible Gene, Encodes a Periplasmic Protein in E.coli. *Journal of Bacteriology*, 174(11) p.3637-3644.

- Young, F. S., & Furano, A. V. (1981). Regulation of the synthesis of E. coli elongation factor Tu. *Cell*, 24(3), 695–706. [https://doi.org/10.1016/0092-8674\(81\)90096-9](https://doi.org/10.1016/0092-8674(81)90096-9)
- Zhou, Z. S., Peariso, K., Penner-Hahn, J. E., & Matthews, R. G. (1999). Identification of the zinc ligands in cobalamin-independent methionine synthase (MetE) from *Escherichia coli*. *Biochemistry*, 38(48), 15915–15926. <https://doi.org/10.1021/bi992062b>

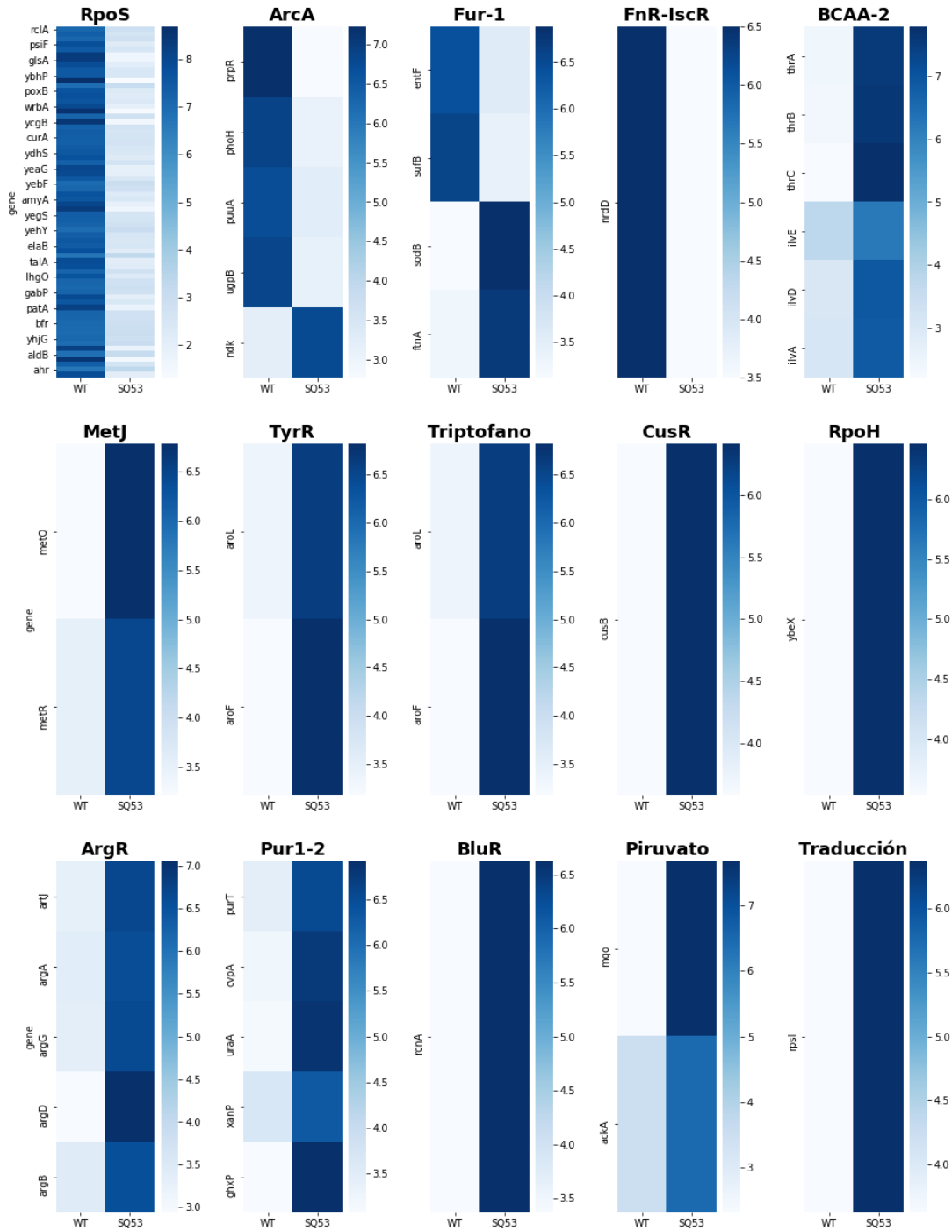
Anexos



Anexo 1. Representación gráfica del análisis por electroforesis para evaluación de calidad de RNA hecho con el software del *Bioanalyzer 2100 Expert*. Este ensayo de análisis de calidad fue corrido en conjunto con otras muestras que fueron trabajadas al mismo tiempo por la unidad de secuenciación masiva del instituto de biotecnología. Las muestras correspondientes a este proyecto son las de los carriles 7 al 10. Los carriles 7 (RIN: 9.3) y 8 (RIN: 9.2) corresponden a las muestras de la cepa WT y los carriles 9 (RIN: 8) y 10 (RIN: 7) corresponden a las muestras de la cepa SQ53.

Original	Abstraccion
Transporte y Metabolismo de Lípidos	Metabolismo
Transporte y Metabolismo de Aminoácidos	Metabolismo
Biosíntesis, Transporte y Metabolismo de M. Secundarios	Metabolismo
Trasnorte y Metabolismo de Carbohidratos	Metabolismo
Transporte y Metabolismo de Coenzimas	Metabolismo
Modificaciones post-traduccionales, chaperonas y reciclaje	Metabolismo
Transporte y Metabolismo de Iones Inorgánicos	Metabolismo
Tráfico Intracelular, Secreción, y Transporte Vesicular	Metabolismo
Transporte y Metabolismo de Nucleótidos	Metabolismo
Producción y Conversión de Energía	Metabolismo
Mecanismos de Defensa	Otros
Control de Ciclo Celular	Otros
Mecanismos de Transducción de Señales	Otros
Sin Anotación	Otros
Biogénesis de Pared Celular/Membrana	Otros
Función Desconocida	Otros
Motilidad Celular	Otros
Traducción, Biogenesis y Ensamblaje de Ribosomas	Traducción
Replicación, Recombinación y Reparación	Transcripción y Replicación
Transcripción	Transcripción y Replicación
Procesamiento y Modificación de RNA	Transcripción y Replicación

Anexo 2. Tabla de equivalencias de Abstracción de COGS. Tabla utilizada para realizar el alineamiento del grupo funcional original (COG) al grupo funcional por abstracción personalizada con el propósito de facilitar la visualización de los sectores del proteoma. Nombre de clasificaciones originales traducidas al español.



Anexo 3. Cambios en iModulones de relevancia para la expresión de proteínas heterólogas, mutante SQ53, versión CuffDiff. Este conjunto de gráficas alternativo para la mutante SQ53 fue hecho con la lista de genes filtrados con un valor de $p < 0.1$ utilizando los valores obtenidos por la prueba de significancia estadística de CuffDiff, la cual no penaliza con el mismo peso la inconsistencia entre réplicas y más bien realiza la prueba entre un promedio calculado de las cuentas normalizadas en FPKMs. Muchos más genes e iModulones aparecen en esta figura,

los cuales tienen un patrón coherente con el fenotipo de crecimiento. Para resolver este conflicto en los resultados de DeSeq2 y CuffDiff debido a la inconsistencia entre réplicas, se sugiere utilizar en el futuro un diseño experimental de RNA-Seq con 3 réplicas por muestra y un método que aproxime mejor la cantidad de proteína traducida como por *Ribosome profiling* para también resolver el conflicto de los cambios en el iModulon de Traducción.

Anexo 4. Avances en estandarización y transferencia de protocolo para Ribosome profiling

Protocolo

Cultivo y Cosecha

0.- Inocular un matraz de 125 ml con 25 mL de medio de cultivo. Para las condiciones experimentales (M9 4 g/L), estos cultivos *overnight* estarán saturados y listos para ser utilizados como inóculos para el cultivo principal

1.- En 1 matraz de 500 mL añadir 420 mL de medio de cultivo e inocular a una densidad inicial (OD₆₀₀) de 0.05, dividir en partes iguales, utilizando una probeta de 100 mL, el volumen en 6 matraces de 250 mL. Agitar el matraz original antes de cada división para asegurar una repartición homogénea. Comprobar OD inicial de los matraces individuales e incubar a la temperatura y agitación correspondientes a las condiciones experimentales.

2.- Monitorear el cultivo hasta que alcance la OD objetivo de cosecha (0.35). Vaciar los ~50 mL de cultivo de cada matraz lo más rápido posible a los tubos falcon con arena (6). Centrifugar a 20,000 rfc por 2 minutos.

3.-Extraer la mayoría del medio de los tubos falcon de 50 mL por decantación, teniendo cuidado de no perder arena/biomasa en la decantación. La mayoría del volumen puede decantarse rápido pero los últimos mililitros deben ser absorbidos con las puntas de 250 µL.

4.-Sumergir los pellets tal cual como quedaron después de extraer el medio por un intervalo breve de tiempo al Dewar (3 segundos aprox.) Colocar los tubos en la hielera con hielo en lo que se termina el procedimiento para todos los tubos.

5.-Una vez teniendo los pellets, estos se pueden almacenar a -80°C o puede procederse a la lisis y extracción.

Lisis y Extracción

1.- Con el tallo de una espátula metálica golpear lateralmente los pellets adheridos al fondo de los tubos falcon, sosteniendo estos boca abajo sobre el mortero pre-enfriado hasta desprender el pellet. El mortero debe pre-enfriarse por dentro y por fuera. Tener cuidado de no romper el tubo, es posible desprender el pellet sin usar fuerza excesiva, pero esto puede suceder, por lo que es recomendable utilizar protección para los ojos y bata por si ocurre una fragmentación accidental.

2.- Con la bala pre-enfriada, después de sumergirla en el nitrógeno líquido y asegurándose que el material también acaba de ser enfriado con algo de nitrógeno, colocar la bala dentro del mortero. Teniendo el mortero recargado en el trape cubierto de toallas de papel, pulverizar el material con fuerza golpeando el martillo con la bala, ser vigoroso. Repetir por lo menos 2 veces.

3.- Manteniendo la misma nota que antes, mantener fría ahora la broca y con el taladro pulverizar el material hasta que éste tenga una consistencia ultra fina y se vea por lo menos de color gris. Podría llegar hasta verse de color negro, pero el cambio en la efectividad de la lisis no es notable.

4.- Transferir utilizando una cuchara de acero pre-enfriada el material pulverizado al tubo falcon etiquetado como "Recolectado". Añadir 250 μL de buffer de lisis de manera inmediata y dejar que el material se disuelva en el buffer hasta que alcance la temperatura ambiente. No dejar pasar mucho tiempo, poner atención al estado del tubo. Centrifugar a 20,000 rfc por 2 min.

5.- Pasar el sobrenadante clarificado al microtubo marcado como "Primer Centrifugado" y añadir Desoxicolato de sodio al 10% para una concentración final en el tubo del 0.3%. Mezclar por inversión y dejar incubar en frío por 3-5 minutos, debe verse lechoso. 6.-Centrifugar a 16,800 rfc por 10 min. Sí se sigue viendo lechoso, dejar incubar en hielo por otros 5 minutos y volver a centrifugar por 10 min.

7.- Transferir este material clarificado al tubo marcado como "Clarificado dd/mm/aa". Cuantificar utilizando 10 ul del material utilizando el kit de cuantificación RNA HS assay del Qubit 4 Fluorometer. Este protocolo está estandarizado para obtener alrededor de 5 ug totales disueltos en el volumen obtenido, lo óptimo es tener una concentración de ~10 ng/ul disueltos en 500 ul.

8.- Para verificar la integridad del material puede procederse a separar una alícuota de volumen equivalente a 1500 ng de material y almacenar a -80°C ambos tubos o proceder al siguiente paso, guardando sólo el de "Clarificado".

Concentración con Trizol para revisión de integridad

1.-Agregar 500 ul de Trizol en el tubo conteniendo la muestra a concentrar. Dejar reposar 10 minutos a temperatura ambiente.

2.-Agregar 100 ul de cloroformo y agitar con vórtex durante 15 s. Incubar por 3 minutos a Temp. Ambiente. Centrifugar a 12,000 rfc por 15 minutos a 4°C.

3.-Colectar el sobrenadante cuidadosamente sin tocar la fase orgánica. Aproximadamente se obtendrán 500 microlitros.

4.-De manera opcional y para asistir en la extracción del sobrenadante sin perturbar el pellet de RNA, utilizar el reactivo Glycoblue coprecipitant. Añadir 0.8 microlitros (50 µg/ml) al sobrenadante colectado en el paso anterior y después 250 ul de isopropanol. Agitar por inversión y dejar incubar durante 10 min a Temp. Ambiente.

5.-Centrifugar a 12,000 rfc por 10 minutos a 4°C. Si se utilizó el Glycoblue, se observará la formación del pellet de color azul, si no, se verá de color blanco.

6.-Retirar el sobrenadante con un movimiento rápido y agregar 1 ml de etanol al 75% frío. Agitar por vórtex hasta que el pellet se desprenda. Centrifugar a 16,800 rfc durante 5 minutos a 4°C. Repetir.

7.Descartar el etanol. Con una centrifugación corta y terminando de retirar el etanol con una micropipeta, de 5 en 5 ul, retirar la mayor parte de etanol restante. Absorbiendolo por las paredes teniendo cuidado de no absorber el pellet. Dejar secar durante 10 minutos.

10.-Resuspender en agua DEPC/Libre de RNAsas a 56°C (15-30 ul).

11.-Analizar la muestra concentrada en un Nanodrop para confirmar pureza y una concentración esperada de alrededor de 100 ng/μL.

12.-Correr un gel de Agarosa al 1% a 90 voltios durante 30-35 minutos. Siempre con buffer nuevo.

Digestión con nucleasa

1.- La cantidad ingresada para el tratamiento debe ser equivalente a mínimo 1 μg de material. Sin embargo, tomando en cuenta posibles pérdidas en el proceso y considerando que del paso anterior tendremos disponibles ~5 μg de material, utilizaremos 4 μg en total.

2.-Preparar la siguiente mezcla de reacción:

MNase Reaction Mix (200 μ L)

Component	Volume
Lisado clarificado (~4 μ g totales)	X μ L
CaCl ₂ , 500 mM	2 μ L
Suprase-In, 20 U/ μ L	3.6 μ L
MNase S7 375 U/ μ L	5 μ L
MNase Buffer (Receta abajo)	q.s. to 200 μ L
Total Volume	200 μL

- X = 1800 ng / (concentración obtenida en ng/ μ L) [Para este caso, el valor podría oscilar entre 100 (18 ng/ μ L) - 130 μ L (14 ng/ μ L), aproximadamente]
- Para resuspender la MNasa S7, agregar 40 μ L de Tris pH 8 10 mM a un vial de roche MNase (15000 U). Esto nos daría una concentración final de 375 U/ μ L.

MNase Buffer (escalar a cuánto se necesite)

Components	Volume
Tris pH 8.0, 1 M	25 μ L
NH ₄ Cl, 1 M	25 μ L
MgOAc, 1 M	10 μ L
Chloramphenicol, 50 mg/mL	10 μ L
5'-guanylyl imidodiphosphate (GMPPNP), 10 mg/mL	1 μ L
RNase-free Water	929 μ L
Total Volume	1 mL

3.-Incubar la mezcla por 1 hora a 25°C

4.-Detener la reacción agregando 2.5 μ L EGTA (500 mM stock). Reservar en hielo.

Recuperación de monosomas, aislamiento de RNA y remoción de tRNAs

1.-Preparar buffer de polisomas:

Buffer de polisomas (10 mL)

Component	Volume
Nonidet P40, 100%	100 μ L
MgCl ₂ , 1 M	500 μ L
EGTA, 0.5 M	500 μ L
Tris pH 8.0, 1 M	500 μ L
NaCl, 5 M	500 μ L
Agua libre de RNAsas	7.9 mL
Total Volume	10 mL

- La concentración final del buffer es de 1 % Nonidet P40, 50 mM MgCl₂, 25 mM EGTA, 50 mM Tris-HCl pH 8.0, 250 mM NaCl

2.-Preparar 2 columnas de Sephacryl S400 MicroSpin para cada muestra. Agitar la columna para resuspender la resina, golpear ligeramente para remover las burbujas que se hayan formado

3.-Aflojar la tapa un cuarto de vuelta y romper el sello plástico del fondo. Poner la columna en un tubo de colección y centrifugar por 1 min a 500 rcf para remover el buffer de almacenamiento en exceso.

4.-Equilibrar la resina pasando 1.5 mL del buffer de polisomas por cada columna en 3 alícuotas de 500 μ L. Vaciar el tubo de colección después de cada lavado. Remover el último lavado por centrifugación por 4 minutos a 500 rcf en una microcentrifuga a 4°C.

5.-Descartar el eluido y colocar un nuevo tubo de colección para cada columna. (Microtubo de 1.5 ml, estéril, libre de RNAsas)

6.-Separar en 2 alícuotas de 100 μ L, la muestra tratada por MNasa en la subsección anterior. Aplicar alícuotas a las columnas y centrifugar por 2 minutos a 4°C a 600 rcf.

7.-Agregar 700 μL de Qiazol al eluido combinado de las 2 columnas, mezclar con vortex inmediatamente y luego seguir el protocolo de aislamiento de RNAs pequeños del miRNeasy Mini Kit de Qiagen. Incubar por 5 min a 25 °C sin agitación. El tiempo de incubación promueve la disociación de complejos nucleoproteicos. Cargar la columna tantas veces como sea necesario hasta pasar todo el volumen de muestra.

9.-Agregar 140 μL de cloroformo a la muestra mezclada con Qiazol del paso 7. Agitar vigorosamente por 15 segundos. No agitar con vórtex, solo agitación vigorosa. Incubar 2 minutos y medio a 22°C.

10.-Centrifugar por 15 min a 12000 rcf a 4°C. Después de la centrifugación configurar la centrífuga a temperatura ambiente (15-25°C)

11.-Después de la centrifugación del paso 10, la muestra se separará en 3 fases, una superior, acuosa y sin color conteniendo el RNA. Una fase blanca y una fase inferior de color rojo, orgánica. El volumen de la fase acuosa debería ser de alrededor de 350 μL .

12.-Transferir la fase acuosa superior a un nuevo microtubo de reacción. Agregar 1 volumen de etanol al 70% (alrededor de 350 μL) y agitar con vortex. No centrifugar, proceder inmediatamente al paso siguiente.

13.-Tomar la muestra (aproximadamente 700 μL), incluyendo cualquier precipitado que se haya formado para introducirlo en una columna RNeasy Mini spin postrada en un tubo de colección. Cerrar la tapa con cuidado (gentilmente) y centrifugar a 8000 rfc por 15s a temperatura ambiente. Tomar el eluido en un tubo de reacción de 2 ml. Descartar la columna, conservar el eluido.

14.-Agregar 450 μL de Etanol al 100% (0.65 volúmenes) al eluido y agitar con vortex. No centrifugar. Proceder inmediatamente al siguiente paso.

15.-Pasar 700 μL de la muestra en una columna RNeasy MinElute spin en un tubo de colección. Cerrar la tapa gentilmente y centrifugar por 15s a 8000 rcf a temperatura ambiente (15-25°C). Descartar el flow-through. Repetir este paso hasta

que toda la muestra haya sido pasada por la spin-column, descartar el FT en cada ocasión.

16.-Pasar 500 μL de Buffer RPE en la columna RNeasy MinElute. Cerrar la tapa gentilmente y centrifugar por 15 segundos a 8000 rcf. Descartar el eluido. Después de la centrifugación, remover la columna RNeasy MinElute spin del tubo de colección con cuidado de modo que se prevenga que algo del etanol en la columna contamine el eluido.

17.-Colocar la columna RNeasy MinElute spin en un tubo de colección de 1.5 mL y añadir 12 μL de agua libre de RNAsas en la membrana de la columna. Cerrar la tapa con cuidado y centrifugar por 1 minuto a 8000 rcf para eluir la fracción con RNAs pequeños. Pasar el eluido una segunda vez por la columna mejora la recuperación. La columna tiene aproximadamente 2 μL de volumen muerto por lo que solo se recuperarán 10 μL .

18.-Medir la concentración de RNA utilizando el Qubit RNA HS Assay Kit. Utilizar 1 μL para medirla. Almacenar a -80°C previo a continuar con el resto del protocolo. La masa total disuelta debe ser de mínimo 1 μg .

10.-La remoción de tRNAs se logra a través de la hibridación de oligos de DNA que son específicos para las mitades 5' y 3' de los tRNAs de E.coli y que han sido modificados con nucleótidos desfosforilados en los extremos para no ser extensibles por PCR. Luego de la hibridación, una RNAsa H termoestable se vuelve capaz de reconocer los tRNAs y de digerirlos después de un periodo de incubación. Preparar la receta para esta reacción como dice a continuación:

Componente	Volumen
Adapter-ligated small RNA	10 μL
Anti-tRNA oligo mix (non-PCR)	10 μL
Superase-In, 20 U/uL	1 μL
10x Hybridase Buffer	3 μL
Thermoestable Hybridase RNase H	1 μL
Agua libre de RNAsas	5 μL
Volumen total	30 μL

11.-En una termociclador programable calentar la muestra a 90°C y mantener esa temperatura por 10 segundos. Enfriar la muestra a 65°C por 1-2 minutos y mantener la temperatura por 5 minutos. Ciclar la reacción 6 veces. (~30-45 minutos totales de reacción)

12.-Concentrar la muestra repitiendo el protocolo de concentración con TRIzol utilizado para la primera verificación de integridad. Resuspender la pastilla obtenida al final en un volumen de 7 µL.

Fosforilación y aislamiento de Ribosome footprints

1.-A los 7 µL del volumen anterior, añadir 1 µL de 10X T4 PNK Buffer, 1 µL de Superase-In 20 U/µL, 1 µL de T4 PNK 10 u/µL y 1 µL de ATP 10 mM. Incubar a 37°C por 30 minutos.

2.-Agregar a la muestra 10 µL de tinte de cargado para RNA y preparar el siguiente gel: 15% acrilamida (29:1 acrilamida:bis), 1x TBE, 8M urea. Calentar y agitar para solubilizar la urea.

3.-En 5 ml del gel en solución, agregar 20 µL de persulfato de amonio al 10% y 4 µL de TEMED. Cargar la muestra en un mini-gel de 0.75 mM sin gel concentrador. Utilizar un peine de 10 pozos y 1x TBE como buffer de corrida.

4.-Lavar urea de los pozos utilizando el mismo buffer y una jeringa, luego inmediatamente cargar la muestra mezclada. Correr junto a un marcador de peso molecular de alta resolución.

5.-Correr el gel hasta que el tinte haya salido del mismo (~1h a 400 voltios)

6.-Colocar el gel en un recipiente de 25 ml con buffer TBE 1x. Agregar tinte SybrGreen 1:10000 (2.5 µL) e incubar por 5 min con agitación.

7.-Bajo luz UV, cortar las bandas entre ~28 y ~40 nt con una navaja estéril y colocar en un microtubo de 1.5 mL. Al tubo agregar 400 µL de NaOAc 400 mM pH 5.2. Congelar la muestra por 10 min a -80°C

8.-Descongelar la muestra a 95°C por 5 min, triturar exhaustivamente con una punta de pipeta. Repetir 3 veces y agitar con vortex.

9.-Centrifugar a máxima velocidad por 10 minutos y coleccionar el sobrenadante. Repetir para remover por completo la acrilamida.

10.-Agregar 1 mL de etanol al sobrenadante, 1 μ L de tinte Glycoblu e incubar por 1 hora a -20°C

11.-Precipitar el RNA por centrifugación a 15 min a máxima velocidad a 4°C. Decantar el sobrenadante y lavar la pastilla precipitada dos veces con etanol frío al 80%, centrifugar por 5 minutos después de cada lavado. Dejar la pastilla secarse por 10 minutos. Resuspender en 20 μ L de agua libre de RNAsas y cuantificar con Qubit la concentración del material.

12.-Utilizar el material como material de entrada para un protocolo de construcción de bibliotecas para small RNA y secuenciar por lecturas pareadas a una profundidad de 5 millones de lecturas.