



**UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS**

**INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS CENTRO
DE INVESTIGACIÓN EN DINÁMICA CELULAR**

**“Análisis comparativo del transcriptoma de *Escherichia coli*
y *Schizosaccharomyces pombe* en dos escenarios metabólicos”**

TESIS PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA

L.C. Joivier Vichi Lozada

DIRECTOR DE TESIS

Dr. Armando Hernández Mendoza

CUERNAVACA, MORELOS

Septiembre, 2019



**UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS**

**INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS CENTRO
DE INVESTIGACIÓN EN DINÁMICA CELULAR**

**“Análisis comparativo del transcriptoma de *Escherichia coli*
y *Schizosaccharomyces pombe* en dos escenarios metabólicos”**

TESIS PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA

L.C. Joivier Vichi Lozada

DIRECTOR DE TESIS

Dr. Armando Hernández Mendoza

CUERNAVACA, MORELOS

Septiembre, 2019

Cómite Revisor

**Dr. José Luis Puente
(IBT- UANM)**

**Dr. Ramón Alberto Batista García
(CIDC-UAEM)**

**Dra. RosaM. Gutierrez Ríos
(IBT-UNAM)**

**Dr. Ricardo Grande Cano
(IBT-UNAM)**

**Dra. Sonia Dávila Ramos
(CIDC-UAEM)**

**Dr. Juan Enrique Morette Sánchez
(IBT-UNAM)**

**Dr. Armando Hermandéz Mendoza
(CIDC-UAEM)**

Publicaciones relacionadas con la Tesis

- Analog circuits of transcriptional regulation of *Escherichia coli* and *Schizosaccharomyces pombe* in respiratory and fermentative growth inferred by comparative analysis of global transcriptional profiles. Journal of Molecular Microbiology and Biotechnology (Research Article in Review : 2019 -11-1).

Authors : Joivier Vichi (Co-author), Emmanuel Salazar (Co-author), Veronica Jimenez (Co-author), Leticia Olvera (Co-author), Ricardo Alfredo Cano (Co-author), Enrique Enrique Morett (Co-author), Edgar Dantán González (Co-author), Armando Hernandez-Mendoza (Corresponding Author).

- **Proposed Title** : Exploration of the non-coding RNAome of *Schizosaccharomyces pombe* from mRNA and small RNA expression profiling analysis. (Artículo en redacción)

Authors : Joivier Vichi (Co-author), Emmanuel Salazar (Co-author), Veronica Jimenez (Co-author), Leticia Olvera (Co-author), Sonia Davila Ramos (Co-author), Ricardo Alfredo Cano (Co-author), Enrique Enrique Morett (Co-author), Edgar Dantán González (Co-author), Armando Hernandez-Mendoza (Corresponding Author).

Resumen

Las técnicas de secuenciación de última generación (*Next Generation sequencing*) han determinado una nueva era en las ciencias biológicas, especialmente en las investigaciones genómicas. Actualmente, la secuenciación masiva de RNA (RNA-seq), ha incrementado el conocimiento acerca del “universo del RNA” a partir de la caracterización global de transcriptomas, estudios de expresión diferencial de genes y el continuo descubrimiento de transcritos reguladores no codificantes como los sRNAs (pequeños RNAs del inglés small RNAs). La regulación post-transcripcional de genes es un mecanismo de control de la expresión génica que está presente tanto en organismos procariontes como eucariontes y regula múltiples procesos biológicos. Uno de esos mecanismos es la regulación por RNAs no codificantes (ncRNAs). Ejemplos de ncRNA que regulan los niveles de transcritos de genes relacionados al metabolismo central del carbono se han descrito en bacterias y en eucariontes superiores. Con el propósito de describir si existen analogías en circuitos regulatorios mediados por ncRNAs en dos patrones celulares, se realizó un análisis comparativo del transcriptoma de *Escherichia coli* y *Schizosaccharomyces pombe* en condiciones similares de crecimiento. Se realizaron 12 librerías de RNA-seq. Los datos de RNA-seq contenían una cantidad de lecturas del orden de 10^6 y las lecturas mapeadas sin ambigüedad sobre el genoma de cada organismo osciló entre 1-1.3 millones. Un total de 739 y 743 genes se detectaron como expresados diferencialmente para *E. coli* y *S. pombe* respectivamente. Entre los genes expresados diferencialmente en *E. coli* se destacan factores de transcripción generales como *fliA*, factores de regulación específicos e incluso 11 sRNAs validados en la literatura y otros como *cyaR* que está sujeto a represión catabólica. En *S. pombe* se destacan como enriquecidos, genes asociados a la síntesis de proteínas, la síntesis de nucleótidos así como transportadores de hexosas y ncRNAs relacionados con estrés celular.

Abstract

Next Generation sequencing techniques have ushered in a new era in life sciences, especially genomic research. Currently, massive RNA sequencing (RNA-seq) has increased knowledge about the "RNA universe" from the global characterization of transcriptomes, differential gene expression studies and the continued discovery of non-coding regulatory transcripts such as sRNAs (small RNAs). Post-transcriptional gene regulation is a mechanism for controlling gene expression that is present in both prokaryote and eukaryote organisms and regulates multiple biological processes. One such mechanism is regulation by non-coding RNAs (ncRNAs). Examples of ncRNAs that regulate the levels of transcripts of genes related to central carbon metabolism have been described in bacteria and higher eukaryotes. In order to describe whether analogies exist in regulatory circuits mediated by ncRNAs in two cell patterns, a comparative analysis of the transcriptome of *Escherichia coli* and *Schizosaccharomyces pombe* in similar growth conditions was performed. Twelve RNA-seq libraries were performed. The RNA-seq data contained a number of readings of the order of 106 and the unambiguously mapped readings on the genome of each organism ranged from 1-1.3 million. A total of 739 and 743 genes were detected as differentially expressed for *E. coli* and *S. pombe* respectively. Among the genes differentially expressed in *E. coli*, general transcription factors such as *fliA*, specific regulatory factors and even 11 sRNAs validated in the literature and others such as *cyaR* which is subject to catabolic repression stand out. In *S. pombe*, genes associated with protein synthesis, nucleotide synthesis as well as hexose transporters and ncRNAs related to cellular stress are highlighted as enriched.

Índice General

CAPÍTULO I : Introducción.....	1
1.2 El mundo infinito del RNA.....	3
1.3 Biogénesis de ncRNA antisentidos.....	5
1.4 ncRNAs codificados en trans.....	7
1.5 Detección y validación de ncRNAs.....	9
1.6 Redes de Interacción biológicas.....	10
1.7 Problema.....	14
CAPÍTULO II : Hipótesis y Objetivos.....	15
2.1 Hipótesis.....	15
2.2 Objetivos.....	15
Objetivo General.....	15
Objetivos específicos.....	15
CAPÍTULO III : Materiales y Métodos.....	16
3.1 Cepas de microorganismos.....	16
3.2 Medios y condiciones de cultivo.....	16
3.3 Extracción y preparación del RNA total.....	17
3.4 Construcción de bibliotecas para la secuenciación masiva de RNA.....	18
3.5 Análisis Bioinformático.....	19
3.5.1 Procesamiento de archivos crudos en formato fastq.....	19
3.5.2 Alineamiento y filtrado de lecturas.....	19
3.5.3 Abundancia de transcritos y Expresión Diferencial de Genes.....	21
3.5.4 Análisis de enriquecimiento de vías metabólicas y procesos biológicos.....	22
CAPÍTULO IV : Resultados.....	23
4.1 Determinación de la expresión diferencial de genes en <i>E. coli</i> K-12 y <i>S. pombe</i> en dos escenarios metabólicos diferentes.....	23
.....	23
4.1.1 Optimización del crecimiento y establecimiento de la fase exponencial.....	23
4.1.2 Extracción de RNA total.....	25
4.1.3 Estadísticas generales de los datos de RNA-seq.....	27
4.1.4 Similitud y Variación entre las réplicas biológicas.....	29
4.1.5 Coeficiente de Variación Biológica.....	33
4.1.6 Análisis de expresión diferencial de genes al utilizar glucosa o glicerol/acetato como fuente de carbono y energía.....	34
4.1.7 Enriquecimiento funcional de procesos celulares y de vías metabólicas.....	38
4.1.8 Análisis de los proceso celulares enriquecidos con la herramienta BINGO.....	40
4.2 Diferenciación celular inducida por la utilización de fuentes de carbono gluconeogénicas.....	46
4.3 Descripción de la abundancia de factores de transcripción y ncRNAs	49
CAPÍTULO V : Discusión General.....	54

5.1 Análisis Global del Transcriptoma de <i>E. coli</i> y <i>S. pombe</i> en dos escenarios metabólicos.....	54
5.2 Abundancia de ncRNAs en fuentes de carbono no glicolíticas.....	59
CAPÍTULO VI : Conclusiones Generales y Perspectiva.....	61
6.1 Conclusiones generales.....	61
6.2 Perspectivas.....	62
Referencias Bibliográficas.....	63
Anexos.....	70
Anexo 1 : Artículo en proceso de Revisión.....	70
Anexo 2 : Artículo en etapa de Redacción.....	71

Índice de Figuras

- Fig. 1: Crecimiento de *E. coli* y *S. pombe* al utilizar glucosa 2 % w/v (curva de color rojo) o glicerol 2% / acetato 0,2% (curva de color azul) como fuente de carbono. Cada punto representa la media de DO a 600 nm de tres cultivos independientes \pm 1 SD.....24
- Fig. 2: Electroforesis de RNA total de *E. coli* y *S. pombe* por diferentes métodos. 1. RNA total de *E. coli*, en gel de agarosa al 1% extraído por columnas de Quiagen. 2 Electroforesis en gel de Agarosa RNA total de *E. coli* extraído con el Kit Mirvana (Ambion). 3 Electroforesis en gel de PAGE al 10 % y Urea de RNA total de *E. coli* extraído con el Kit Mirvana (Ambion). 4. Electroforesis de RNA total de *S. pombe* en gel de agarosa al 1%.....26
- Fig. 3. Cuantificación e integridad de RNA total de *E. coli* (superior) y *S. pombe* (inferior). Cada muestra representa el análisis en Bioanalyzer 2100, Chip RNAnano (Agilent) de 1 μ l de RNA total....26
- Fig. 4: Generación de bibliotecas y análisis global de archivos de secuenciación. Las bibliotecas se realizaron de acuerdo al protocolo de Illumina y se secuenciaron por multiplexing. Los cuadros muestran el total de las lecturas originales por réplica en fastq (tabla) y las lecturas filtradas y alineadas en el genoma de referencia , archivos bam (cuadro azul).....27
- Fig. 5: Correlación de Spearman de los datos crudos de conteo cadena específicos entre réplicas biológicas para datos de RNA-seq de *E. coli* y *S. pombe*.....29
- Fig. 6: Gráfico de escala multidimensional (MDS) de los archivos de RNA-seq de *E. coli* (sección superior) y *S. pombe* (sección inferior). La muestras se agruparon en la dimensión 1 por tratamiento biológico (Glucosa vs. Glicerol/acteto) mientras que la dimensión 2 por la variación entre réplicas biológicas. Cada réplica biológica se obtuvo de cultivos independientes.....31
- Fig. 7: Coeficiente de Variación Biológica para el conteo de lecturas en *E. coli*. Cada punto representa el coeficiente de dispersión específico de cada gen entre las réplicas biológicas.....32
- Fig. 8: Coeficiente de Variación Biológica para el conteo de lecturas en *S. pombe*. Cada punto representa el coeficiente de dispersión específico de cada gen entre las réplicas biológicas.....33
- Fig. 9: Expresión diferencial de genes en *E. coli*. Para la realización del análisis se seleccionaron genes con una sumatoria de $\log_{2} \text{CPM} > 3$ en *E. coli* (4538). Los genes con un $\log_{2} \text{FC} > 1.5$ y un $\text{FDR} <$

0.05 se consideran como diferencialmente expresados.....	34
Fig. 10: Expresión diferencial de genes en <i>S.pombe</i> . Para la realización del análisis se seleccionaron genes con una sumatoria de $\log\text{CPM} > 1$ en <i>E. coli</i> (4538). Los genes con un $\text{Log}_2 \text{FC} > 1.5$ y un $\text{FDR} < 0.01$ se consideran como diferencialmente expresados.....	35
Fig. 11: Distribución de genes expresados diferencialmente en <i>E. coli</i> (superior) y <i>S pombe</i> (Inferior). TF (genes que codifican a factores de transcripción) son representados con un área grisoscuro, ncRNA son representados con un área negra mientras que genes que codifican a proteínas que no codifican a factores de transcripción están representados con un área gris claro.....	36
Fig. 12: Comparación de GOs enriquecido funcionalmente en el crecimiento en glucosa y glicerol/actetato en <i>E. coli</i>	38
Fig. 13: Comparación de GOs enriquecido funcionalmente en el crecimiento en Glucosa y Glicerol/Actetato en <i>S.pombe</i>	38
Fig. 14: Procesos biológicos enriquecidos en <i>E. coli</i> al crecer en glucosa. El diagrama se realizó empleando la herramienta BINGO de Cytoescape utilizando las a redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.	41
Fig. 15: Procesos biológicos enriquecidos en <i>E. coli</i> al crecer en glicerol/acetato. El diagrama se realizó empleando la herramienta BINGO de Cytoescape utilizando las a redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.	42
Fig. 16: Procesos biológicos enriquecidos en <i>S. pombe</i> al crecer en glucosa. El diagrama se realizó empleando la herramienta BINGO de Cytoescape utilizando las a redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.	43
Fig. 17: Procesos biológicos enriquecidos en <i>S. pombe</i> al crecer en glicerol/acetato. El diagrama se	

realizó empleando la herramienta BINGO de Cytoescape utilizando las a redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.

.....44

Fig. 18: Empleo de la herramienta KEEG mapper&color para indicar los genes sobreexpresados que estan asociados al ensamblaje flagelar. Genes sobre expresados se ddestacan en color rosa.....45

Fig. 19: Tinción de células de E. coli mediante el Método de Leifson. Las fechas indican la presencia del Flagelo de forma polar. A Bacterias obtenidas en glucosa. B y C bacterias crecidas en Glicerol-acetato. La barra indica la extensión de 5 micrometros.....46

Fig. 20: Diferencias morfológicas en S. pombe al crecer en glucosa (superior) y en glicerol/ acetato (inferior).....47

Fig. 21: Indice de log FC de los factores sigma de Escherichia coli para las condiciones estudiadas.. 49

Fig. 22: Agrupamiento mediante el uso de la distancia euclidiana de los ncRNAs de.....49

Fig. 23: Correspondencia entre genes análogos relacionados con el procesamiento de la glucosa a piruvato.....53

Índice de tablas

Tabla 1: Análisis de región promotora de ncRNA en S. pombe.....50

CAPÍTULO I : Introducción

En la actualidad el número de publicaciones acerca de elementos regulatorios del tipo de ncRNA se eleva cada año de forma geométrica. Razones que explican este hecho es su potencial uso en la terapia génica para el tratamiento de distintas patologías en plantas y animales, o simplemente en la descripción de mecanismos de regulación de genes que en décadas pasadas no se sospechaba de su existencia.

Particularmente los microRNAs y siRNAs y sus análogos bacterianos los sRNAs se consideran que son elementos ubicuos y tienen potencialmente varios genes blancos bajo su control. Dichos elementos al igual que los factores transcripción, constituyen importantes componentes de las redes de regulación de genes y ejercen su efecto sobre distintos procesos celulares ¹(Costa, Aprile et al. 2013).

Por otra parte, las redes celulares no son sistemas biológicos aislados y han estado sujetos a una presión selectiva durante la evolución. Por ejemplo, las levaduras, como *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe* son capaces por un lado de fermentar glucosa a alcohol cuando hay exceso de ésta, aún en presencia de oxígeno, fenómeno conocido como efecto Crabtree. Asimismo llevan a cabo un metabolismo oxidativo cuando la concentración de glucosa es baja o cuando la fuente de carbono no es fermentable, como es el caso del succinato, etanol, glicerol y otros compuestos ²(Fendt, Oliveira et al. 2010). De forma similar *Escherichia coli*, cuando crece en medios con una concentración de glucosa alta, produce

acumulación de acetato a partir de acetyl Co-A, fenómeno conocido como “efecto Crabtree bacteriano”, y reprime la expresión de genes que intervienen en la asimilación de fuentes alternas de carbono en un efecto conocido como represión catabólica. Al igual que en las levaduras, cuando la glucosa es escasa, *E. coli* puede utilizar otras fuentes de carbono como piruvato, glicerol, succinato entre otros ³(Paliy and Gunasekera 2007).

Analizar y comparar el genoma de dos organismos (como por ejemplo el de una bacteria y una levadura) permite determinar características comunes en términos de genes ortólogos de secuencia y/ o función, e incluso inferir si existen puntos comunes en redes de regulación transcripcional. Sin embargo, el transcriptoma de un organismo presenta mayor complejidad y dinamismo que el genoma, provee información acerca de la expresión de genes codificantes de proteínas y de regiones no codificantes del DNA como distintos elementos de ncRNAs. La afirmación anterior conlleva a que analizar el transcriptoma de dos o más organismos sometidos a iguales condiciones de estudio es una alternativa para comparar la expresión genes ortólogos y discernir cómo han evolucionado las redes de interacción biológicas que controlan un mismo proceso celular para distintos organismos modelos.

Datos recientes de secuenciación masiva de RNA (RNA-seq) indican que las porciones transcritas de los genomas son más extensas y complejas que las que han sido previamente estudiadas por técnicas de hibridación. El desarrollo de las tecnologías de secuenciadores de última generación (NGS) permite secuenciar el RNA total de un organismo en un momento y bajo una condición determinada, lo que posibilita determinar la expresión diferencial de genes conocidos, de secuencias no anotadas, así como las marcas de asociación de los ácidos nucleicos con distintas proteínas. De igual manera, es posible con ayuda de

herramientas computacionales, caracterizar estructuras secundarias de transcritos y realizar enfoques sobre las redes regulatorias, sus componentes, propiedades topológicas y su evolución partir de datos de RNA-seq (Hoeijmakers, Bartfai et al. 2013).

Consecuentemente la presente investigación tiene como propósito caracterizar la población de mRNAs y de sRNAs a partir de datos de RNA-seq de una bacteria y una levadura, como modelo de célula procariota y eucariota respectivamente, en condiciones similares de crecimiento. Los resultados del análisis de datos de RNA-seq permitirán inferir si existen analogías en circuitos de regulación post-transcripcional que se hayan conservado en la evolución o que sean puntos de evolución convergente en dos dominios de la vida.

1.2 El mundo infinito del RNA

Durante los últimos años la visión del RNA como biomolécula ha cambiado drásticamente. En el presente se conoce que el RNA funciona como regulador, catalizador y controlador de varios procesos celulares que antes se atribuían exclusivamente a las proteínas. El universo de todas las especies de RNA transcritos de un genoma se denomina "transcriptoma" y el análisis del mismo a través de diferentes herramientas de bioinformática ha revelado que componentes de los mecanismos de regulación de la expresión genes es específica de cada condición de estudio ⁴(Hawkins and Morris 2008).

Tanto en eucariotas como en procariotas la regulación de la expresión génica ocurre a múltiples niveles. La combinación y acción concertada entre los factores de transcripción y los sRNAs aseguran una respuesta precisa ante determinado estímulo. De forma canónica, a

nivel transcripcional los factores de transcripción se unen a promotores y potencian la activación o represión de un gen. A nivel post-transcripcional de forma general los sRNAs reprimen la expresión de un gen por degradación o inhibición de su traducción ⁵(Beisel and Storz 2010).

La nomenclatura utilizada para los RNAs regulatorios desafortunadamente no ha sido lo suficientemente uniforme. Generalmente el término sRNAs es utilizado para los RNAs pequeños bacterianos mientras que ncRNA o RNA no codificante es ampliamente utilizado para sRNAs eucarióticos. Los sRNAs poseen una extensión variable y por lo general son menores a 200 nt en bacterias mientras que en eucariontes generalmente son menores que 30 nt y las tres clases más conocidas son los siRNAs (del inglés small interfering RNAs), los miRNAs (microRNAs) y los piRNAs RNAs (del inglés Piwi dependent small RNA) ^{6,7}(Mattick and Makunin 2006; Watanabe and Kanai 2011).

En el 1995, Delihias realiza una comparación entre los RNAs regulatorios micF, dicF descritos en *E. coli* desde 1984 y el recientemente encontrado en ese momento lin-4 de *Caenorhabditis elegans* por Lee, Feinbaum et al. 1993, ⁸. Delihias menciona el término “RNAs antisentidos que actúan en trans” para nombrar estos elementos ⁹ (Delihias 1995). En esta comparación, se destacan las diferencias considerables entre los sRNAs de origen bacteriano y los de origen eucariótico. Sin embargo, se concluye que a pesar de su divergencia filogenética, su mecanismo de regulación es el mismo; formación de un dúplex de RNA por complementariedad de secuencias con un mRNA blanco o diana. Una revisión más profunda acerca de la abundancia de los sRNAs y las semejanzas y diferencias entre especies bacterianas y eucariontes fue publicada por Storz et. al., 2005. Los autores

describen los distintos modos de interacción y regulación de los sRNAs, y discuten como estas interacciones están presentes tanto en procariontes como en eucariontes a pesar de la divergencia de su origen ¹⁰(Storz, Altuvia et al. 2005).

En el año 2012 M. de la Fuentes et. al. 2012; comparan las características de los ncRNAs involucrados con la termorregulación en procariontes y eucariontes, y concluyen que si bien a nivel molecular la divergencia es evidente, algunas características básicas parecen estar preservadas a través de la evolución ¹¹(de la Fuente, Valera et al. 2012). El examen de los mecanismos de acción muestra puntos comunes y similitudes en los RNAs regulatorios en todos los tipos de organismos. El mecanismo más común por el cual los sRNAs regulan la expresión génica es por apareamiento de bases o complementariedad con transcritos blancos o dianas, por lo que se conoce que actúan como RNAs antisentidos de transcritos mensajeros. Los sRNAs que actúan por apareamiento de bases se pueden agrupar en dos amplias clases: los cis antisentido RNA, que están codificados en la misma localización génica, pero en la cadena opuesta a los RNAs dianas o blancos y por lo tanto su interacción se basa en una complementariedad perfecta, y los sRNAs que actúan en trans que se codifican en una localización distinta en el cromosoma a los RNAs dianas y no presentan una perfecta complementariedad de bases ^{12,13}(Alkan, Karakoc et al. 2006; Bleda, Medina et al. 2012).

1.3 Biogénesis de ncRNA antisentidos.

La síntesis de transcritos antisentidos está presente en los tres dominios de la vida, aunque se ha investigado principalmente en eucariontes. El desarrollo de las tecnologías de NGS ha

revelado que los transcriptomas bacterianos han resultado ser inesperadamente complejos en cuanto a frecuencias de transcripción antisentido. Es destacable mencionar que la transcripción antisentido no había sido objeto de análisis de estudios genómicos principalmente por tres problemas técnicos; el carecer de algoritmos bioinformáticos robustos para su predicción, el hecho de que se consideran como artefactos experimentales generados durante la síntesis de cDNA en el análisis de microarreglos y su posterior interpretación en los datos experimentales; puesto que es difícil diferenciar ncRNAs con funciones regulatorias de artefactos transcripcionales ¹⁴(Georg and Hess 2011).

Una alternativa efectiva para la validación experimental de ncRNAs codificados en cis es la realización de análisis tipo CHIP-chip mediante la inmunoprecipitación utilizando el factor Rho seguido de ensayos de microarreglos o RNA-seq ¹⁵(Peters, Mooney et al. 2009). Los sRNAs codificados en cis en bacterias también denominados asRNAs (antisense RNAs) son tan diversos que no existen características compartidas entre ellos excepto que su transcripción ocurre a partir de la cadena antisentido de alguna unidad transcripción. De acuerdo a su localización son clasificados como divergentes (sobrelapamiento en el extremo 5', head to head), convergentes (sobrelapamiento con el extremo 3', tail to tail) o internamente localizados como asRNAs. El tamaño de los asRNAs es diverso y tienen un rango de 100 a 300 nt como por ejemplo SymR, GadY, y SyR7, pero otros son substancialmente largos incluso mayores que 700 nt ¹⁶(Stazic, Lindell et al. 2011).

La interacción de un asRNAs con su RNA diana resulta en la formación de un dúplex de RNA o RNA de doble cadena (dsRNA). Estos cambios estructurales influyen en la estabilidad y en el tiempo de vida media de ambos RNAs con varias consecuencias para el RNA blanco,

como pueden ser la terminación de la transcripción, la estabilización o degradación del transcripto y la modulación de la traducción ¹⁷(Sorek and Cossart 2010).

Dentro del dominio eucariota los siRNAs comprenden una subclase de RNAs antisentido con una extensión de 21-25 nt que se derivan de RNA de doble cadena de origen exógeno y/o endógeno y han recibido especial atención como reguladores de la expresión de genes a nivel transcripcional y post-transcripcional ¹⁸(Okamura 2012). La biogénesis de los siRNAs en el interior de la célula ocurre cuando la interacción entre dos RNAs de sentidos opuestos, genera un dsRNA que es cortado en fragmentos aleatorios de 21-25 nt de doble cadena por un complejo de proteínas que contienen a Dicer, una endonucleasa específica de tipo RNasa III. Controlan procesos celulares tales como la metilación del DNA en plantas, la inactivación en mamíferos del cromosoma X, la formación de heterocromatina y mecanismos locus específicos como la represión de genes en animales, plantas y recientemente en *S. pombe* ¹⁹ (Carthew and Sontheimer 2009).

1.4 ncRNAs codificados en trans

En contraste a los ncRNAs codificados en cis o antisentidos, los sRNAs codificados en trans se han descrito ampliamente y se ha demostrado su participación en la regulación de genes tanto en bacterias y eucariontes. Dentro de sus características fundamentales es importante mencionar que son expresados como unidades independientes de transcripción, y son inducidos por condiciones específicas del entorno lo que sugiere que funcionan como integrantes de diferentes redes de regulación. Por ejemplo los sRNAs RhyB /SraI que se induce en condiciones limitantes de hierro y SgrS/RyaA que se induce ante elevados niveles

de glucosa-6- fosfato (Collins 2011).

Cerca de un tercio de sRNAs descritos en *E. coli* interaccionan con la proteína Hfq que actúa en el apareamiento de bases entre el sRNA y la molécula diana. Esta proteína es homóloga de la proteína Sm y Sm-I (Sm like proteins) que forman el núcleo del splicing y los complejos de degradación de RNAs mensajeros en las células eucarióticas y archeas ²⁰(Moll, Leitsch et al. 2003).

De forma análoga a muchos sRNAs bacterianos codificados en trans los miRNAs modulan la estabilidad del RNA mensajero y la traducción mediante el apareamiento de bases. Los miRNAs tienen una extensión de 22 nt y se originan a partir de cortes en RNAs con una estructura peculiar en forma de tallo y lazo e intervienen en su biogénesis gran parte de los complejos proteicos involucrados en la generación de los siRNAs. Una de las principales diferencias entre los siRNAs y los miRNAs es que los miRNAs tienen especificidad de cadena y extremos definidos Los miRNAs y por tanto las microRNPs reconocen su diana a través de patrones de apareamiento de bases, que dirige el tipo de respuesta funcional. Una complementariedad perfecta entre el miRNAs y su diana conduce a la escisión del RNA mensajero y su degradación. Sin embargo un apareamiento parcial conduce a regulación mediada por eventos de eliminación de la cola de poliA y modulación traduccional ^{21, 22}(Malphettes and Fussenegger 2006; Racz, Kaucsar et al. 2011).

Diferentes estudios han concluido que la mayoría de los miRNAs están altamente conservados y que los miRNAs son raramente perdidos en la evolución una vez establecidos en las especies, lo que ha servido de base para investigar la filogenia en organismos eucariontes. Por consiguiente, la conservación de miRNAs incluye la identificación de conservadas interacciones miRNA-mRNA y una estrategia apropiada para

su identificación incluye determinar conservación de secuencias de miRNAs, de región codificante de mRNAs y de la región 3' no traducida del mRNA ²³(Lin, Mitra et al. 2014).

1.5 Detección y validación de ncRNAs

Puesto que los procesos celulares se regulan cooperativamente por un extenso número de genes, es importante asumir una visión global de las interacciones sRNA-mRNA. Para este propósito métodos experimentales de alto rendimiento combinados con herramientas computacionales son necesarios para analizar los archivos de expresión de genes. El estudio de sRNAs implica herramientas predictivas basadas en análisis de secuencias, que tenga como salida la interpretación de la estructura secundaria y la interacción entre el sRNA y el mRNA diana. Igualmente las técnicas de biología molecular como la secuenciación masiva, la fusión transcripcional con genes reporteros y los microarreglos de DNA entre otras, han sido cruciales para generar, probar y validar diversos algoritmos computacionales ²⁴(Goutsias and Lee 2007).

Actualmente los secuenciadores de última generación (NGS) tienen una capacidad tal que en una sola línea de una celda de flujo de la plataforma Solexa-Illumina se puede secuenciar el total de RNA (RNA-seq) de un organismo en un momento y bajo una condición determinada. Durante la secuenciación masiva de RNA se generan millones de lecturas que abarcan el transcriptoma completo de un organismo, posteriormente se realiza el mapeo de transcriptos sobre el genoma de referencia o el ensamblaje de novo del transcriptoma, y es seguido por el análisis cualitativo y cuantitativo de los fragmentos de RNA. Las lecturas de secuencias pueden incluir ncRNAs de diferentes familias, RNA que codifica a proteínas y RNA

degradado.

Esta técnica permite detectar la expresión diferencial de genes conocidos, y de regiones no codificantes como regiones intergénicas, intrones y transcripción antisentido ²⁵(Ozsolak and Milos 2011). Además, al igual que los ensayos de microarreglos, el procesamiento y análisis de datos de RNA-seq permite inferir modelos de redes de interacción biológica ¹³(Bleda, Medina et al. 2012).

1.6 Redes de Interacción biológicas

Los modelos de redes biológicas y en particular las redes de regulación de genes son generalmente contexto específico, ya que se basan en datos estructurales, cualitativos y cuantitativos obtenidos en condiciones específicas de células o tejidos. Una alternativa para discernir las similitudes y diferencias presentes en redes de interacción biológica es el estudio de la expresión diferencial de genes en diferentes organismos sometidos a una misma condición o relacionados con el un mismo proceso. Existen varios ejemplos en la literatura, aunque la mayoría compara especies pertenecientes a un mismo dominio evolutivo.

Un ejemplo de estudios relacionados con análisis de evolución convergente es la presencia de circuitos análogos en la regulación de la división celular entre la bacteria gemante *Caulobacter crescentus* y *Saccharomyces cerevisiae* ²⁶. Otro estudio es acerca de la conservación y comparación de redes metabólicas presentes en procariontes y eucariontes al comparar las redes presentes en la bacteria *Helicobacter pylori* y *S. cerevisiae* ²⁷.

Adicionalmente, en un análisis comparativo funcional existe el problema de decidir el punto de corte para los valores de los cambios de expresión en diferentes organismos. Esto se puede evitar determinando un grupo de genes definidos a priori (vías metabólicas, vías de señalización, grupos de genes de interés) que muestran diferencias concordantes estadísticamente significativas, entre dos estados biológicos (fenotipos). Ante este hecho, se reportó la conservación evolutiva de una vía de transducción de señales y una ruta metabólica que responde a la deficiencia de biotina en tres modelos eucariontes (levadura, nemátodo e hígado de rata), analizando transcriptomas y observando la expresión coordinada de ciertos grupos de genes y su comparación con el programa GSEA (Gene Set Enrichment Analysis) ^{28,29}(Velazquez-Arellano, Ortega-Cuellar et al. 2011).

Por otra parte, las redes celulares no son sistemas biológicos aislados y han estado sujetos a una presión selectiva durante la evolución. Por ejemplo, la mayoría de los organismos unicelulares son capaces de usar principalmente la glucosa como fuente de carbono. Sin embargo, cuando las células son crecidas en medio de cultivo, alteran su medio ambiente inmediato y aquellos organismos que son capaces de adaptarse rápida y eficientemente a cambios ambientales, tienen una ventaja selectiva adquirida durante la evolución. Las levaduras, como *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe* han evolucionado mecanismos muy efectivos para adaptarse a diversas concentraciones de diferentes fuentes de carbono. Son capaces por un lado de fermentar glucosa a alcohol cuando hay exceso de ésta, aún en presencia de oxígeno, el cuál es un fenómeno conocido como efecto Crabtree o por el otro lado de llevar a cabo un metabolismo oxidativo cuando la glucosa escasea o cuando la fuente de carbono no es fermentable, como en el caso del

succinato, etanol, y otros carbohidratos ³⁰.

De forma similar la especie bacteriana *Escherichia coli* cuando es crecida en medios ricos con glucosa como fuente de carbono, la glucólisis produce la acumulación de acetato a partir de acetyl Co-A mientras que reprime la expresión de genes para la utilización de fuentes alternas de carbono y así optimizar el uso de glucosa, en un efecto conocido como represión catabólica de carbono . Al igual a como ocurre en levaduras, cuando el carbono es escaso, *E. coli* puede utilizar otras fuentes de carbono como piruvato, glicerol, succinato y fumarato entre otros mientras que reprime los genes de la glucólisis ^{31,32}.

De acuerdo las características estructurales de sRNAs en bacterias y los siRNAs/miRNAs en eucariontes buscar conservación entre los mismos es un reto con pocas probabilidades de éxito. Sin embargo, analizar si genes ortólogos de secuencia y/o función están sujetos a mecanismos post-transcripcionales, formularía nuevas preguntas si los mecanismos de regulación por sRNAs sobre genes específicos son conservados en la evolución o si son un ejemplo especial de adaptación a través de eventos de evolución convergente.

Numerosos programas están disponibles libremente para la detección sRNAs en eucariontes y procariontes, no obstante, la predicción no es completamente eficiente y depende en gran medida de cuán correcto este anotado el genoma y si existen librerías validadas de interacción sRNAs-mRNAs. MirDeep es un ejemplo de software que puede emplearse para la construcción de mapas de secuencias cortas en un genoma de referencia ³³(An, Lai et al. 2013). Por otra parte, TargetScan se ha reportado como el más robusto en la predicción para varios sistemas celulares. La principal razón de su superioridad es que utiliza la información

de la conservación entre especies con lo que reduce el número de falsos positivos, además de que realiza un algoritmo de alineamiento global entre el genoma y librerías de segmentos de las regiones 3' no traducidas^{34, 35}(Thomas, Saito et al. 2011).

Actualmente cerca de 15,000 miRNAs derivados de 142 especies están registrados en la base de datos de miRNAs, miRBase, sin embargo no existen conjuntos de microRNAs provenientes de hongos y levaduras, por lo que hallazgos en el estudio de elementos involucrados en la regulación postranscripcional en *S. pombe* como modelo constituyen nuevos aportes al estado del arte³⁶.

La expresión diferencial entre las diferentes fases de un cultivo no solo se restringe a RNAs codificadores, sino también a small RNAs (sRNAs). Los sRNAs que se asocian a la proteína Hfq constituyen una de las clases más extensas de reguladores post-transcripcionales y su mecanismo de acción es similar a los microRNAs detectados en organismos eucariontes. En un reciente estudio en *Salmonella typhimurium*, se demostró que las regiones 3' no traducidas de algunos mRNAs constituyen un reservorio genómico para la acción de Hfq y la biogénesis de nuevos sRNAs, también este estudio evidenció que la expresión de sRNAs asociados a Hfq muestran marcadas diferencias entre las fases del cultivo, tanto en la cantidad de genes que codifican para sRNAs como en la cantidad de lecturas^{40,41}(Chao, Papenfort et al. 2012).

Teniendo en cuenta los elementos mencionados anteriormente, el objetivo de esta primera etapa de experimentación es establecer el crecimiento de los dos microorganismos modelos seleccionados *E. coli* y *S. pombe*, así como determinar la fase exponencial y el punto de toma de la muestra, que se utilizará como fuente para la obtención de RNA y la posterior

generación de bibliotecas de cDNA a partir de RNA.

1.7 Problema

En las dos últimas décadas en la literatura se ha incrementado el número de publicaciones relacionadas con la identificación y caracterización de ncRNAs. Dentro de las especies de especies de ncRNAs, los sRNAs regulan la expresión de genes a niveles transcripcional y post-transcripcional a partir de la interacción por complementariedad con otra molécula de ácido nucleico o modificando la actividad de una proteína o un complejo de proteínas. La presencia de mecanismos de regulación post-transcripcional de genes se ha descrito tanto en organismos eucariontes como procariontes y estudios de acerca de los RNAs regulatorios se han extendido de forma amplia y paralela en sistemas eucariontes, bacterianos y se encuentra en etapas tempranas en archeas. No obstante, la existencia de analogías en circuitos de regulación con componentes de sRNAs y la relación entre genes ortólogos continúa siendo una interrogante.

Considerando las premisas anteriores, el presente proyecto tiene la finalidad de determinar similitudes y diferencias en las redes de interacción biológica relacionadas al metabolismo central del carbono, donde se encuentre elementos de sRNAs en dos organismos modelos *Escherichia coli* y *Schizosaccharomyces pombe* en dos escenarios metabólicos (crecimiento en fuentes de carbono fermentable y no fermentable). Los resultados generados brindarán datos cualitativos y cuantitativos para comprender la evolución de circuitos regulatorios entre organismos procariontes y eucariontes, priorizando el uso de las herramientas bioinformáticas para la de detección de sRNAs y realizar inferencias acerca de los cambios que estos elementos regulatorios han tenido en organismos de diferentes reinos.

CAPÍTULO II : HIPÓTESIS Y OBJETIVOS

2.1 Hipótesis

El análisis del transcriptoma de *E. coli* y *S. pombe* durante el crecimiento en glucosa y glicerol/acetato constituye una herramienta que permite inferir analogías en circuitos de regulación de la expresión génica, con especial atención a aquellos genes relacionados al metabolismo central del carbono y donde participen elementos regulatorios de tipo ncRNAs.

2.2 Objetivos

Objetivo General

Caracterizar la expresión de mRNAs y sRNAs de *E. coli* y *S. pombe* a partir de datos de RNA-seq obtenidos en dos condiciones de estudios; crecimiento en glucosa y crecimiento en glicerol/acetato.

Objetivos específicos

1. Determinar el conjunto de genes expresados de forma diferencial en *E. coli* y *S. pombe* mediante la metodología de RNA-seq durante el crecimiento en medio YE suplementado con glucosa al 2% y glicerol/acetato 2% (10:1).
2. Describir el comportamiento de factores de transcripción y la abundancia de transcritos de tipo de ncRNA en *E. coli* y *S. pombe* en condiciones similares de crecimiento.
3. Inferir analogías en circuitos de regulación mediadas por ncRNAs sobre genes ortólogos de rutas metabólicas asociadas al metabolismo central del carbono.

CAPÍTULO III : MATERIALES Y MÉTODOS

3.1 Cepas de microorganismos

En el presente estudio se empleo la cepa de *Escherichia coli* K-12 MG1655 donada por el Dr. Enrique Morett (IBT, UNAM) y la cepa de *Schizosaccharomyces pombe* 972 h- donada por la Dr. Elena Hidalgo (Universitat Pompeu Fabra, Barcelona, España). El estudio se llevo a cabo en el laboratorio de Ingeniería en biocatálisis del Instituto de Biotecnología de la UNAM.

3.2 Medios y condiciones de cultivo

La cepa de *E. coli* se cultivo rutinariamente en agar LB para su mantenimiento y conservación a 4°C mientras que la cepa de *S. pombe* se cultivo en agar YES para este mismo procedimiento. Previo a la inoculación en el medio de obtención de biomasa para la extracción de RNA, los organismos se cultivaron en los medios de cultivo limitados en nutrientes agar R2A y agar YE suplementado con glucosa al 1%.

Los cultivos celulares para lo obtención de RNA se llevaron a cabo en volúmenes de 50 ml y se crecieron aeróbicamente en matraces de 250 ml con medio YE en dos tratamientos o condiciones de estudio; medio suplementado con glucosa al 2 % w/v como fuente de carbono y medio adicionado con glicerol 2% y acetato de sodio al 0.2 % w/v. La plataforma de rotación se mantuvo a 220 rpm. La temperatura de incubación y el pH del medio se

ajustaron a los parámetros óptimos descritos en la literatura; 37°C y 7.2 ± 0.2 para *E. coli* y 30°C y 6.5 ± 0.4 para *S. pombe*.

Las células crecidas en agar R2A o YE según sea el caso, se cosecharon y fueron lavadas dos veces en solución tampón de sales fosfato (PBS) y colectadas por centrifugación a 5000 g. Posteriormente se suspendieron en PBS y se determinó la densidad óptica a 600 nm de longitud de onda (Zeiss Spectrophotometer, model PMQ II). El volumen de inóculo se ajustó en cada momento para que la densidad óptica inicial del cultivo estuviera en un rango de 0.02-0.05 DO. El incremento de la biomasa se determinó mediante la turbidez del medio posterior a la inoculación. Las lecturas se realizaron tomando muestras de cultivo (0.5 ml) periódicamente. En dos ocasiones se realizó la siembra de diluciones seriadas de muestras para observar la conservación de las características culturales y realizar conteo de unidades formadoras de colonias (ufc) por mililitro. En cada caso los experimentos se realizaron con tres cultivos independientes como réplicas biológicas.

Para la extracción de RNA total, los cultivos se cosecharon en la fase media exponencial y posteriormente se les añadió RNA Later (Ambion) para preservar la integridad del RNA y se incubaron por un minuto en hielo. Posteriormente se colectó la biomasa mediante centrifugación a 5000 g por un espacio de cinco minutos y se almacenaron a -80°C.

3.3 Extracción y preparación del RNA total

El RNA total se extrajo empleando fenol/tiocianato de guanidinium provisto por el reactivo Tri-Reagent (Trizol) de Zymo Research, seguido de precipitación con isopropanol. El

procedimiento se ejecuto con la indicaciones del fabricante. Las muestras obtenidas se procesaron con DNAsa I (Fermentas, life technologies) para la eliminación del ADN residual. y se reconcentraron con el kit RNA Clean & Concentrator™ (Zymo Research). La calidad del RNA se determinó electroforesis en gel de agarosa empleando MOPS y formamida como agente desnaturalizante. Además, la concentración y calidad se evaluó espectrofotométricamente con la tecnología de Nanodrop y Bioanalyzer 2100, Chip RNAnano (Agilent). Para cada organismos se obtuvieron tres réplicas biológicas por condición : Gluc_1, Gluc_2, Gluc_3 (Crecimiento en Glucosa) y Glyc_1, Glyc_2, Glyc_3 (Crecimiento en Glicerol/ Acetato).

3. 4 Construcción de bibliotecas para la secuenciación masiva de RNA.

Con el propósito de maximizar la diversidad de secuencias correspondiente a mRNAs en las librerías de cDNA se procedió a reducir a cantidad de rRNA por muestra biológica. Según el caso se removieron los transcritos 16S y 23S, para bacteria y 18S y 26S para levaduras, para ello se utilizó el kit comercial RiboMinus™ Yeast/ Bacteria Transcriptome Isolation Kit y se procedió de acuerdo a las indicaciones del fabricante. Las bibliotecas se prepararon utilizando el kit TruSeq RNA sample Prep Kit (Illumina, San Diego, CA, USA), de acuerdo a las instrucciones del fabricante con la siguiente modificación el RNA total se fragmento a mediante metodos químicos a una extensión de 150 pb (extensión correspondiente a un exon promedio) solo realizaron 10 ciclos de amplificación de PCR y se selecciono en gel un abanda de 200- 250 p. El tamaño seleccionado corresponde a un inserto de 130-180pb (adaptadores añaden ~70pb) y se repurificó con el *kit Quiagen Gel Extraction* siguiendo las indicaciones del fabricante y se resuspendió en un volumen de 30µl. El material amplificado

se diluyó a una concentración de 4pM y se cargó en una línea de *flow-cell* del *NextSeq 500* . La secuenciación se llevó a cabo en una corrida de 36 ciclos. En total, se construyeron 12 librerías con lecturas tipo pair-end con una extensión de 75 pb. Las librerías se secuenciaron en la Unidad de Secuenciación Masiva Ácidos Nucleicos del IBT-UNAM.

3.5 Análisis Bioinformático

3.5.1 Procesamiento de archivos crudos en formato fastq

Las secuencias obtenidas (secuencias en formato fastq), se procesaron con la aplicación *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y se generó un reporte acerca de la cantidad, calidad y longitud de las mismas. Se utilizó la aplicación *Trimmomatic* () para el corte y eliminación de adaptadores , barcodes , filtración de secuencias que presenten baja calidad y ambigüedad (no definición de nucleótidos y repetición de una misma base).

3.5.2 Alineamiento y filtrado de lecturas

El alineamiento se realizó con el programa Bowtie ⁴²(Langmead, Trapnell et al. 2009) y Tophat2 ⁴³ para *E. coli* y *S. pombe* respectivamente. Los genomas de referencias se obtuvieron de la base de datos Genebank del NCBI (<ftp://ftp.ncbi.nih.gov/genbank/genomes> [/Bacteria/Escherichia_coli_K_12_substr_MG1655uid225/](ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/Escherichia_coli_K_12_substr_MG1655uid225/)) para *E. coli* y del sitio

(ftp://ftp.ebi.ac.uk/pub/databases/pombase/pombe/Chromosome_Dumps/fasta/) para la obtención de datos en *S. pombe*.

El filtrado de lecturas se realizó con el empleo de las herramientas dispuestas en *bmap* y *samtools* ⁴⁴(Li, Handsaker et al. 2009). Se ejecutaron los siguientes comandos para el filtrado de lecturas procesadas.

1. Filtro inicial :

1. `Trimomatic` (Eliminación de adaptadores y recorte de regiones de menor calidad)

2. Alineamiento :

1. `bowtie2 -p 2 -x gen_eco -U input_eco.fastq | samtools view -Sb -> $out_path/ out.bam`

2. `tophat2 -G spo.gff -library-type fr-firststrand -p 1 -o $out_path gen_spo input_spo.fastq`

3. Filtro final

1. `samtools sort -n file.bam file_s.bam #` (Ordenamiento de las lecturas por nombres)

2. `samtools fixmate -r file_s.bam file_fs.bam #` (ajuste de extensión de lecturas y eliminación de alineamientos secundarios).

3. `samtools rmdup -S file_s.bam file_fs.bam #` (Eliminación de duplicados ópticos de PCR).

4. `samtools sort file_fs.bam file_ns.bam #` (Ordenamiento por coordenadas).

```
5. samtools view -bq 30 file_ns.bam file_mapped.bam #  
   (Eliminacion de alineamiento de baja calidad)  
6. samtools index file_mapped.bam file_mapped.bai #  
   Creación de índice de ordenamiento.
```

3.5.3 Abundancia de transcritos y Expresión Diferencial de Genes

La información de las posiciones de mapeo de los datos de secuenciación se obtuvo mediante la función *coverage* de la aplicación *Bedtools* y la cantidad de lecturas mapeadas será considerada como la cantidad de transcritos o expresión de un gen.

El análisis de expresión diferencial se llevó a cabo con el paquete *edgeR* de la librería Bioconductor, plataforma R. Se determinó la dispersión de los datos a partir de la normalización por el tamaño de las librería (entre réplicas y entre tratamiento) mediante el método TMM disponible en el paquete *edgeR* de Bioconductor ⁴⁵. En ambos casos se determinó el coeficiente de variación biológica y se analizó la expresión diferencial considerando la variación gen-específica. Los genes con un logaritmo de razón de cambio (log FC del inglés log *foldchange*) mayor que 1.5 y con un FDR < 0.05 se consideraron como diferencialmente expresados al comparar las condiciones de crecimiento.

3.5.4 Análisis de enriquecimiento de vías metabólicas y procesos biológicos.

Para localizar los genes expresados diferencialmente en sus correspondientes rutas metabólicas, se realizó un análisis de las listas de DEGs obtenidas para ambas condiciones experimentales en la Enciclopedia de Kyoto de Genes y Genomas (KEGG) Pathway Color. De igual forma el análisis de los procesos biológicos enriquecidos se realizó por la base de datos STRING y la herramienta BINGO de la plataforma de Cytoscape.

CAPÍTULO IV : RESULTADOS

4.1 Determinación de la expresión diferencial de genes en *E. coli* K-12 y *S. pombe* en dos escenarios metabólicos diferentes.

4.1.1 Optimización del crecimiento y establecimiento de la fase exponencial

Con el objetivo de crecer ambos microorganismos en condiciones homogéneas, se inoculó la cepa *E. coli* K-12 MG1655 en el medio de cultivo YE, que es ampliamente utilizado en el crecimiento vegetativo de *S. pombe*. Para cada organismo, se utilizó la misma formulación de medio de cultivo y se analizaron dos tratamientos; medio YE con glucosa 2 % y medio YE con glicerol al 2% y acetato de sodio al 0.2%.

Tanto en *E. coli* como *S. pombe* se realizó un crecimiento primario en un medio sólido reducido en nutrientes, posteriormente un preinóculo en las condiciones mencionadas y a partir de este crecimiento, se ajustó la densidad celular a 0.06 de DO en medio fresco para iniciar el cultivo experimental y tomar la muestra en la fase exponencial temprana. La figura 1, muestra las curvas de crecimiento obtenidas para cada microorganismo en las condiciones objeto de estudio y cada punto representa la media de tres cultivos independientes.

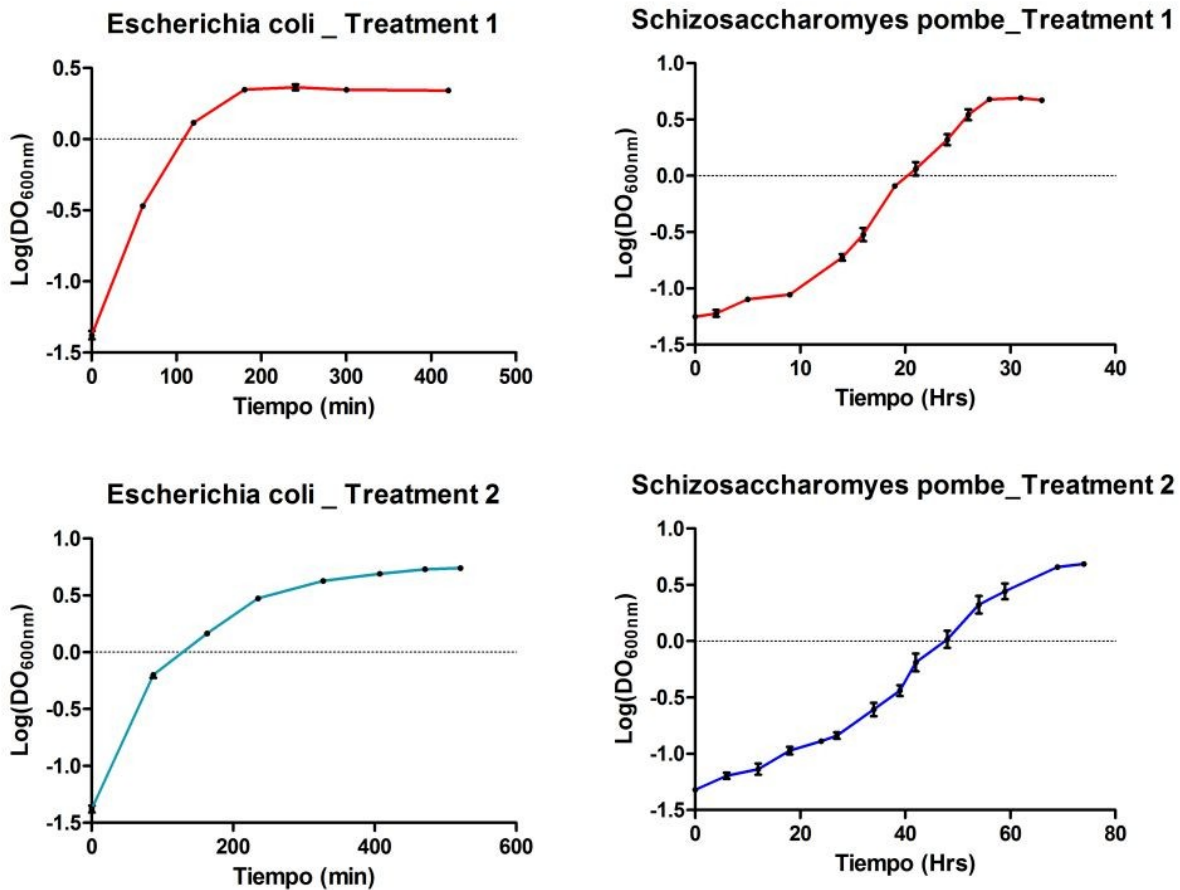


Fig. 1: Crecimiento de *E. coli* y *S. pombe* al utilizar glucosa 2 % w/v (curva de color rojo) o glicerol 2% / acetato 0,2% (curva de color azul) como fuente de carbono. Cada punto representa la media de DO a 600 nm de tres cultivos independientes \pm 1 SD.

Las curvas de crecimiento muestran que ambos organismos utilizan satisfactoriamente las fuentes de carbono empleadas, sin embargo ciertas diferencias son observadas al utilizar glucosa o glicerol/acetato como fuente de carbono. En el caso de *E. coli*, el tiempo de duplicación está en el rango entre los 20 – 25 min y el crecimiento se inhibe por completo a los 200 min post-inoculación al emplear glucosa como fuente de carbono. El tiempo de duplicación en glicerol/actetato es ligeramente mayor, 38 – 45 min, y la fase estacionaria se alcanza después de los 400 min post-inoculación. Los valores de DO en la fase estacionaria

son mayores que 3 unidades de DO al utilizar glicerol/acetato como fuente de carbono.

Por su parte, la levadura *S. pombe* en experimentos iniciales mostró una fase de adaptación mayor a 20 h en el crecimiento en glicerol/acetato. En contraste, esta fase lag no se aprecia durante el crecimiento en glucosa. Bajo las condiciones experimentales de estudio, los rangos del tiempo de duplicación fueron entre 2-3 h y 7-8 h para el medio YE adicionado con glucosa y glicerol/acetato respectivamente. La entrada en fase estacionaria se alcanza a las 30 h (glucosa) y a 70 h (glicerol/acetato) sin diferencias en la DO máxima alcanzada.

4.1.2 Extracción de RNA total

Para la extracción y purificación de RNA total se utilizaron tres métodos diferentes: columnas de QUIAGEN, Kit Mirvana y extracción con Trizol con posterior precipitación empleando isopropanol, Figura 2. Al final, se seleccionó el método de Trizol por la razón de que permitía el tratamiento de mayor cantidad de biomasa, mayor eficiencia de lisis y menor tiempo de manipulación de la muestra biológica. No obstante, a las muestras se les adicionó el aditivo de microRNA, presente en el kit de Ambion, e inhibidores de RNAsas para garantizar la integridad de RNAs de diferentes extensiones. Para la extracción de RNA total de *E. coli* se empleó un volumen de 10 ml de cultivo mientras que en el caso de *S. pombe* fue de 15 ml.

Las muestras de RNA total se mostraron libres de contaminantes fenólicos y proteínas al ser cuantificadas con el equipo Nanodrop. Igualmente, al ser analizadas por Bioanalyzer 2100, se mostró la integridad de los RNA ribosomales y se alcanzó un RIN entre 7 – 8.5,

Figura 3. El perfil de los RNA totales mostró un pico en la región de RNAs menores a 200 nucleótidos para ambos organismos.

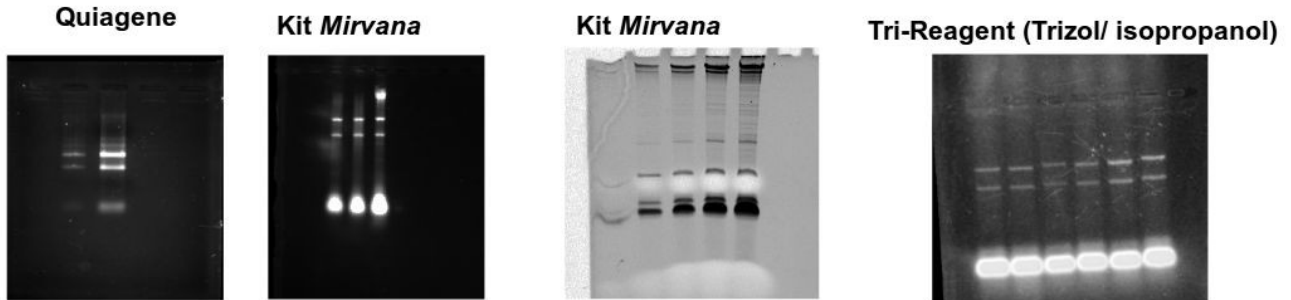


Fig. 2: Electroforesis de RNA total de *E. coli* y *S. pombe* por diferentes métodos. 1. RNA total de *E. coli*, en gel de agarosa al 1% extraído por columnas de Quiagen. 2 Electroforesis en gel de Agarosa RNA total de *E. coli* extraído con el Kit Mirvana (Ambion). 3 Electroforesis en gel de PAGE al 10 % y Urea de RNA total de *E. coli* extraído con el Kit Mirvana (Ambion). 4. Electroforesis de RNA total de *S. pombe* en gel de agarosa al 1%.

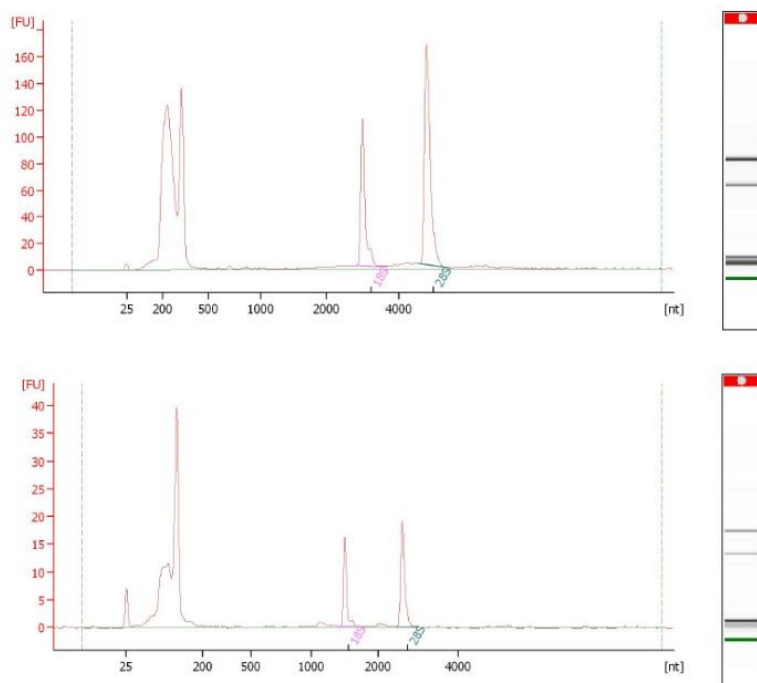


Fig. 3. Cuantificación e integridad de RNA total de *E. coli* (superior) y *S. pombe* (inferior). Cada muestra representa el análisis en Bioanalyzer 2100, Chip RANano (Agilent) de 1 μ l de RNA total.

4.1.3 Estadísticas generales de los datos de RNA-seq.

En el presente estudio se generaron doce bibliotecas de RNA-Seq. Las bibliotecas se realizaron con lecturas pareadas, cadena específicas y las lecturas tienen una extensión de 75p. Las bibliotecas se construyeron de la siguiente forma : seis para cada organismo, donde tres réplicas biológicas independientes se realizaron para cada condición. Una vez realizado el llamado de bases, las cantidades de lecturas en formato fastq oscilaron entre 7.8-9.3 millones y 10.1-13.6 millones para *E. coli*, en el crecimiento empleando glucosa y glicerol/acetato, respectivamente. En cambio para *S. pombe*, se generaron bibliotecas con un rango entre 6.3 -7.3 millones de lecturas en ambas condiciones (Figura 4). Hay que hacer notar, que en todas las bibliotecas el índice de calidad de los archivos fastq tuvo un valor promedio de $Q_{\text{phred}} > 34$. Así mismo, las lecturas de sobrerrepresentación de adaptadores, artefactos y secuencias con bases ambiguas no excedió el 1 % del tamaño de la biblioteca.

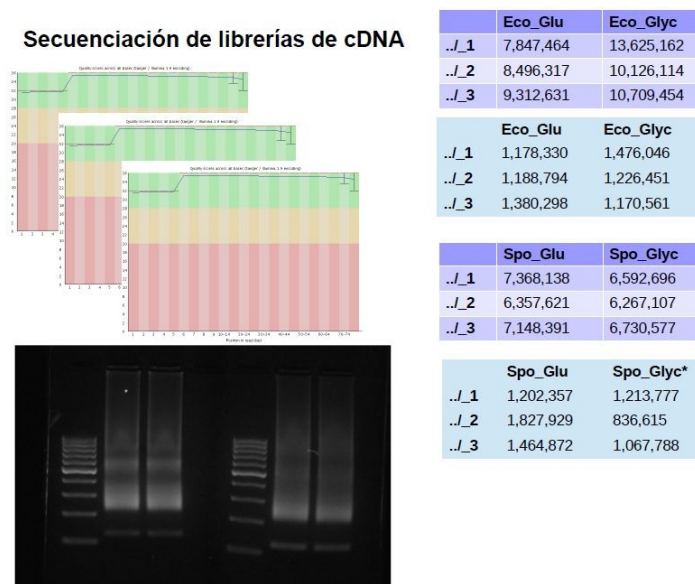


Fig. 4: Generación de bibliotecas y análisis global de archivos de secuenciación. Las bibliotecas se realizaron de acuerdo al protocolo de Illumina y se secuenciaron por multiplexing. Los cuadros muestran el total de las lecturas originales por réplica en fastq (tabla) y las lecturas filtradas y alineadas en el genoma de referencia, archivos bam (cuadro azul).

El alineamiento con Bowtie2 (*E. coli*) y Tophat2 (*S. pombe*) se realizó sin admitir errores de complementariedad. Una vez obtenidos los archivos de alineamiento en formato *.bam, se eliminaron aquellas lecturas que corresponden a duplicados ópticos o duplicados potenciales de PCR (lecturas con idénticas coordenadas externas) y los alineamientos que presentaban un valor de MAPQ (mapping Quality, Phred Scale) inferior a 30.

Considerando lo anterior, el número de lecturas final descendió considerablemente para cada réplica biológica, (Figura 4, recuadros azules). La media geométrica de lecturas sin ambigüedad sobre el genoma de referencia de *E. coli* K-12 MG1655 fue de 1 245 801.36 (\pm 113 706.017) y 1 284 442.64 (\pm 162656.31) para el crecimiento en glucosa y glicerol/acetato respectivamente. Los valores de las lecturas mapeadas en *S. pombe*, para el crecimiento en glucosa, fueron de 1 341 882.69 (\pm 52 610.18) mientras que al emplear glicerol/acetato se mapearon 1 147 150 (\pm 119 184.98). La observación de las lecturas mapeadas sobre los archivos de alineamiento (archivos tipo BAM) mostró diferencias de conteo para genes en las distintas condiciones estudiadas en ambos microorganismos (Fig. 5).

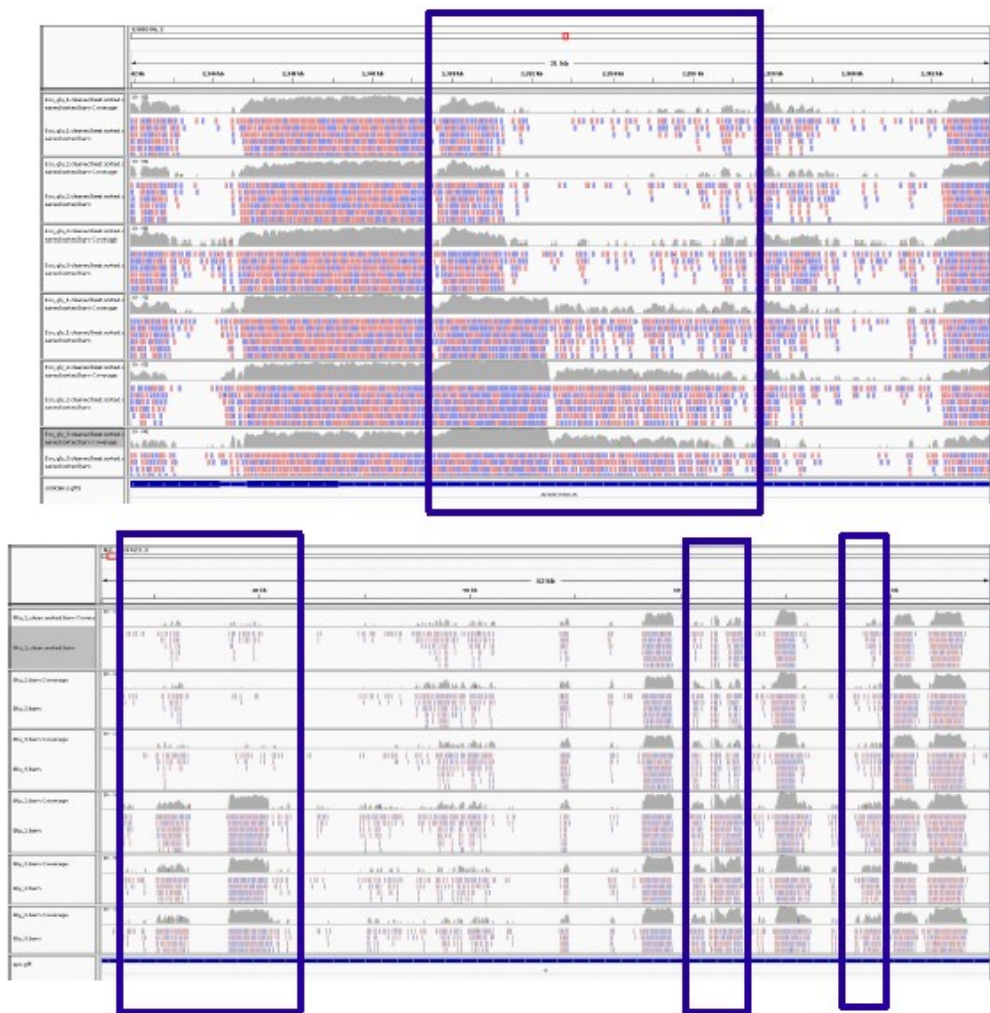


Fig. 5: Diferencias en la cantidad de transcritos a partir del análisis de los archivos de alineamiento sobre los genomas de referencias. Los archivos bam se visualizaron en el programa IGV. Fragmento del cromosoma de *E. coli* (arriba) y fragmento del cromosoma I de *S. pombe* (abajo). El orden de las filas es el siguiente; de arriba hacia abajo: Crecimiento de la glucosa (1-3) y crecimiento en glicerol / acetato (4-6). Los genes con diferente expresión son resaltados en las cajas azules.

4.1.4 Similitud y Variación entre las réplicas biológicas.

Las lecturas que se alinearon al archivo de anotación de características genéticas con dependencia al sentido de los genes se tabularon como conteo de transcritos. Para verificar la reproducibilidad de los experimentos, se calcularon los valores de correlación pareada por

el método de Spearman. Los resultados indicaron que las tres réplicas biológicas analizadas para cada condición poseen altos valores de correlación entre ellas.

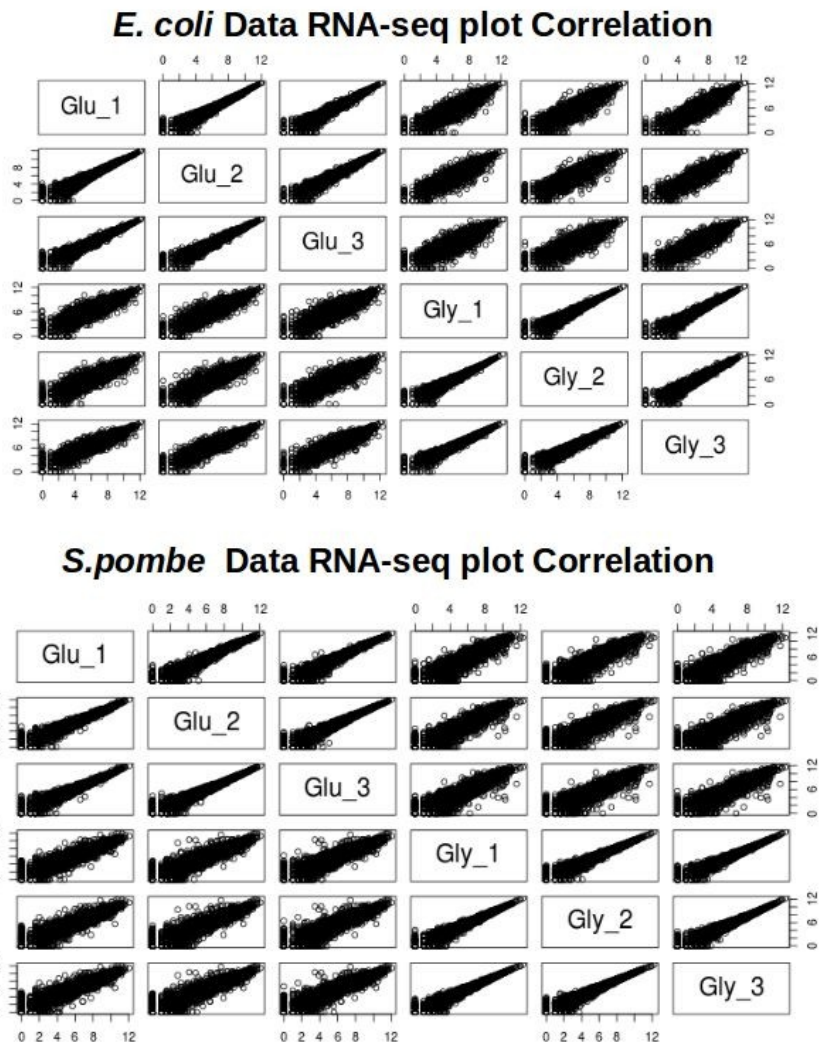


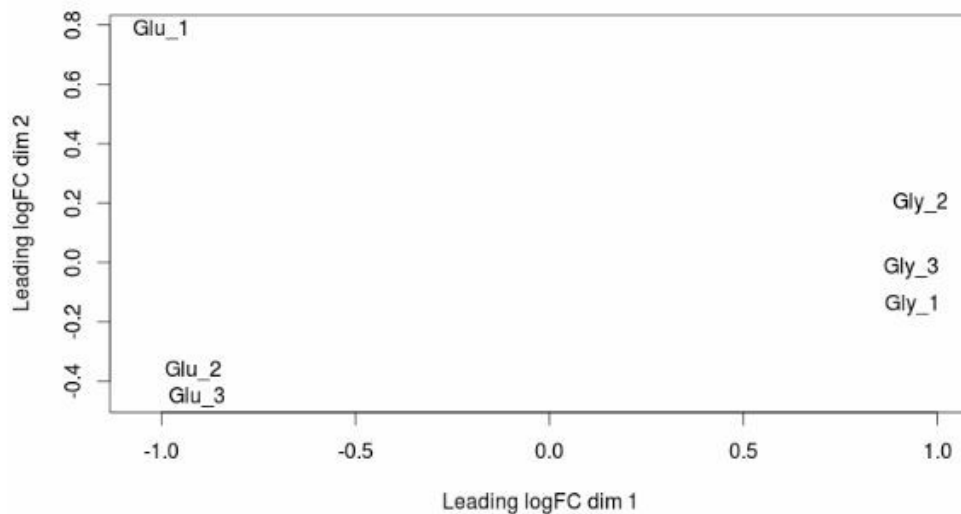
Fig. 6: Correlación de Spearman de los datos crudos de conteo cadena específicos entre réplicas biológicas para datos de RNA-seq de *E. coli* y *S. pombe*.

Los índices de correlación (r^2) fueron mayores a 0.98 para las muestras biológicas pertenecientes al mismo tratamiento en *E. coli*, mientras que la comparación para muestras

de diferentes condiciones estuvo entre 0.89-0.93. En *S. pombe*, el rango de valores de correlación fue de 0.97-0.98, para muestras pertenecientes al mismo tratamiento en tanto que las comparaciones entre condiciones diferentes arrojaron valores entre 0.87 y 0.91 (Figura 5).

Las tablas de conteo para cada biblioteca se normalizaron mediante el paquete de edgeR, de Bioconductor a través del método de TMM. El gráfico de escala multidimensional mostró que el principal factor de variación entre las muestras biológicas es el tratamiento objeto de estudio (la utilización de diferentes fuentes de carbono). La segunda dimensión de variación entre las muestras es precisamente la variación entre las muestras biológicas pertenecientes al mismo tratamiento (Figura 6). Hay que hacer notar que en ambos organismos las réplicas pertenecientes al tratamiento en glucosa tuvieron una mayor dispersión que las muestras crecidas en glicerol/acetato.

E. coli Data RNA-seq plot MDS



S. pombe Data RNA-seq plot MDS

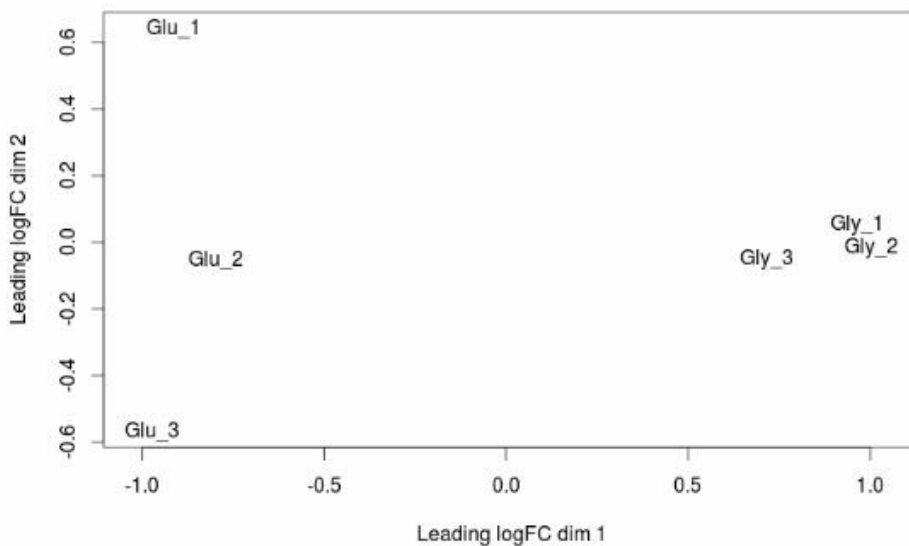


Fig. 7: Gráfico de escala multidimensional (MDS) de los archivos de RNA-seq de *E. coli* (sección superior) y *S. pombe* (sección inferior). Las muestras se agruparon en la dimensión 1 por tratamiento biológico (Glucosa vs. Glicerol/acteto) mientras que la dimensión 2 por la variación entre réplicas biológicas. Cada réplica biológica se obtuvo de cultivos independientes.

4.1.5 Coeficiente de Variación Biológica.

En un conjunto de datos de RNA-seq, el índice de dispersión está asociado al coeficiente de variación biológica. La estimación de la dispersión se ejecutó a través de la función `estimateDisp` implementando la similitud ponderada empírica de Bayes. El índice de dispersión gen específica para *E. coli* tuvo un rango de 0.0041-0.029 con una dispersión general de 0.01089. En los datos obtenidos para *S. pombe* se obtuvo un rango de 0.0022-0.3991 con una dispersión general de 0.055. El coeficiente de variación biológica (BCV) constituye la raíz cuadrada de los índices de dispersión. En la Figura 8 y 9 se muestra como el BCV es una función del logCPM en ambos organismos. El BCV disminuye hasta alcanzar un valor constante cuando se incrementan los niveles de expresión hecho que se asocia a la disminución de la variación técnica.

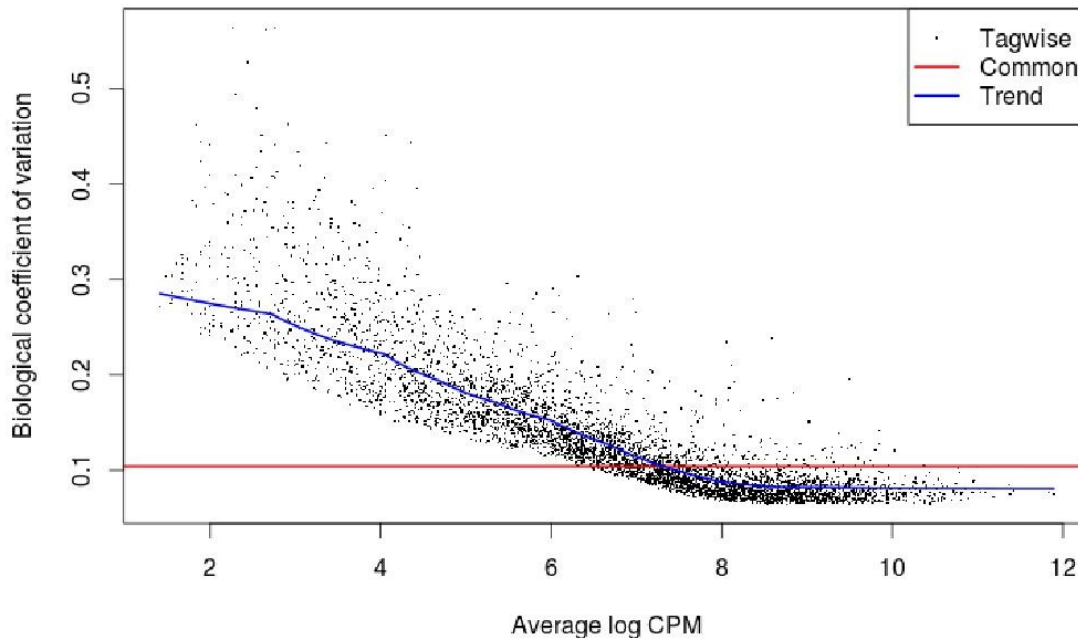


Fig. 8: Coeficiente de Variación Biológica para el conteo de lecturas en *E. coli*. Cada punto representa el coeficiente de dispersión específico de cada gen entre las réplicas biológicas.

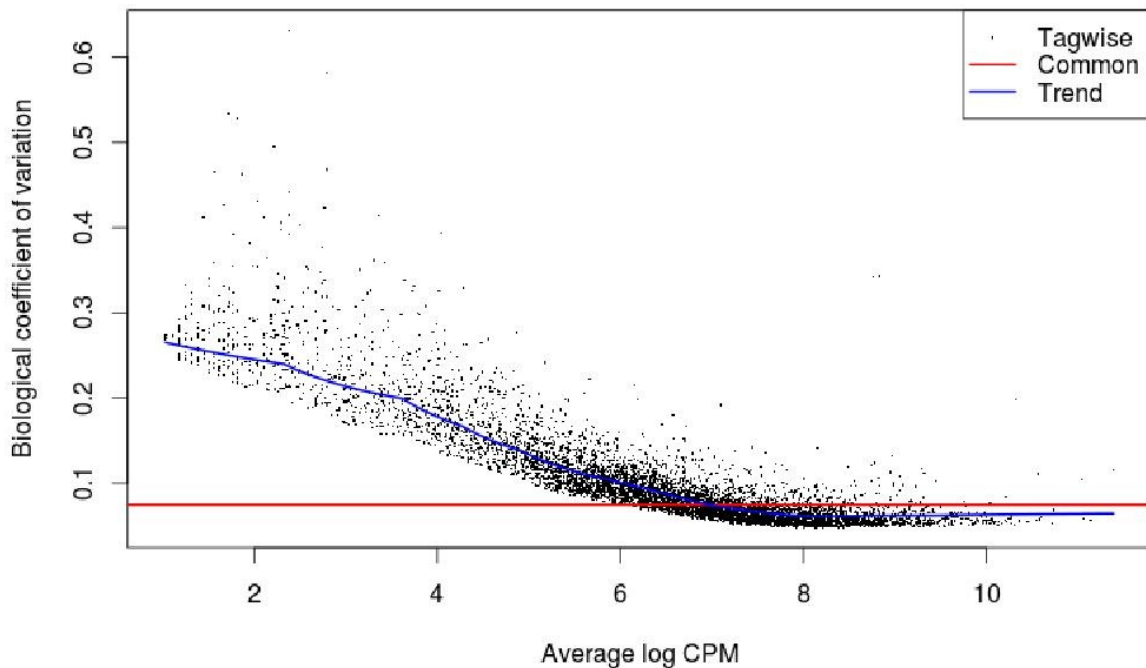


Fig. 9: Coeficiente de Variación Biológica para el conteo de lecturas en *S. pombe*. Cada punto representa el coeficiente de dispersión específico de cada gen entre las réplicas biológicas.

4.1.6 Análisis de expresión diferencial de genes al utilizar glucosa o glicerol/acetato como fuente de carbono y energía.

La expresión específica de los genes, mostró que existe una fuerte asociación entre la expresión relativa de los mismos y su dispersión, hecho que corresponde a una distribución binomial negativa de los datos obtenidos. Para la detección de genes se expresan diferencialmente, se seleccionaron aquellos genes que tuviesen un valor de CMP >3. De acuerdo al parámetro anterior, se conformo un conjunto de 4538 genes para *E. coli* y 6648 genes para *S. pombe*.

El análisis de la expresión diferencial de genes se llevó a cabo utilizando el paquete edgeR a través de métodos estadísticos basados en modelos lineales generalizados. Solo los genes con un módulo de $\log_2 FC > 1.0$ y un valor de $FDR < 0.05$ se consideraron como diferencialmente expresados. El análisis estadístico identificó un total de 779 genes expresados diferencialmente en *E. coli* (Figura 10) y 793 en *S. pombe* (Figura 11).

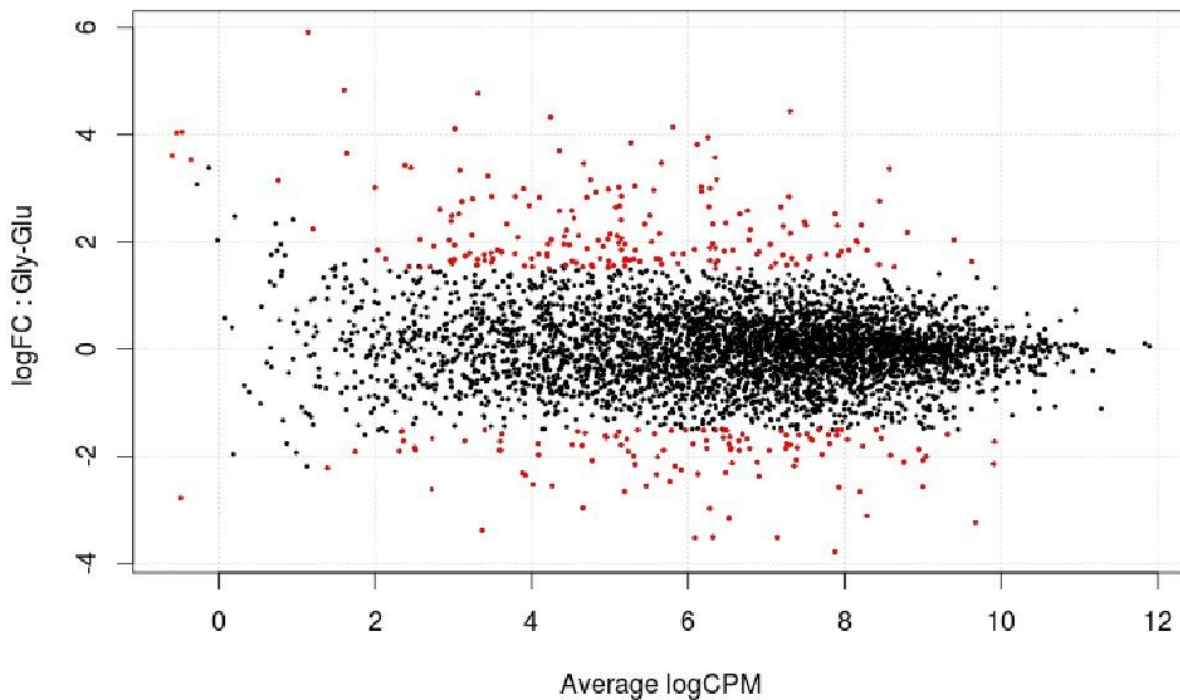


Fig. 10: Expresión diferencial de genes en *E. coli*. Para la realización del análisis se seleccionaron genes con una sumatoria de $\log_2 CPM > 3$ en *E. coli* (4538). Los genes con un $\log_2 FC > 1.0$ y un $FDR < 0.05$ se consideran como diferencialmente expresados.

De los genes expresados en *E. coli*, 438 estuvieron sobre expresados durante el crecimiento en glicerol/acetato y 351 se durante el crecimiento en glucosa. En cuanto *S. pombe* se

detectaron 462 genes en el crecimiento glicerol/ acetato y 331 en glucosa. Además, los valores de logFC fueron mayores en el crecimiento en glicerol/acetato en ambos organismos.

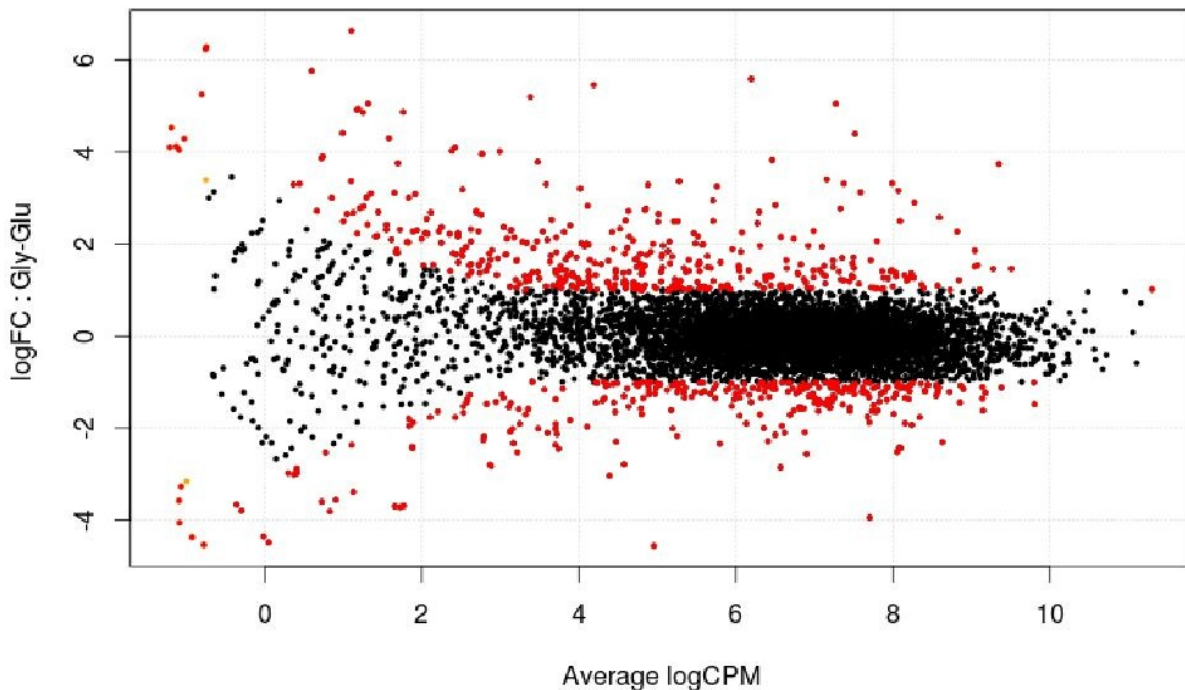


Fig. 11: Expresión diferencial de genes en *S.pombe*. Para la realización del análisis se seleccionaron genes con una sumatoria de logCPM > 3. Los genes con un $\text{Log}_2 \text{FC} > 1.0$ y un $\text{FDR} < 0.01$ se consideran como diferencialmente expresados.

La Figura 12 muestra la distribución de los genes expresados diferencialmente en ambas condiciones en tres categorías seleccionadas; genes que codifican a factores de transcripción (color gris), genes codificadores a ncRNAs (color negro) y genes codificadores a proteínas, que no se encuentran anotados como factores de transcripción, (entramado a rayas). Los genes cuyos productos que regulan la expresión génica (ncRNA y factores de transcripción) corresponden al 3% y 5% durante el crecimiento en glucosa y glicerol/acetato en *E. coli*. Por su parte, esta sección en *S. pombe*, representa el 23 % y el 56 % de los

genes diferencialmente expresados. De forma general los ncRNAs tienen mayor sobre expresión en glicerol/acetato que en glucosa.

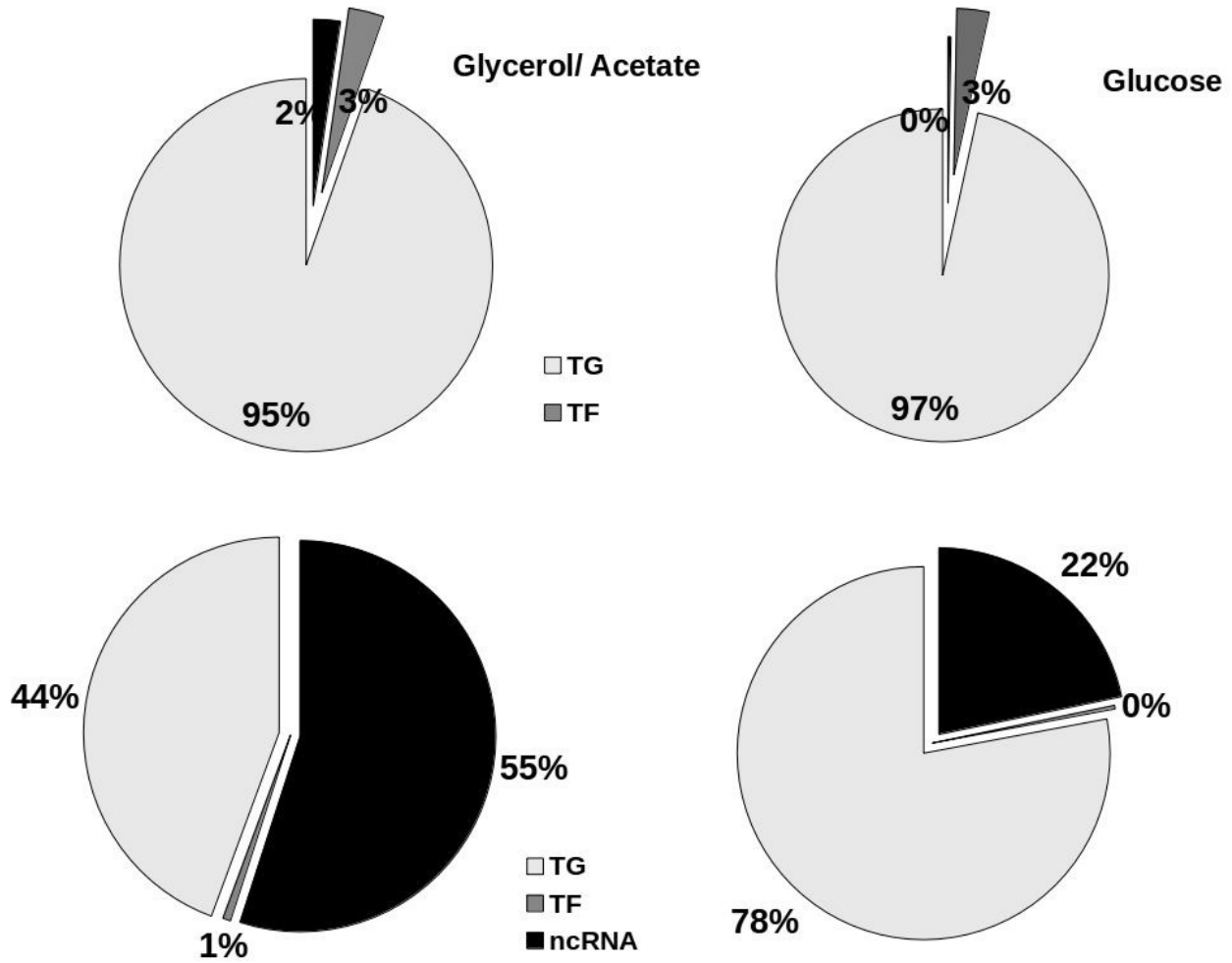


Fig. 12: Distribución de genes expresados diferencialmente en *E. coli* (superior) y *S. pombe* (Inferior). TF (genes que codifican a factores de transcripción) son representados con un área gris oscuro, ncRNA son representados con un área negra mientras que genes que codifican a proteínas que no codifican a factores de transcripción están representados con un área gris claro.

Teniendo en cuenta las dos condiciones o tratamiento analizadas, los genes *fruB*, *nirB* y *pyrB* son los de mayor logFC en glucosa en *E. coli* con menor índice de FDR. De forma similar, en *S. pombe* se detectaron los genes SPAC821.09, SPAC26H5.09c, y SPAC19D507

que están involucrados en los procesos de citocinesis, y en la generación de precursores metabólicos y energía.

Por otra parte en relación con el crecimiento en glicerol/acetato, los genes con mayor logFC y menor probabilidad de corresponder a un faso positivo son; *glpD*, que codifica a la glicerol 3-fosfato deshidrogenasa aeróbica, *frmB* y *frmA* que están involucrados en la oxidación formaldehído endógeno en *E. coli*. De forma análoga, en *S. pombe* se encuentran SPBC1348.14c y SPCC548.07c, relacionados con la entrada de hexosas a la célula, y el gen SPAC13F5.03c cuyo producto es la glicerol deshidrogenasa mitocondrial, enzima relacionada con el metabolismo del glicerol y la obtención de energía.

4.1.7 Enriquecimiento funcional de procesos celulares y de vías metabólicas.

La categorización funcional de los genes expresados diferencialmente en *E. coli* y en *S. pombe* se realizó utilizando la base de datos STRING. Las figuras 13 y 14 muestran el número de GO agrupados en categorías funcionales. De forma general, ambos organismos tuvieron un mayor número GO asociados a vías metabólicas, funciones moleculares y procesos biológicos durante el crecimiento en glucosa. En esta condición, la biosíntesis de aminoácidos, ribosomas, el metabolismo del nitrógeno entre otros, tuvieron una relevancia común tanto en *E. coli* como en *S. pombe* relacionando el crecimiento de la glucosa con la proliferación celular.

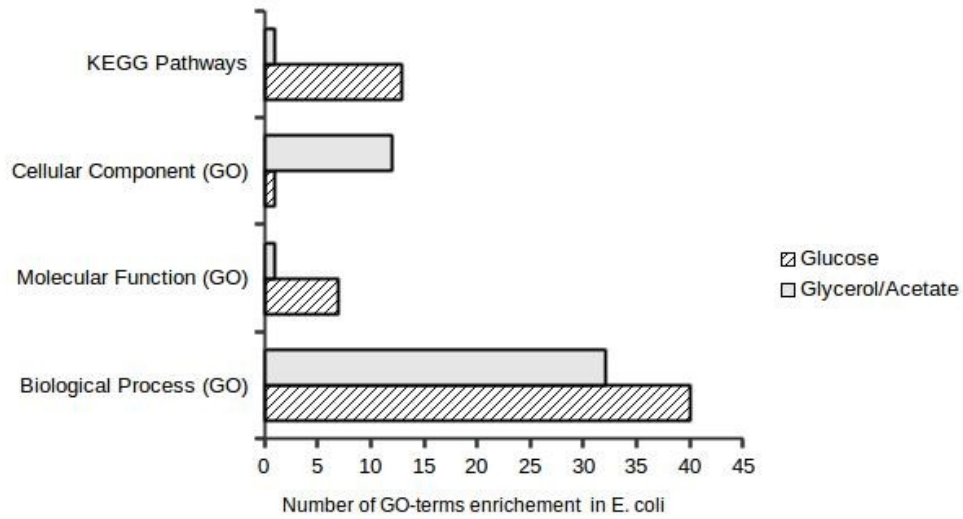


Fig. 13: Comparación de GOs enriquecido funcionalmente en el crecimiento en glucosa y glicerol/acetato en *E. coli*.

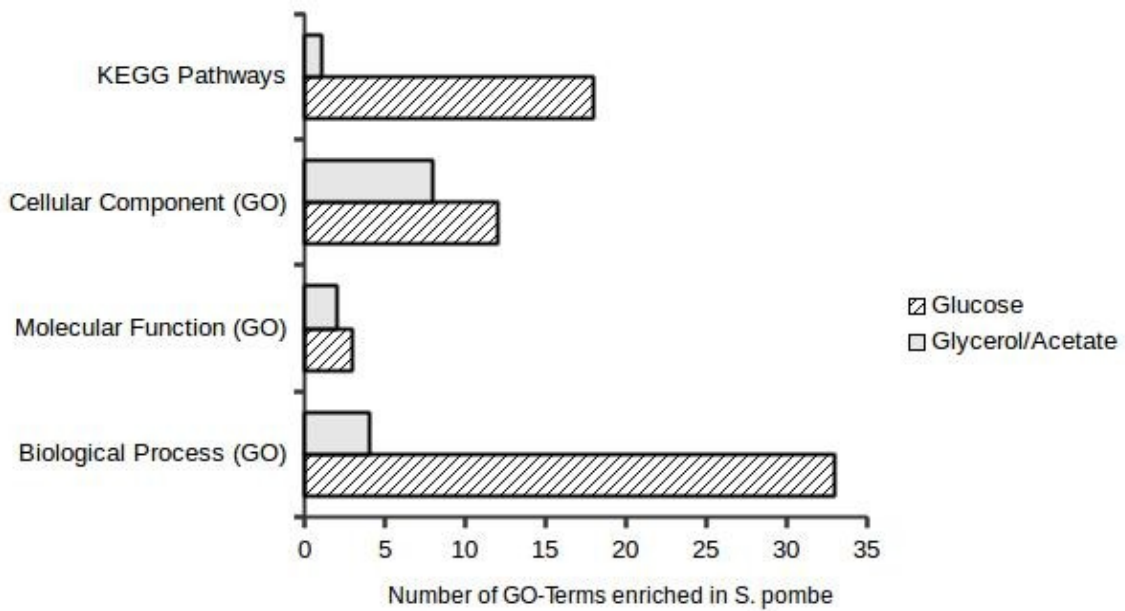


Fig. 14: Comparación de GOs enriquecido funcionalmente en el crecimiento en Glucosa y Glicerol/Actetato en *S.pombe*.

A su vez, con respecto al crecimiento en glicerol; la formación del flagelo se detectó estadísticamente significativo en *E. coli*. Cabe señalar, que esta estructura celular es indicada en los procesos biológicos (GO: 0044781; organización flagelar tipo bacteriano; GO: 0048870; motilidad celular), en los componentes celulares (GO: 0009288, flagelo celular tipo bacteriano), en la función molecular (GO:0003774; actividad motora) y en vías metabólicas (0240, ensamblaje flagelar). Otras vías encontradas fueron la síntesis de metabolitos secundarios con 18 genes y la expresión de sistemas de dos componentes con 15 genes.

Por su parte las categorías funcionales de genes que en *S. pombe* se detectaron en esta condición de crecimiento fueron el transporte de hexosas, el metabolismo de la galactosa, y cambios en los componentes integrales de la membrana. Esta última con 127 genes, entre las que sobresalen proteínas específicas de *S. pombe*, como las proteínas mug.

4.1.8 Análisis de los procesos celulares enriquecidos con la herramienta BINGO.

Con motivo de esclarecer las conexiones entre los procesos biológicos en cada condición de estudio; los grupos de genes que codifican a proteínas que no constituyen factores de transcripción, se analizaron con la herramienta BINGO del programa Cytoscape. Esta herramienta, permite integrar las listas de genes obtenidas, el valor de logFC y FDR de cada gen con la información de redes de interacción biológicas presentes en la base de datos BIOGRID. El análisis detecta la respuesta fisiológica global para cada organismo en cada condición de crecimiento, como se observan en la Figura 15, Figura 16, Figura 17 y Figura

18.

En el caso de *E. coli* cerca de 47 procesos celulares se detectaron como estadísticamente significativo durante el crecimiento en glucosa. Los procesos están relacionados con el metabolismo de carbohidratos, la síntesis de biomoléculas, de aminoácidos, de proteínas y la obtención de energía. En cambio durante el crecimiento en glicerol/acetato se detectaron solamente 17 procesos enriquecidos, relacionados con el transporte y metabolismo de fuentes de carbono como el glicerol y otros poliacoholes.

Igualmente, el análisis en *S. pombe* durante el crecimiento en glucosa, mostró 22 procesos estadísticamente significativos relacionados con la proliferación celular y el incremento de la biomasa. Los procesos mencionados fueron semejantes en categorías a los detectados en *E. coli* en esta condición. Por lo que se refiere al crecimiento en una fuente de carbono gluconeogénica como la mezcla glicerol/acetato se detectaron procesos relacionados con la internalización de hexosas, el uso de galactosa como fuente de carbono y respuestas relacionadas a estrés celular.

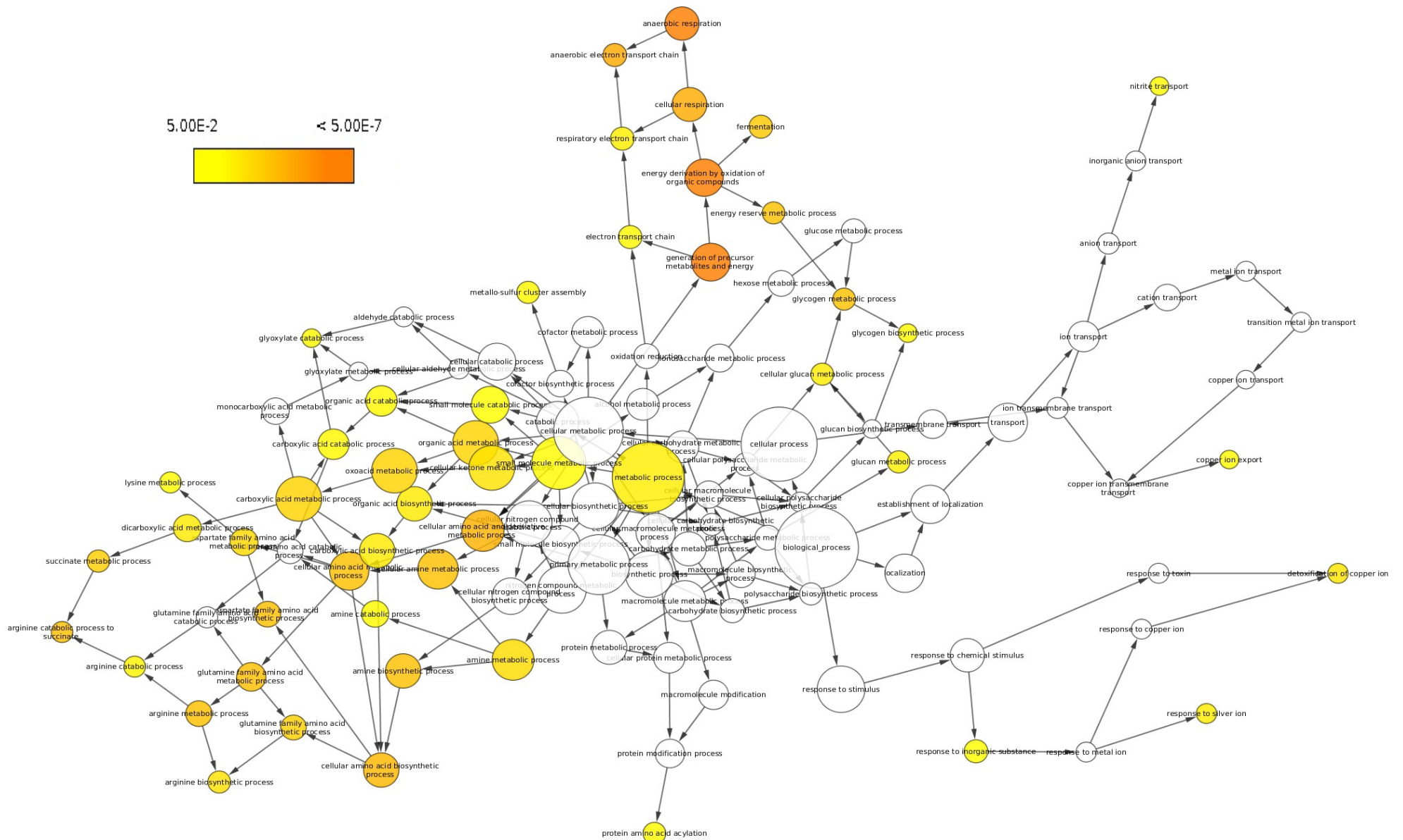


Fig. 15: Procesos biológicos enriquecidos en *E. coli* al crecer en glucosa. El diagrama se realizó empleando la herramienta BINGO de Cytoscape utilizando las redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.

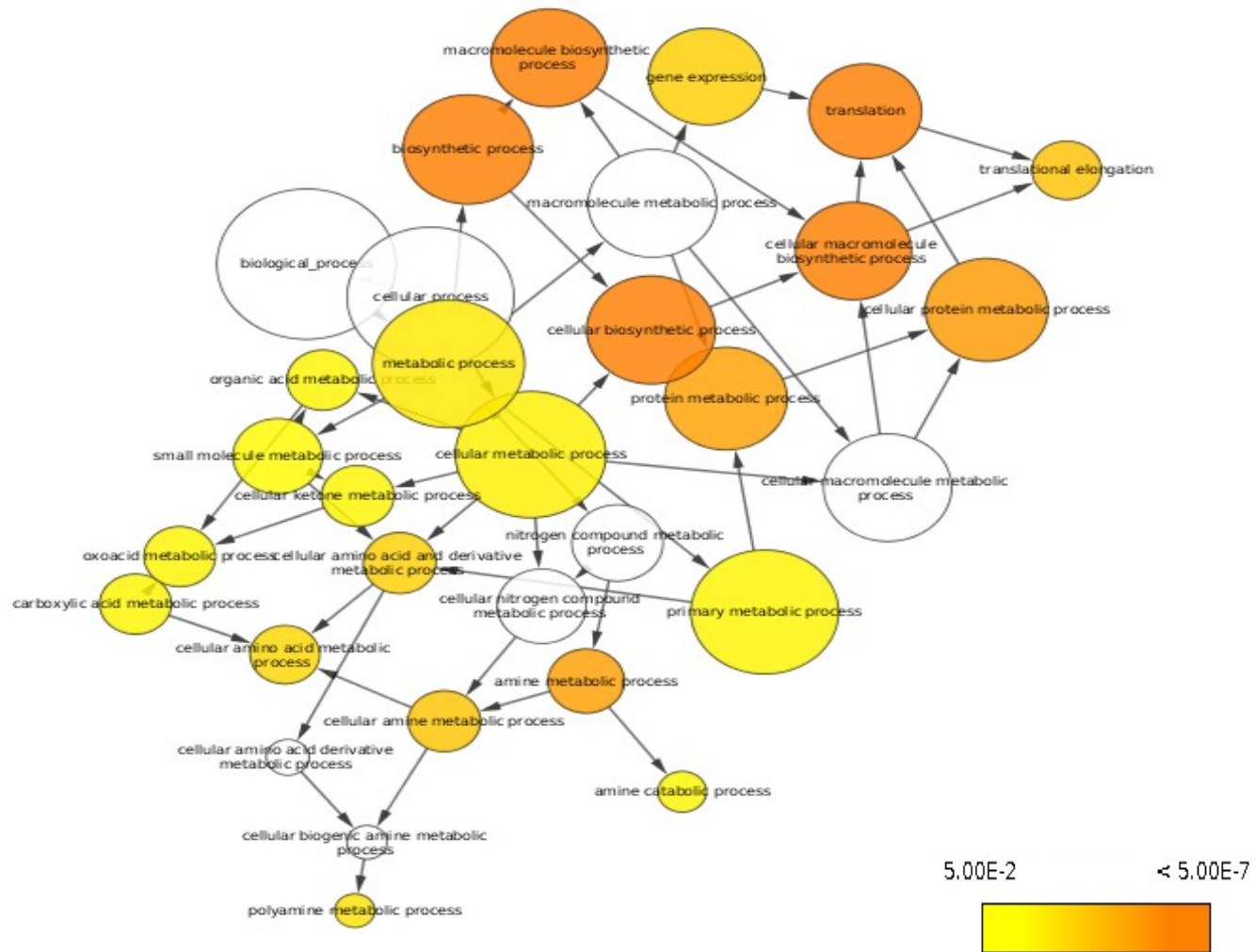


Fig. 17: Procesos biológicos enriquecidos en *S. pombe* al crecer en glucosa. El diagrama se realizó empleando la herramienta BINGO de Cytoscape utilizando las redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.

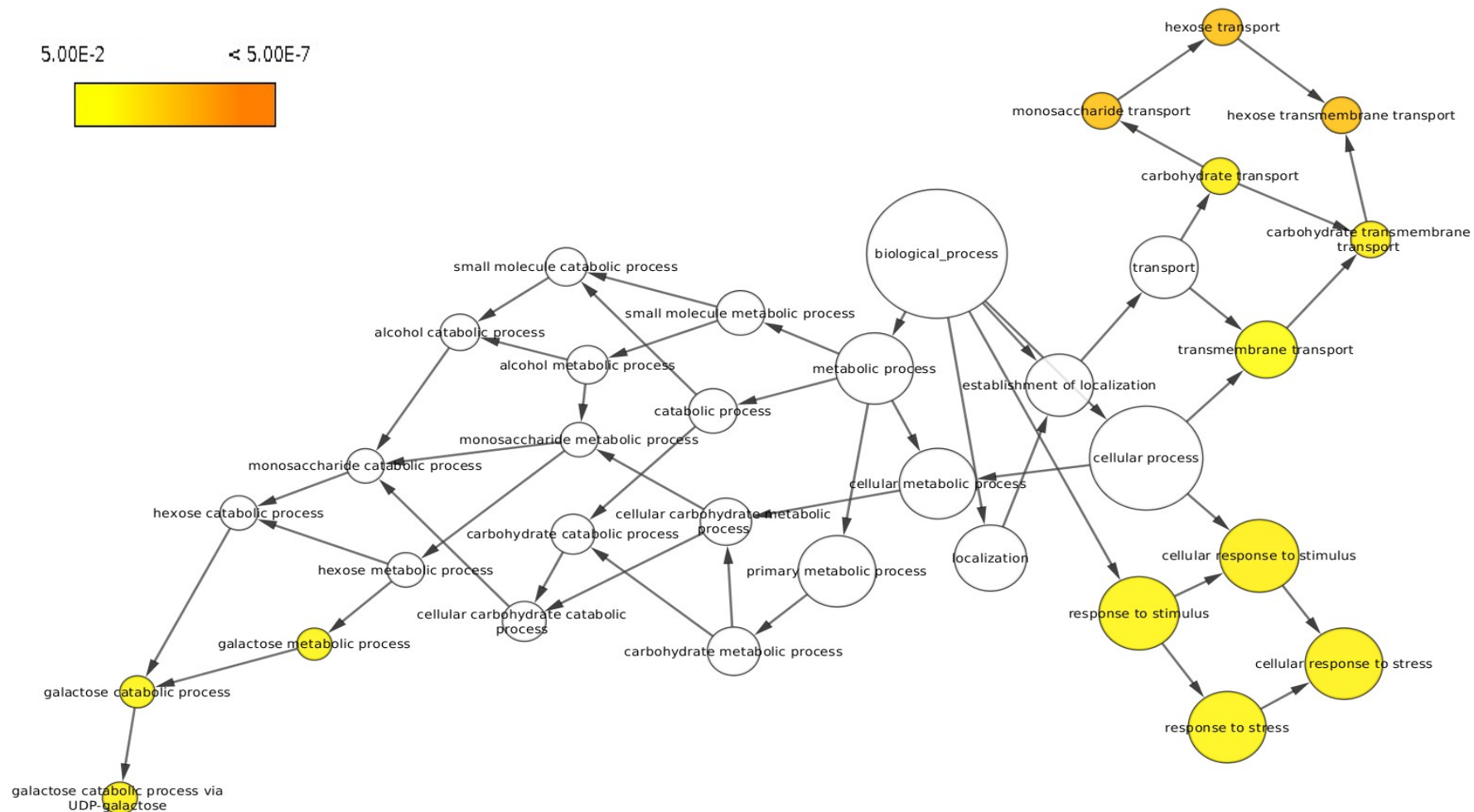


Fig. 18: Procesos biológicos enriquecidos en *S. pombe* al crecer en glicerol/acetato. El diagrama se realizó empleando la herramienta BINGO de Cytoscape utilizando las redes de interacción presentes en BIOGRID y la lista de genes DE. El diámetro de los círculos corresponden a la cantidad de genes presentes en los procesos. El color se corresponde con un valor de $p < 0.05$ en el test hipergeométrico.

4.2 Diferenciación celular inducida por la utilización de fuentes de carbono gluconeogénicas

Teniendo en cuenta que el crecimiento en glicerol/acetato detecto cambios en las proteínas de membrana y formación de estructuras celulares a nivel de estudios de transcriptoma , se realizó un examen microscópico para identificar cambios a nivel morfológicos en los organismos modelos investigados.

En la Figura 19 se muestra la expresión de forma *in silico* de la estructura flagelar mediante la herramienta *KEEG mapper* a partir de los datos de RNA-seq. Para demostrar la presencia de esta estructura flagelar se procesaron las muestras de *E. coli* mediante el método de Leifson y se detecto la formación de estructuras similares a flagelo en el medio adicionado con glicerol/acetato como fuente de carbono (Fig. 20).

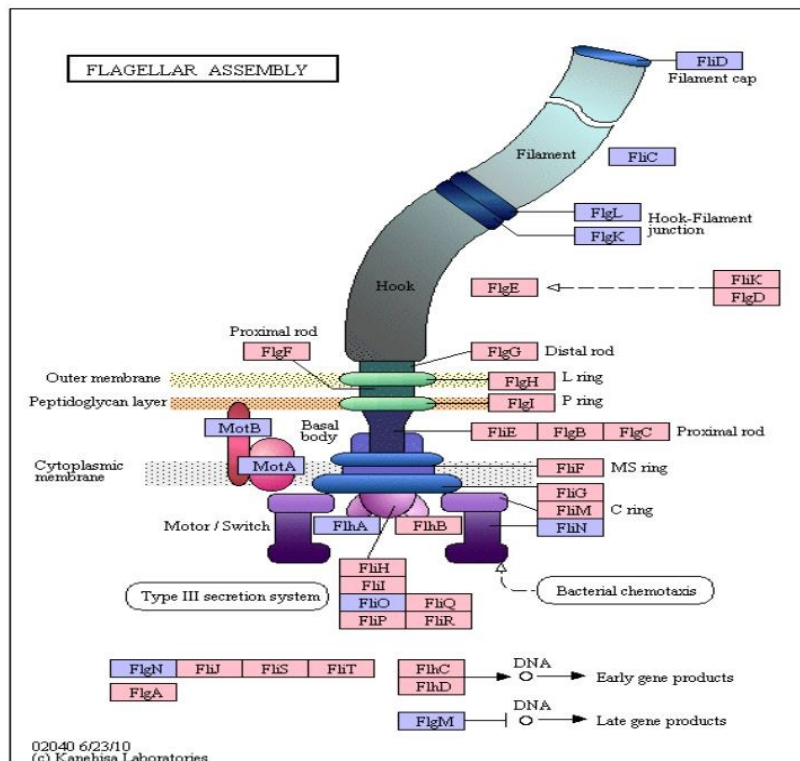


Fig. 19: Empleo de la herramienta KEEG mapper&color para indicar los genes sobreexpresados que están asociados al ensamblaje flagelar. Genes sobre expresados se destacan en color rosa.

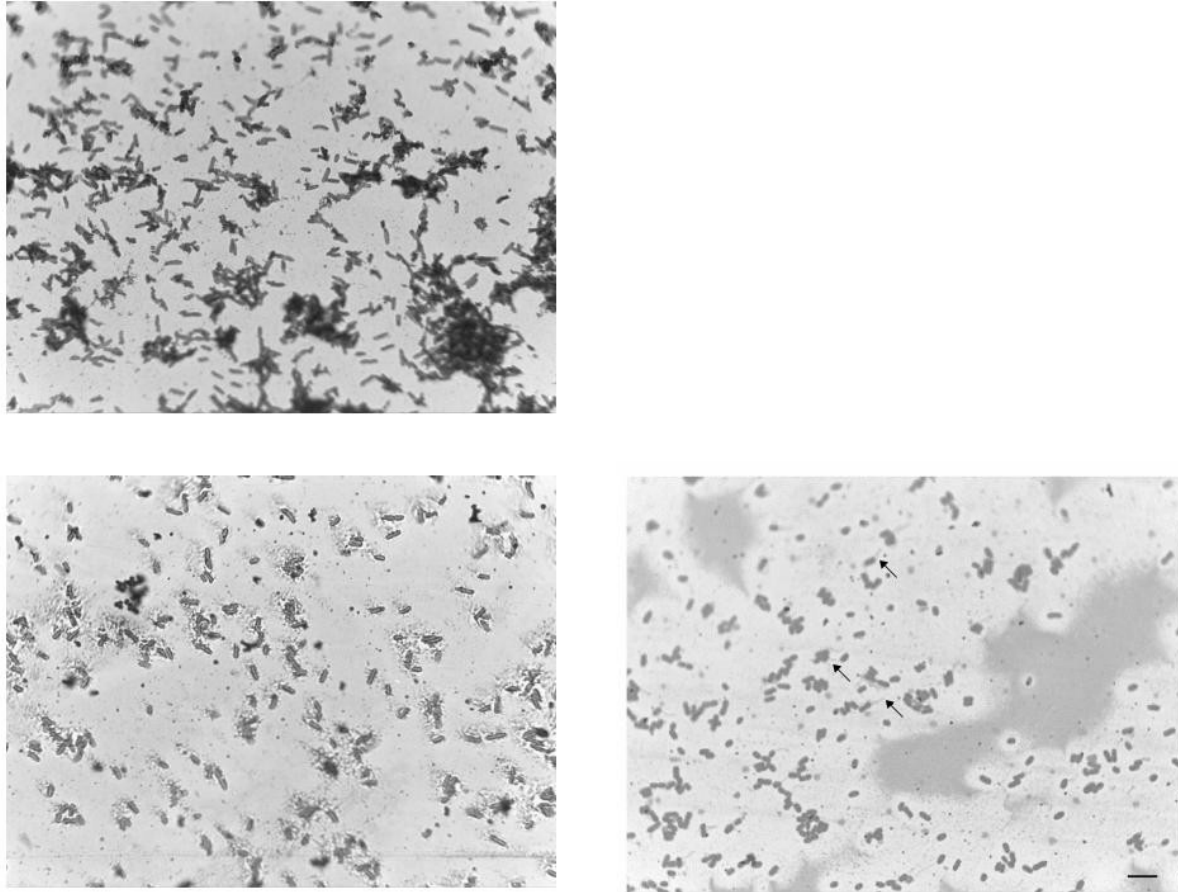


Fig. 20: Tinción de células de *E. coli* mediante el Método de Leifson. Las flechas indican la presencia del Flagelo de forma polar. A Bacterias obtenidas en glucosa. B y C bacterias crecidas en Glicerol-acetato. La barra indica la extensión de 5 micrometros.

En relación a *S. pombe* la observación microscópica refleja diferencias en la morfología celular en el crecimiento en glucosa (forma de bastoncillo con células alargadas) con respecto a el crecimiento en glicerol/acetato. En este último se observó células más pequeñas y redondeadas, Figura 21.

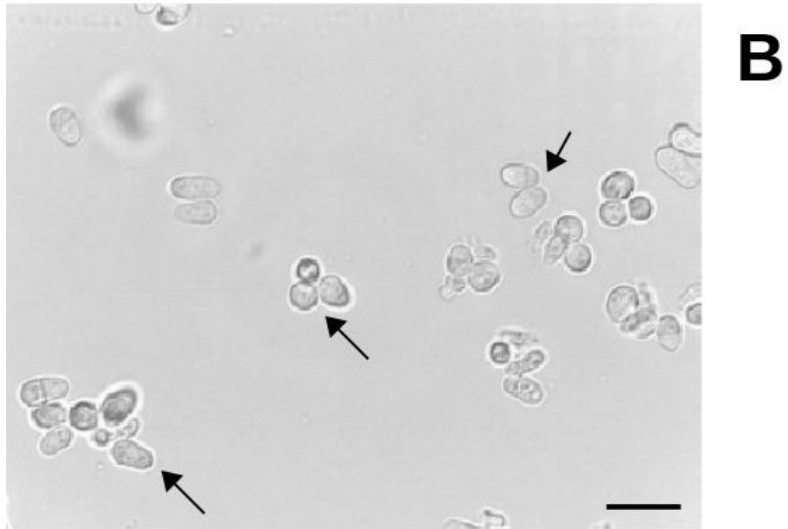


Fig. 21: Diferencias morfológicas en S. pombe al crecer en glucosa (superior) y en glicerol/ acetato (inferior).

4.3 Descripción de la abundancia de factores de transcripción y ncRNAs .

El análisis de la densidad de lecturas o de expresión mostró que los genes que codifican para ncRNAs tuvieron una media de valores de expresión menor que los genes que codifican a proteínas en ambos organismos. Igualmente para ambas condiciones y en ambos organismos los factores de transcripción mostraron un perfil de expresión muy similar con un rango estrecho de valores (Fig 24.).

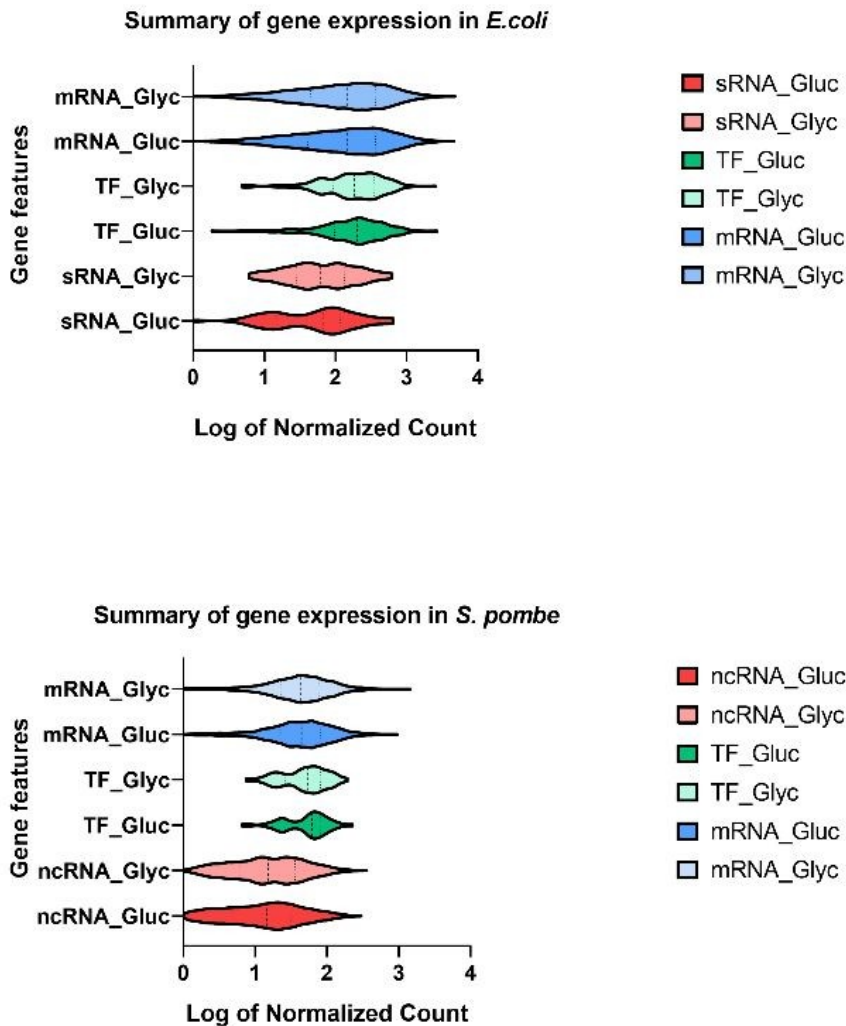


Fig. 22: Gráfico de Violin con los valores de conteo normalizados para gene que codifican a Factores de transcripción (TF) genes que codifican a proteínas no factores de transcripción (mRNA) y genes que codifican a RNAs no codificante (sRNA y ncRNA).

En el presente ensayo genes como *mlc*, *fur*, *arcA*, *crp* y *cra*, que tienen un efecto directo en el metabolismo central del carbono, no mostraron expresión diferencial. Otros reguladores maestros como *fnr*, *soxR*, *soxS*, y *fadR* mostraron un p-FDR ~ 0.05 pero un valor del módulo de logFC < 1.

Similar comportamiento estuvo presente en *spf* (SpoT42) y *sgrs* (*sgrS*), ncRNAs vinculados a la regulación de genes asociados al metabolismo del carbono. Por lo contrario *cyaR*, que constituye un sRNA que se expresa bajo la acción de CRP-cAMP y está sujeto a represión catabólica, se encuentra diferencialmente expresado en el crecimiento glicerol/ acetato.

Uno de los genes que muestra expresión diferencial en el crecimiento en glicerol es el factor transcripcional *fliA* o factor sigma 28, que regula la actividad de los promotores de operones asociados con la motilidad y señalización celular. El gen *fliA* tiene un valor absoluto de logFC = 2,97 expresándose en glicerol/acetato con un comportamiento diferente al resto de los factores sigmas descritos en *E. coli* (Figura 23) .

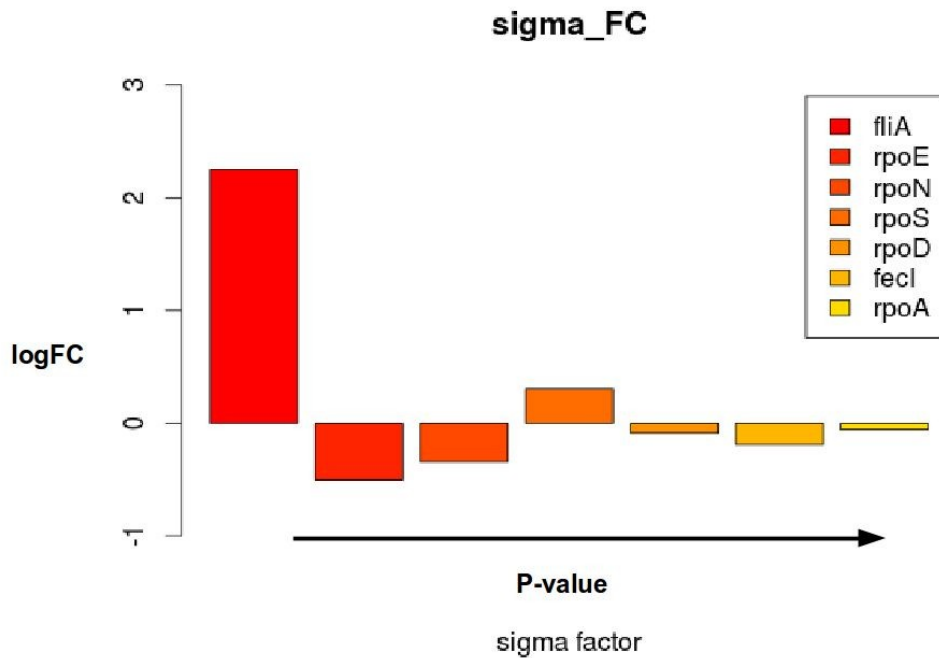


Fig. 23: Índice de log FC de los factores sigma de *Escherichia coli* para las condiciones estudiadas.

En el caso de *S. pombe* la mayoría de ncRNAs no tiene función asignada en la base de datos Pombase, y se infiere que participan en el control de la expresión de genes a través de diferentes mecanismos. En el estudio de genes expresados diferencialmente se detectaron 219 genes de ncRNA expresados marcadamente en el crecimiento en glicerol/acetato en contraste al crecimiento en glucosa donde se expresaron solo 63 genes. Uno de los ncRNAs que detectados es el gen SPNCRNA.1324, que se expresa de forma antisentido al gen *fbp1*, sin embargo valor de expresión con respecto al resto de los 218 ncRNAs sobre expresados, no forma ningún grupo a partir de sus datos de conteo (Figura 24). Al analizar la región de ADN corriente arriba de los ncRNAs pertenecientes al cluster 2 (color ver en la Figura 24), se determino que existían motivos de unión de factores de transcripción asociados a estrés celular (tabla 1).

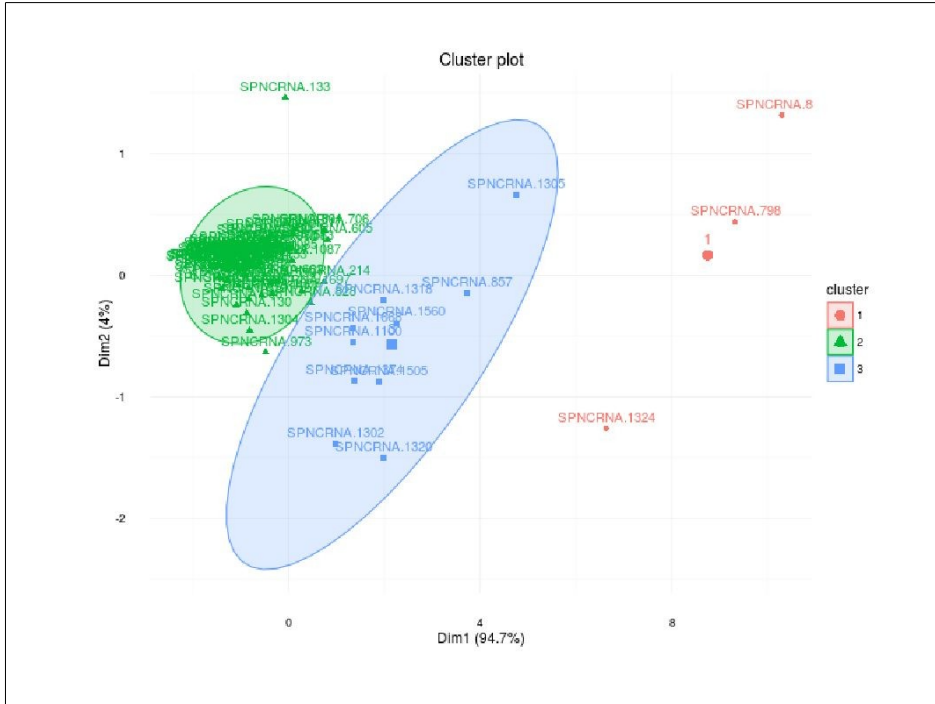


Fig. 24: Agrupamiento mediante el uso de la distancia euclidiana de los ncRNAs de *S. pombe* expresados diferencialmente.

Tabla 1: Análisis de región promotora de ncRNA en *S. pombe*.

name1	name2	logoDP	consensus1	consensus2	match_rank	<i>S. pombe</i>	Description
assembly_4	REI1	10520	...rCAGGGG..	tCAGGGG	1	SPCC550.15c	
assembly_4	MSN4_MSN2	8358AGGGG..	AGGGG	2	SPAC3H1.11	stress factor
assembly_4	RGM1	8265AGGGG..	AGGGG	3	X	
assembly_4	USV1	8478AGGGG..	AGGGG	4	X	
assembly_4	YER130C_MSN2	9233	..ArCAGGGGh.	dAwAGGGGk	5	SPAC3H1.11	stress adaptation
assembly_2	CAD1	3599	..ATTAGTTAkt	ATTAGTAAKC	6	X	
assembly_6	MOT3	5361	..AAGGTA...	hAGGyA	7	X	
assembly_6	ZAP1	5348	swAAGGT....TAAAGGT	8	X	
assembly_6	NRG1	5276	...AGGTACw.	AGGCACA	9	X	
assembly_5	HMS1	3000	...CCACGTAy.	HCACGTMS	10	X	
assembly_4	GIS1_RPH1	9713	...rCAGGGGh.	TwAGGGGt	11	X	
assembly_2	PHO2	2127	tTATTA.....	WTAWTW	12	X	
assembly_2	YAP1	716	tTATTAGTTAkt	MTATTAGTMAKM	13	SPAC1783.07c	oxidative stress tolerance
assembly_5	PHO4	6821	...CCACGTAy.	sCACGTGy	14	X	
assembly_6	RME1	4177	swAAGGTA...	..yAAAGGaA	15	X	
assembly_4	YER130C_MSN4_MSN2	3943	...CAGGGGh.	maGGrrb	16	SPAC3H1.11	
assembly_2	YAP3	6459	...TTAGTTA..	TTAGTAA	17	X	
assembly_4	CUP2_HAA1	7207	...CAGGGG..	smGGsG	18	SPCC584.02	Copper-binding transcription factor
assembly_4	YER130C_MSN2	7054	...rCAGGGGh.	AtAGGGGg	19	SPAC3H1.11	stress adaptation
assembly_4	YER130C_MSN2	1651	.dArCAGGGG..	rdwwagrr	20	SPAC3H1.11	stress adaptation
assembly_5	HAC1	10182	..GCCACGTA..	gMCACGTr	21	X	
assembly_5	CRZ1	531	swGCCAC.....	..MMGCCMC	22	SPAC4G8.13c	activates transcription of stress response genes
assembly_4	USV1	8233	...rCAGGGGh.	yCAGGGGc	23	X	
assembly_2	YHP1_YOX1	7329	...TTAGTTAk.	TTAATTAA	24	SPBC21B10.13c	restricting ECB-mediated transcription to the M/G1 interval
assembly_2	YAP5	509	..ATTAGTTAkt	ATTAKTAATW.	25	X	
assembly_1	SPT10	1930	.hGAACASCAvW	HGATCASCWGA....	26	SPAC8C9.19	sequence-specific activator of histone genes
assembly_5	SKO1	4692ACGTAyk	ACGTAWT.	27	X	
assembly_4	GIS1_RPH1	6853	.dArCAGGGGh.	wwwAGGGGt	28	X	
assembly_4	NHP10	4544	...rCAGGGGh.	GCCGGGGA	29	SPAC10F6.08c	Non-essential INO80 chromatin remodeling complex subunit
assembly_4	GIS1_RPH1	8764	..ArCAGGGGh.	wwwAGGGGw	30	X	
assembly_4	RGM1	8154	...rCAGGGGh.	yCAGGGGc	31	X	

CAPÍTULO V : Discusión General

5.1 Análisis Global del Transcriptoma de *E. coli* y *S. pombe* en dos escenarios metabólicos.

En la obtención de datos de RNA-seq, la condición ideal es que el material biológico (RNA) provenga de una de una sola célula. Una alternativa de esta condición es precisamente lograr que la mayoría de unidades en un cultivo sean detenidas en una misma etapa del ciclo celular con el fin de disminuir la heterogeneidad en la muestra biológica. El crecimiento típico de un cultivo microbiano o incluso de líneas de celulares, muestra cinco fases distintas de crecimiento: fase lag, o fase de adaptación que antecede a la fase exponencial; fase exponencial o fase log, donde la división celular presenta una velocidad constante de crecimiento; fase estacionaria donde las condiciones del medio dejan de ser óptimas y la fase de muerte celular, donde la mayoría de la células pierden la viabilidad.

La principal diferencia entre las fases de un cultivo es precisamente la velocidad de crecimiento, parámetro que está relacionado con el tiempo de duplicación de las células. Este parámetro también muestra diferencias en un cultivo no solo a través del tiempo sino también en el espacio considerando el mismo como el entorno en el que crece la célula. Las diferentes condiciones de un entorno celular determinan su expresión específica de genes en esta condición y consecuentemente su tiempo de duplicación.

La velocidad de crecimiento toma valores positivos y aproximadamente constantes durante la fase exponencial ³⁷(Kjeldgaard, Maaloe et al. 1958) con lo que esta relacionado a la duplicación celular y a una población de células homogéneas.

En el campo de la fisiología microbiana se ha establecido desde 1958 que la composición macromolecular de una célula cambia con la velocidad de crecimiento ³⁷(Kjeldgaard, Maaloe et al. 1958). Las células que crecen más rápido contienen proporcionalmente más rRNAs, tRNAs y mRNAs más estables. La razón de esta abundancia de RNAs estables es simple: en el orden de crecer más rápido las células deben sintetizar proteínas más rápido. Se conoce que siete operones de rRNA de *Escherichia coli* están bajo el control de promotores que dependen de la velocidad de crecimiento y que la expresión de proteínas ribosomales, factores de traducción y el aparato de transcripción están ligados a la concentración de rRNA (Dennis, Ehrenberg et al. 2004).

Nuevos estudios han demostrados que la partición (estado libre o unida al DNA) de la RNA polimerasa también se encuentra asociada a la velocidad de crecimiento ³⁸(Klumpp and Hwa 2008). Sobre esta perspectiva, Zaslaver y colaboradores en 2009, detectaron que existen al menos dos clases distintas de promotores que muestran un comportamiento diferente entre condiciones de crecimiento: los promotores ribosomales y los promotores metabólicos. La clase de promotores ribosomales está correlacionada con las condiciones de crecimiento, mientras que los promotores de proteínas metabólicas (enzimas, proteínas de transporte y factores de regulación) se expresan en niveles bajos-intermedios y varían entre diferentes condiciones de crecimiento, con marcada expresión diferencial durante la fase exponencial y

no así durante la fase estacionaria ³⁹(Zaslaver, Kaplan et al. 2009).

La realización de un análisis comparativo global del transcriptoma entre las especie *E.coli* y *S. pombe* permite obtener una descripción de que respuestas fisiológicas han cambiado y cuáles se conservan entre un organismos procarionte y un organismos eucarionte. En ambos organismos el análisis del transcriptoma y los genes expresados indican una conservación en la respuesta metabólica e incluso en la respuesta transcripcional.

Tanto en *E. coli* como en *S. pombe* el crecimiento en glucosa favorece la proliferación celular o formación de biomasa, y en una fuente gluconeogénica, como el glicerol, se encendieron genes típicos del metabolismo respiratorio como la enzima gliceraldehído fosfato deshidrogenasa. Estos puntos comunes hace plantear que la respuesta a un mismo estímulo (condición de crecimiento) en especies separadas evolutivamente pudiese ser un patrón conservado donde la respuesta transcripcional juega un papel central. Tal es el caso de la expresión de las enzimas vinculadas a la glicólisis y la gluconeogénesis en ambos organismos (Figura 25) .

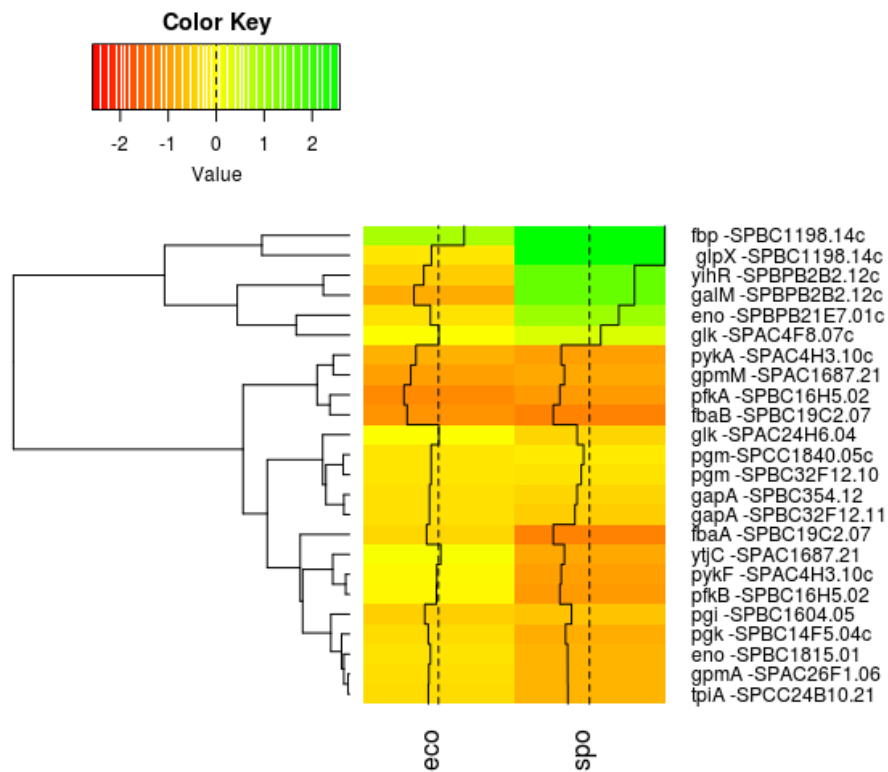


Fig. 25: Correspondencia entre genes análogos relacionados con el procesamiento de la glucosa a piruvato.

En ambos casos, se observan cambios en la estructura de la célula ante la utilización del glicerol/acetato como fuente de carbono y energía. Cabe mencionar que en *E. coli* no se detectaron sobre expresados ninguno de los genes descritos a procesos de restricción nutricional, a pesar de que Zhang y colaboradores describió que la respuestas a fuentes de carbono gluconeogénicas como el glicerol , el succinato , y la glicina se corresponden a una respuesta de “patrones de forrajeo” del inglés *foreaging behavior*. Esta conducta es descrita en ecología para animales superiores y describe todas los métodos por los cuales un organismo busca, adquiere y utiliza fuentes de nutrientes y energía.

De manera similar en *S. pombe* tampoco se observó una respuesta de restricción nutricional, no se detectaron genes relacionados con la formación del autofagosoma; complejo molecular que se encuentra asociado a escenarios de restricción calórica y déficit energético. Por otro lado se detectaron categorías funcionales de genes asociados a estrés celular y genes relacionados con reconocimiento y señalización a nivel de membrana, como es el caso de la familia de proteínas *mug*.

Estas proteínas preparan a *S. pombe* para la formación del diploide y la fase de meiosis. La meiosis si bien constituye un evento de reproducción sexual se induce en medios con déficit de nitrógeno. Esta descrito en la literatura que la formación del diploide y la consecuente esporulación, tanto en *S. cerevisiae* como en *S. pombe* permite la reparación del ADN, eventos de recombinación homóloga, reestructuración del citoesqueleto y ventajas adaptativas de sobrevivencia. Por lo que podemos relacionarlo como una diferenciación celular o respuesta tipo *foreaging behavior* que garantiza la utilización de fuentes alternativas de nutrientes y energía, de forma análoga a la respuesta observada en *Escherichia coli*..

Nuestro análisis nos permite inferir como la respuesta metabólica a partir del análisis del transcriptoma y como la misma es conservada en dos organismos perteneciente a dominios de la vida diferentes. También nos indica que la glucosa es una fuente de energía de preferencia para la célula y que su presencia en el medio determina una respuesta transcripcional conservada en la evolución.

5.2 Abundancia de ncRNAs en fuentes de carbono no glicolíticas.

Un aspecto interesante con respecto al tipo de respuesta observada muestra que en el caso del crecimiento de glicerol existe una mayor detección de transcritos del tipo ncRNAs en ambos organismos.

Los RNAs regulatorios controlan la expresión de genes a nivel post-transcripcional y transcripcional. En *E. coli* han sido descritos más de 170 y algunos de los encontrados en el presente estudio han sido reportados y validados en literatura en respuesta a estrés celular como *ryhB* y *micF*. Los sRNAs mencionados anteriormente regulan la expresión ARN mensajeros de proteínas vinculadas al estrés oxidativo y estrés en la envoltura de membrana. Sin embargo, el control postranscripcional de genes es una respuesta que involucra a múltiples factores (Richards y Vanderpoll 2011).

El sRNAs con menor FDR en nuestro estudio fue *cyaR*, con un valor de cambio de 2.68. El transcrito *cyaR* es estabilizado por Hfq, y tiene genes blancos como *ompx*, *yqaE*, *nadE*, y *luxS*. Este sRNAs esta sujeto a represión catabólica y su expresión depende del efecto combinatorio del factor sigma E y el sistema de dos componentes CpxAR cuales están asociados a estrés de membrana.

Desde el punto de vista de reguladores maestros el gen *fliA* o sigma 28 tiene una marcada expresión en el crecimiento en glicerol/acetato. Es necesario destacar que *fliA* regula genes relacionados con la expresión del flagelo y otros unidades transcripcionales como el regulador maestro *flhDC* y el operón *glpABD*, que permite el metabolismo del glicerol. Además *fliA* al igual que sigma E compite por los sitios de unión de sigma 70 en la respuesta transcripcional

y controla en combinatoria con genes de tipo sRNA los genes corriente abajo a su regulación.

En el caso de *S. pombe* se han anotado más de 1500 ncRNA en el genoma y Atkinson y col. 2018 reportó cerca de 5000 nuevos ncRNAs. Adicionalmente uno de los ncRNAs asociados a la respuesta a estrés es SPNCRNA.1164 que se detectó con un valor de $\log_{2}FC$ de 0.7003 y un valor de probabilidad de FDR de 0.0237 este ncRNA se especula que actúa en *trans*. Curiosamente el gen blanco de SPNCRNA.1164 es *atf1*, factor transcripcional que constituye un regulador maestro en la respuesta asociada al estrés celular en *S. cerevisiae*, en humanos y en *S. pombe*. Sin embargo no es detectado como sobre expresado. No obstante a este hecho, el estudio de *clusterización* de ncRNA sobre expresados en glicerol mostró que cierto grupo presenta motivos de unión a factores de transcripción relacionados con estrés celular. Por lo que podemos suponer que al igual que en *E. coli* la expresión de ncRNAs en fuentes no glicolíticas está relacionada estrechamente con condiciones de estrés y / o diferenciación celular.

CAPÍTULO VI : CONCLUSIONES GENERALES Y PERSPECTIVA

6.1 Conclusiones generales

- El análisis del transcriptoma de *E. coli* y *S. pombe* obtenido durante el crecimiento en glucosa comparado con el crecimiento en glicerol/acetato permitió determinar los genes expresados diferencialmente y los procesos biológicos enriquecidos en cada organismo modelo, mediante el empleo de herramientas bioinformáticas. Hecho que brindó la posibilidad de inferir que la respuesta fisiológica de ambos organismos es análoga; proliferación celular al crecer bajo glicólisis aeróbica y estrés y/o diferenciación celular al crecer mantenido en metabolismo respiratorio.
- Los factores de transcripción tuvieron un logcpm menor que los ncRNAs tanto en *E. coli* como en *S. pombe*; lo que está relacionado con el rol que cumplen en la regulación de genes. Adicionalmente, la cantidad de transcritos correspondientes a ncRNAs expresados diferencialmente y la abundancia de los mismos fue mayor en el crecimiento de glicerol/acetato que en el crecimiento en glucosa lo cual sugiere que la respuesta del ncRNAoma está sujeta al control del estrés y a la represión catabólica al crecer en ausencia de glucosa.
- No se detectaron analogías en los circuitos de regulación mediados por ncRNAs sobre genes ortólogos en rutas metabólicas asociadas al metabolismo central del carbono. Nuestro análisis indicó que la respuesta fisiológica está estrechamente vinculada con

el control transcripcional que con un nivel posttranscripcional al menos en genes cuyo producto son enzimas del metabolismo primario.

6.2 Perspectivas

Debido a que este es el primer estudio para la descripción y/o exploración de circuitos regulatorios conservados desde la perspectiva del análisis del transcriptoma en un organismo procarionte y un organismo eucarionte, surgen varias preguntas científicas o líneas de investigación para profundizar en la caracterización de este tópico:

1. Establecer si al igual que en *E. coli* en *S. pombe* existen genes de tipo ncRNA sujetos a represión catabólica.
2. Validar genes ortólogos expresados diferencialmente en ambos organismos a través de RT-qPCR.
3. Ampliar la búsqueda de ncRFs y fragmentos pequeños parecidos a microRNAs partir de diferentes herramientas bioinformáticas .
4. Realizar experimentos de tipo RIP-seq o CLIP-seq para poder determinar la unión de estos transcritos a mensajeros y/o proteínas específicas.

Referencias Bibliográficas

1. Costa, V., Aprile, M., Esposito, R. & Ciccodicola, A. RNA-Seq and human complex diseases: Recent accomplishments and future perspectives. *Eur. J. Hum. Genet.* **21**, 134–142 (2013).
2. Fendt, S. M. *et al.* Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Mol. Syst. Biol.* **6**, 1–11 (2010).
3. Paliy, O. & Gunasekera, T. S. Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl. Microbiol. Biotechnol.* **73**, 1169–1172 (2007).
4. Hawkins, P. G. & Morris, K. V. RNA and transcriptional modulation of gene expression. *Cell Cycle* **7**, 602–607 (2008).
5. Beisel, C. L. & Storz, G. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.* **34**, 866–882 (2010).
6. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15 Spec No**, 17–29 (2006).
7. Watanabe, Y. & Kanai, A. Systems biology reveals microRNA-mediated gene regulation. *Front. Genet.* **2**, 1–13 (2011).
8. Lee, R. C. Lee-Et-Al-Microrna-1993. **75**, 843–854 (1993).
9. Delihias, N. Regulation of gene expression by trans-encoded antisense RNAs. *Mol. Microbiol.* **15**, 411–414 (1995).
10. Storz, G., Altuvia, S. & Wassarman, K. M. AN ABUNDANCE OF RNA REGULATORS. *Annu. Rev. Biochem.* **74**, 199–217 (2005).

11. De La Fuente, M., Valera, S. & Martínez-Guitarte, J. L. NcRNAs and thermoregulation: A view in prokaryotes and eukaryotes. *FEBS Lett.* **586**, 4061–4069 (2012).
12. Alkan, C. *et al.* RNA Secondary Structure Prediction Via Energy Density Minimization. in *Research in Computational Molecular Biology* (eds. Apostolico, A., Guerra, C., Istrail, S., Pevzner, P. A. & Waterman, M.) 130–142 (Springer Berlin Heidelberg, 2006).
13. Bleda, M. *et al.* Inferring the regulatory network behind a gene expression experiment. *Nucleic Acids Res.* **40**, 168–172 (2012).
14. Georg, J. & Hess, W. R. cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 286–300 (2011).
15. Peters, J. M. *et al.* Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 15406–11 (2009).
16. Stazic, D., Lindell, D. & Steglich, C. Antisense RNA protects mRNA from RNase e degradation by RNA-RNA duplex formation during phage infection. *Nucleic Acids Res.* **39**, 4890–4899 (2011).
17. Sorek, R. & Cossart, P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* **11**, 9 (2009).
18. Okamura, K. Diversity of animal small RNA pathways and their biological utility. *Wiley Interdiscip. Rev. RNA* **3**, 351–368 (2012).
19. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
20. Moll, I., Leitsch, D., Steinhauser, T. & Bläsi, U. RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep.* **4**, 284–289 (2003).

21. Malphettes, L. & Fussenegger, M. Impact of RNA interference on gene networks. *Metab. Eng.* **8**, 672–683 (2006).
22. Rácz, Z., Kaucsár, T. & Hamar, P. The huge world of small RNAs: Regulating networks of microRNAs (Review). *Acta Physiol. Hung.* **98**, 243–251 (2011).
23. Lin, C. C., Mitra, R. & Zhao, Z. A tri-component conservation strategy reveals highly confident microRNA-mRNA interactions and evolution of microRNA regulatory networks. *PLoS One* **9**, (2014).
24. Lee, J. G. and N. H. Computational and Experimental Approaches for Modeling Gene Regulatory Networks. *Current Pharmaceutical Design* **13**, 1415–1436 (2007).
25. Oszolak, F. & Milos, P. M. Transcriptome Profiling Using Single-Molecule Direct RNA Sequencing BT - High-Throughput Next Generation Sequencing: Methods and Applications. in (eds. Kwon, Y. M. & Ricke, S. C.) 51–61 (Humana Press, 2011).
doi:10.1007/978-1-61779-089-8_4
26. Brazhnik, P. & Tyson, J. J. Cell Cycle Control in Bacteria and Yeast. *Cell Cycle* 522–529 (2006).
27. Kelley, B. P. *et al.* Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci.* **100**, 11394–11399 (2003).
28. Ortega-Cuellar, D. *et al.* Biotin starvation with adequate glucose provision causes paradoxical changes in fuel metabolism gene expression similar in rat (*Rattus norvegicus*), nematode (*Caenorhabditis elegans*) and yeast (*Saccharomyces cerevisiae*). *J. Nutrigenet. Nutrigenomics* **3**, 18–30 (2010).
29. Velazquez-Arellano, A., Ortega-Cuellar, D., Hernandez-Mendoza, A. & Moreno-Arriola,

- E. A heuristic model for paradoxical effects of biotin starvation on carbon metabolism genes in the presence of abundant glucose. *Mol. Genet. Metab.* **102**, 69–77 (2011).
30. Skinner, C. & Lin, S. J. Effects of calorie restriction on life span of microorganisms. *Appl. Microbiol. Biotechnol.* **88**, 817–828 (2010).
 31. Martínez-Gómez, K. *et al.* New insights into Escherichia coli metabolism: Carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microb. Cell Fact.* **11**, 1–21 (2012).
 32. Zhao, K., Liu, M. & Burgess, R. R. Adaptation in bacterial flagellar and motility systems: From regulon members to 'foraging'-like behavior in E. coli. *Nucleic Acids Res.* **35**, 4441–4452 (2007).
 33. An, J., Lai, J., Lehman, M. L. & Nelson, C. C. MiRDeep*: An integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res.* **41**, 727–737 (2013).
 34. Thomas, L. F., Saito, T. & Sætrom, P. Inferring causative variants in microRNA target sites. *Nucleic Acids Res.* **39**, (2011).
 35. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
 36. Kozomara, A. & Griffiths-Jones, S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, 68–73 (2014).
 37. Kjeldgaard, N. O., MaalOe, O. & Schaechter, M. The Transition Between Different Physiological States During Balanced Growth of Salmonella typhimurium. *J. Gen.*

- Microbiol.* **19**, 607–616 (1958).
38. Klumpp, S. & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20245–20250 (2008).
 39. Zaslaver, A. *et al.* Invariant distribution of promoter activities in Escherichia coli. *PLoS Comput. Biol.* **5**, (2009).
 40. The Mendeley Support Team. Getting Started with Mendeley. *Mendeley Desktop* 1–16 (2011).
 41. Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M. & Vogel, J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).
 42. Langmead, B. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma.* 1–14 (2010). doi:10.1002/0471250953.bi1107s32
 43. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562 (2012).
 44. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 45. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
 46. Neidhardt, F. C. GUEST COMMENTARY Bacterial Growth: Constant Obsession with dN/dt. *J. Bacteriol.* **181**, 7405–7408 (1999).

47. Bochner, B. R., Gadzinski, P. & Panomitros, E. Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* **11**, 1246–1255 (2001).
48. Davidson, C. J., Narang, A. & Surette, M. G. Integration of transcriptional inputs at promoters of the arabinose catabolic pathway. *BMC Syst. Biol.* **4**, (2010).
49. Richards, G. R. & Vanderpool, C. K. Molecular call and response: The physiology of bacterial small RNAs. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1809**, 525–531 (2011).
50. Hardiman, T., Lemuth, K., Keller, M. A., Reuss, M. & Siemann-Herzberg, M. Topology of the global regulatory network of carbon limitation in Escherichia coli. *J. Biotechnol.* **132**, 359–374 (2007).
51. Lemuth, K. *et al.* Global transcription and metabolic flux analysis of Escherichia coli in glucose-limited fed-batch cultivations. *Appl. Environ. Microbiol.* **74**, 7002–7015 (2008).
52. Moreno, S., Klar, A. & Nurse, P. B. T.-M. in E. [56] Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. in *Guide to Yeast Genetics and Molecular Biology* **194**, 795–823 (Academic Press, 1991).
53. Wilhelm, B. T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243 (2008).
54. Wilhelm, B. T., Marguerat, S., Goodhead, I. & Bähler, J. Defining transcribed regions using RNA-seq. *Nat. Protoc.* **5**, 255–66 (2010).
55. Rhind, N. *et al.* Comparative functional genomics of the fission yeasts. *Science (80-.)*. **332**, 930–936 (2011).

56. Lackner, D. H., Schmidt, M. W., Wu, S., Wolf, D. A. & Bähler, J. Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome Biol.* **13**, (2012).
57. Forsburg, S. L. & Rhind, N. Basic methods for fission yeast. *Yeast* **23**, 173–183 (2006).
58. Suzuki, S., Matsuzawa, T., Nukigi, Y., Takegawa, K. & Tanaka, N. Characterization of two different types of UDP-glucose/-galactose 4-epimerase involved in galactosylation in fission yeast. *Microbiology* **156**, 708–718 (2010).
59. Klement, T., Dankmeyer, L., Hommes, R., van Solingen, P. & Büchs, J. Acetate-glycerol cometabolism: Cultivating *Schizosaccharomyces pombe* on a non-fermentable carbon source in a defined minimal medium. *J. Biosci. Bioeng.* **112**, 20–25 (2011).
60. Klein, T., Heinzle, E. & Schneider, K. Metabolic fluxes in *Schizosaccharomyces pombe* grown on glucose and mixtures of glycerol and acetate. *Appl. Microbiol. Biotechnol.* **97**, 5013–5026 (2013).
61. Saitoh, S. & Yanagida, M. Does a shift to limited glucose activate checkpoint control in fission yeast? *FEBS Lett.* **588**, 2373–2378 (2014).

Anexos

Anexo 1 : Artículo en proceso de Revisión.

Journal of Molecular Microbiology and Biotechnology

Manuscript:	MMB-2019-11-1
Title:	Analog circuits of transcriptional regulation of Escherichia coli and Schizosaccharomyces pombe in respiratory and fermentative growth inferred by comparative analysis of global transcriptional profiles
Authors(s):	Joivier Vichi (Co-author), Emmanuel Salazar (Co-author), Veronica Jimenez (Co-author), Leticia Olvera (Co-author), Ricardo Alfredo Cano (Co-author), Enrique Enrique Morett (Co-author), Edgar Dantán González (Co-author), Armando Hernandez-Mendoza (Corresponding Author)
Keywords:	Escherichia coli, Evolution, Gene regulation, High throughput, Yeast
Type:	Research Article

Anexo 2 : Artículo en etapa de Redacción

Análisis de los resultados obtenidos de las librerías de ncRNAs pequeños en *S. pombe*

Comunicación Corta

Putative Title : **Exploration of the non-coding RNAome of *Schizosaccharomyces pombe* from mRNA and small RNA expression profiling analysis**

Authors: Vichi-Lozada J.1, Salazar-Bustamante E.1, Grande-Cano R.A.2, Olvera-Rodríguez L.2, Jiménez-Jacinto V.2, Dantán-González E.1, Morett-Sánchez E.2, and Hernández-Mendoza A.1*

Los genes que codifican proteínas constituyen una fracción muy pequeña del genoma en comparación con sus contrapartes no codificantes, a las que se hace referencia en la literatura como ncRNA. Estudios recientes estiman un total de 15.000 moléculas de ARN no codificantes (incRNAs) en humanos, sin embargo, todavía es difícil cuantificar con precisión el número de incRNAs y su papel en la fisiología celular. Los IncRNAs se definen por la ausencia de un marco de lectura (ORF), y presentan una longitud superior a 200 nt. Estas transcripciones están generalmente localizadas en el núcleo, presentan cola de poliA y son producidas por la ARN polimerasa II. De acuerdo a su ubicación con respecto a los genes que codifican las proteínas, los IncRNAs se clasifican como intragenicos o intrónicos, intergenicos y antisentido (cuando se superponen con uno o más exones). El descubrimiento de IncRNA en unicelulares como *Saccharomyces cerevisiae* y *Schizosaccharomyces pombe* dio lugar a un universo increíble para la exploración de nuevos mecanismos de regulación

mediados por el RNA. Los esfuerzos iniciales para caracterizar estas moléculas mostraron que funcionan en cis, como el gen XIST y AIR, que secuestran el complejo de modificación de la cromatina para silenciar los sitios adyacentes, y la regulación trans como HOTAIR, otras hipótesis apoyan que juegan como precursores de pequeños ncRNAs como los micrRNAs y siRNAs.

RNA-seq mostró que los lncRNAs se transcriben con cierta relación a las regiones codificantes, especialmente los lncRNAs antisentido y de acuerdo a los patrones de expresión varios lncRNAs sugieren una funcionalidad a nivel post-transcripcional. Algunos lncRNAs se superponen a otros pequeños RNAs y se cree que son precursores de estos pequeños RNAs reguladores. En este contexto, es de gran interés investigar, identificar y extraer información sobre nuevas familias posibles de pequeñas familias de reguladores de ARN diferentes a los microARNs y siRNAs en la levadura de fisión. Actualmente los costos de secuencia permiten obtener una mayor cobertura del genoma más simple y revelar todo su repertorio de transcripción. Los organismos modelo unicelulares como la levadura, han sido fundamentales para revelar la naturaleza del ncRNA largo y pequeño en ciertos escenarios metabólicos, por ejemplo en la levadura de fisión se ha mostrado la regulación postranscripcional mediada por el ncRNA y las asociaciones con factores de transcripción como el ATF1, sin embargo su función subyacente es la de la mayoría de los ncRNAs anotados en el genoma.

Para obtener más información sobre el transcriptoma no codificante de la levadura de fisión, interrogamos al transcriptoma de la levadura de fisión en dos condiciones de crecimiento aplicando tecnología de secuenciación profunda para ARN largo y pequeño. Este método ofrece un poderoso medio para identificar nuevos ARNs reguladores. Se determinó la

expresión diferencial de genes de ncRNA en fuentes de carbono glicolíticas y gluconeogénicas y el uso de herramientas computacionales para distinguir de forma fiable los productos de degradación de ARN, identificando pequeños ARN con bajas tasas de expresión, o para definir a lncRNA como precursores de pequeños ncRNA.

Métodos

La cepa utilizada en este trabajo fue *S. pombe* 972 h- (donada por la Dra. Elena Hidalgo, Universitat Pompeu Fabra, Barcelona, España). YE-Glu se refiere al medio de extracto de levadura, fue hecho del extracto de levadura 5 g.L⁻¹ (BD, Difco) suplementado con KH₂PO₄ 200 mg.L⁻¹ y glucosa 2% (p/v) para la fuente de carbono. El mismo medio con glicerol al 2 % y acetato de sodio al 0,2 % como fuente de carbono se denominó YE-Gly. Todos los cultivos (cultivos por lotes de 250 ml de matraz Erlenmeyer con 100 ml de medio) se incubaron a 30 °C en un agitador rotatorio de incubadora a 250 rpm en la fase inicial de crecimiento exponencial (DO 0,5). En el presente estudio, las réplicas son cultivos cultivados a partir de colonias individuales independientes. Se realizaron cinco réplicas independientes para cada tratamiento.

El ARN Total fue extraído por el reactivo TRI (Zymo Research), seguido de la precipitación de isopropanol sin filtración para retener todos los ARN pequeños. Las muestras fueron tratadas con DNAasa I (Thermo) y el ARN ribosomal fue removido con RiboMinus™ Levadura (Life Technologies), la cual fue usada de acuerdo al protocolo recomendado por el fabricante. Las muestras se volvieron a concentrar con el kit RNA Clean Concentrator™ (Zymo Research) y se evaluó la integridad del ARN utilizando un kit Agilent Bioanalyzer 2100 y RNA Nano 6000 Labchip Kit (Agilent Technologies). Las librerías de secuenciación enriquecidas con ARNm fueron preparadas usando el kit de preparación de ARNm trenzado Illumina

TruSeq según las instrucciones del proveedor. En el caso de la preparación de pequeñas bibliotecas de ncRNA, 10 microg de ARN total fueron tratados con el kit RNA Clean Concentrator™-5, y se obtuvo una fracción enriquecida de 17-200 nt de ARN. Se seleccionaron fragmentos con una longitud de 24-30 nt en gel. Se procedió a hacer las bibliotecas de cDNA con el Illumina TruSeq Small RNA Library Preparation kit, de acuerdo a las recomendaciones del fabricante. La secuenciación del ARN fue realizada por la Unidad Universitaria de Secuenciación Masiva de DNA (UUSMD)-UNAM (<http://www.uusmd.unam.mx/>), utilizando un instrumento de Illumina NextSeq 500.

Análisis de los datos del ARN-secuencia y evaluación de genes expresados diferenciales

El coeficiente de variación del porcentaje de cobertura del cuerpo genético se determinó con RseQC (<http://rseqc.sourceforge.net>) y los perfiles de expresión del recuento bruto se procesaron por el número total de lecturas de la biblioteca para cada afección utilizando herramientas de recuento htseq (http://htseq.readthedocs.io/en/release_0.9.1/). Las comparaciones de genes entre tratamientos para la expresión diferencial se realizaron utilizando el paquete EdgeR en R (<http://bioconductor.org/packages/release/bioc/html/edgeR.html>). Realizamos nuestro análisis de acuerdo con las instrucciones proporcionadas. Para el recuento de genes se utilizó el método TMM para normalizar el tamaño efectivo de la biblioteca y la prueba de cociente de probabilidad del modelo lineal generalizado (GLM) con una tasa de descubrimiento falso del 1% se utilizó como curado para llamar a la expresión diferencial.

Exploración de pequeños datos de ARN-secuencia e identificación de pequeños ncRNAs generados a partir de ncRNAs

Cuatro pequeñas bibliotecas de ARN fueron preparadas y las lecturas fueron alineadas usando Bowtie (<http://bowtie-bio.sourceforge.net>). Los parámetros de mapeo se establecieron para alinear las lecturas de mejor calidad (- phred64-quals), reportando la mejor alineación sin desajustes. Se analizaron pequeños datos de ARN-secuestro con herramientas Galaxy para detectar la expresión del nivel y la correlación entre la réplica biológica y el tratamiento diferente. Para identificar nuevos ncRNAs pequeños, utilizamos los gráficos visualizados en IGV de pequeñas lecturas de datos de ncRNA (ver arriba) para identificar manualmente las regiones no codificantes transcritas. Las regiones no codificantes transcritas se verificaron utilizando el procedimiento descrito por Ilnytsky y Bilichak en 2016 y se evaluaron para identificar qué cadena podría ser transcrita y predecir posibles nuevos miARNs y/o siRNAs. Se utilizaron búsquedas en la Herramienta de Búsqueda de Alineación Local Básica (BLAST; <http://blast.ncbi.nlm.nih.gov>) para verificar que las regiones no coincidían con proteínas anotadas en otros genomas. Este enfoque se utilizó en conjunto con las búsquedas en las bases de datos Rfam (<http://rfam.sanger.ac.uk/>) y miRBase (<http://www.mirbase.org/>).

Las librerías presentan lecturas de alta calidad con un rango de extensión entre 15-30 nt y cerca de 14 millones de lecturas promedio. De acuerdo al análisis bioinformático descrito por Ilnytsky y Bilichak, 2017 (23), se determinó la eficiencia del mapeo de lecturas Fig.3. Las librerías de crecimiento en glicerol tienen mayor eficiencia que las obtenidas en glucosa e incluso menor dispersión que las mismas.

Los archivos de alineamiento solo se generaron con las lecturas que presentan posiciones

únicas en el genoma (aproximadamente 1 millón de lecturas), y los mismo se analizaron para conocer la cobertura sobre el percentil génico Fig.4 y ensayos posteriores. Como se observa en el gráfico, los picos de cobertura son casi coincidentes a lo largo del gen excepto para dos fragmentos, el 65 percentil y el segmento entre el 88-92 percentil. En cada caso existe una relación casi del doble entre una condición de tratamiento con respecto a otra. En el caso del 65 percentil los datos de las librerías obtenidas en el crecimiento en glucosa presentan su valor máximo de cobertura, mientras que en el segmento del 88-92 se alcanzan los valores máximos de cobertura para los datos de obtenidos en el crecimiento glicerol/acetato.

La visualización de los datos de secuenciación muestran solamente una parte de su complejidad para el procesamiento computacional y para su interpretación biológica Fig.5. Lecturas son mapeadas en regiones no anotadas del genoma, en genes codificantes a^[OBJ] proteínas, tRNAs, rRNAs, ncRNAs , en regiones corriente arriba al sitio de inicio de transcripción etc. Igualmente se destaca que ciertas regiones muestras expresión diferencial, por lo que puede sr considerado e hecho que no es un proceso azaroso, o de transcripción generalizada. Sin embargo es necesario demostrar que no constituyen un artefacto experimental.

Imágenes de resultados preliminares a partir de los datos de las bibliotecas de *small RNA*.

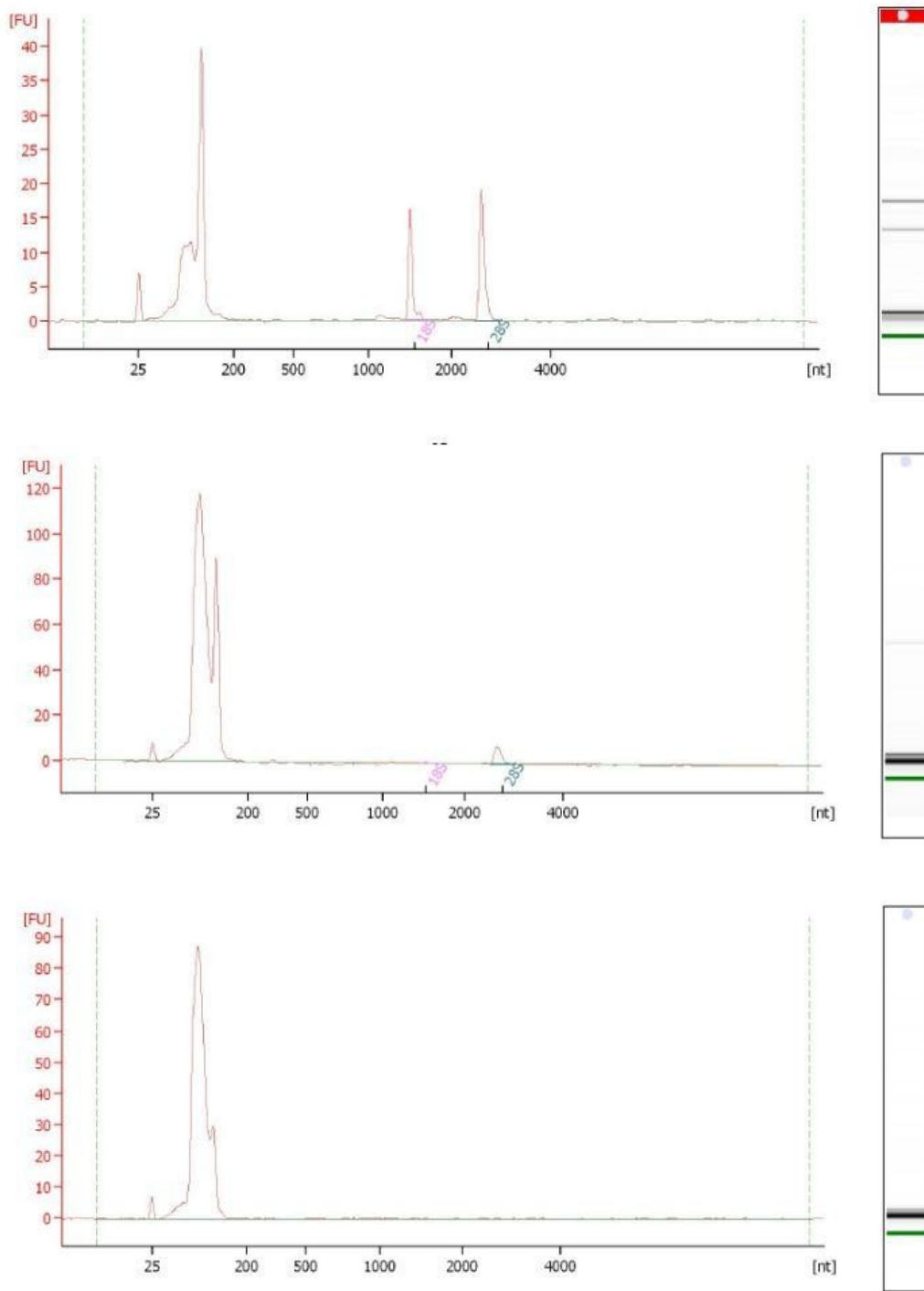


Fig. 1: Cuantificación e integridad de RNA. RNA total (superior), RNA total procesado con el kit Ribominus Yeast (media) y fracción enriquecida con RNAs > 200 nt.

Descripción de las librerías.

Total de lecturas en las ibrerías y distribución de lecturas por tamaño o extensión de nucleótidos.

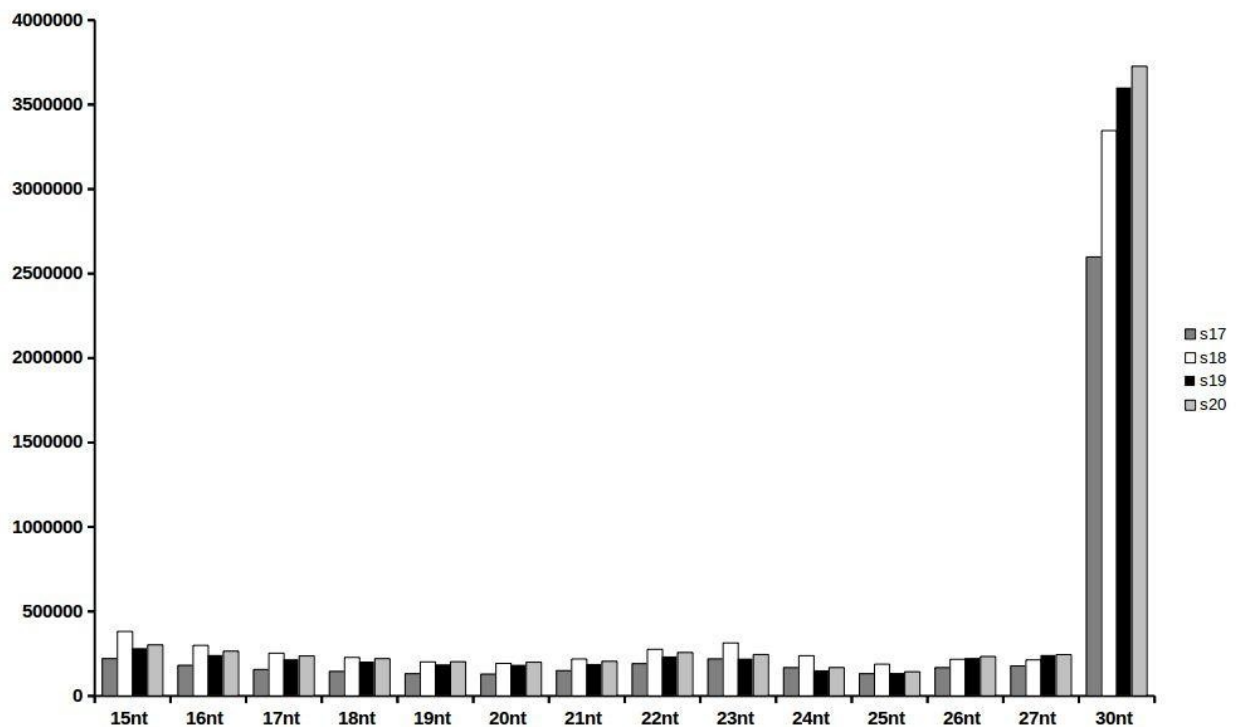


Fig. 2: Distribución de las lecturas de cada biblioteca de acuerdo a su extensión. Las muestras s17 y s18 corresponden réplicas biológicas (cultivos independientes) obtenidos en glucosa y las muestras s19 y 120 al trascurso en glucosa.

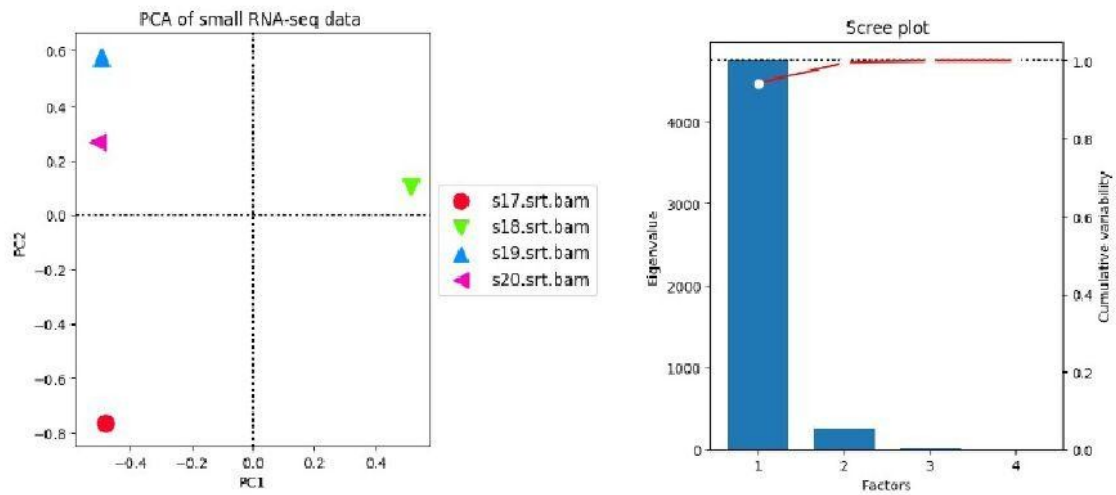


Fig. 3: Análisis de componentes principales de las bibliotecas obtenidas. Gráfico de PCA (Izquierda) .

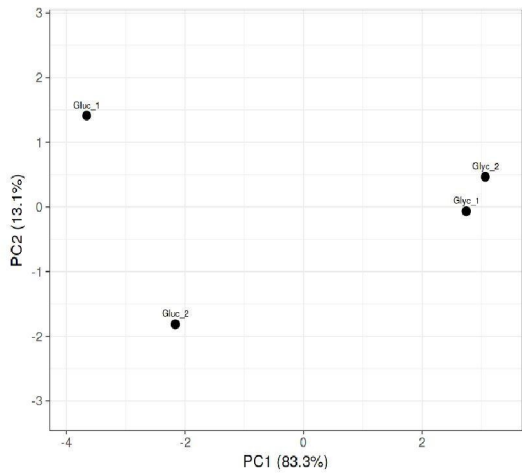
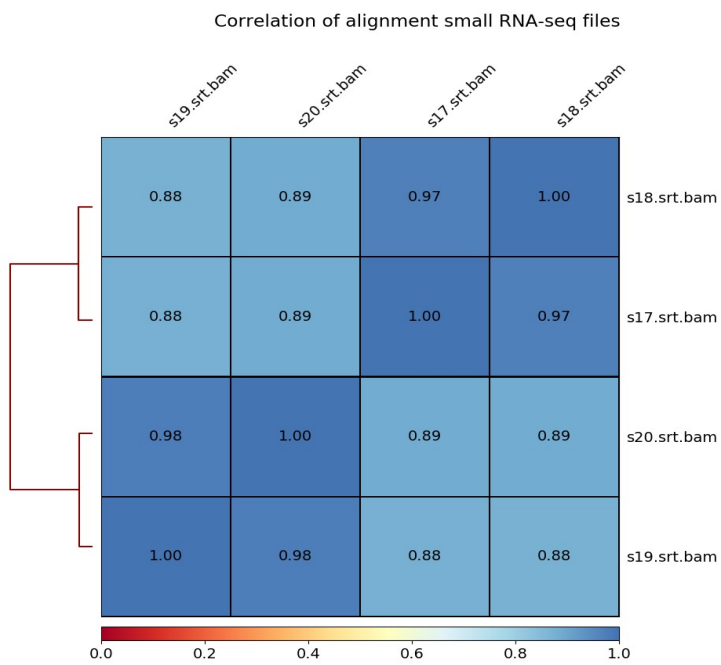
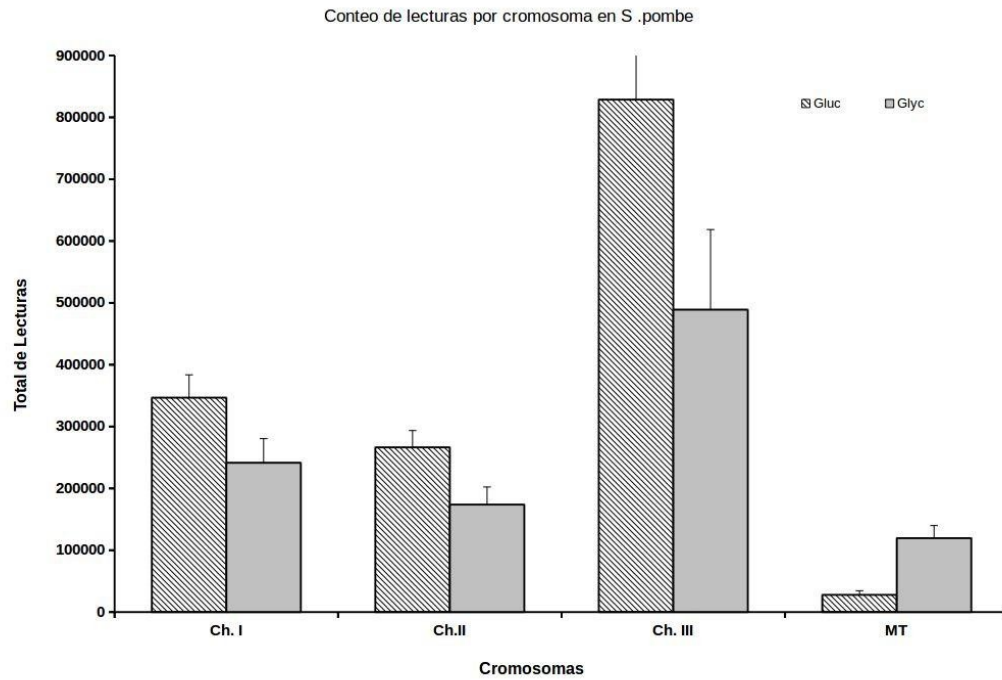


Fig. 4: Gráfico de componentes Principales en R de las librerías normalizadas. Izquierda crecimiento en Glucosa y derecha crecimiento en glicerol





*Fig. 5: Análisis de lecturas mapeadas a los tres cromosomas de *S. pombe* y al cromosoma mitocondrial. Las lecturas obtenidas en el crecimiento en glucosa corresponden a las barras con rayas. Las lecturas correspondientes a las barras grises constituyen a crecimiento en glicerol/ acetato.*

Lecturas sobre los tipos de ncRNAs.

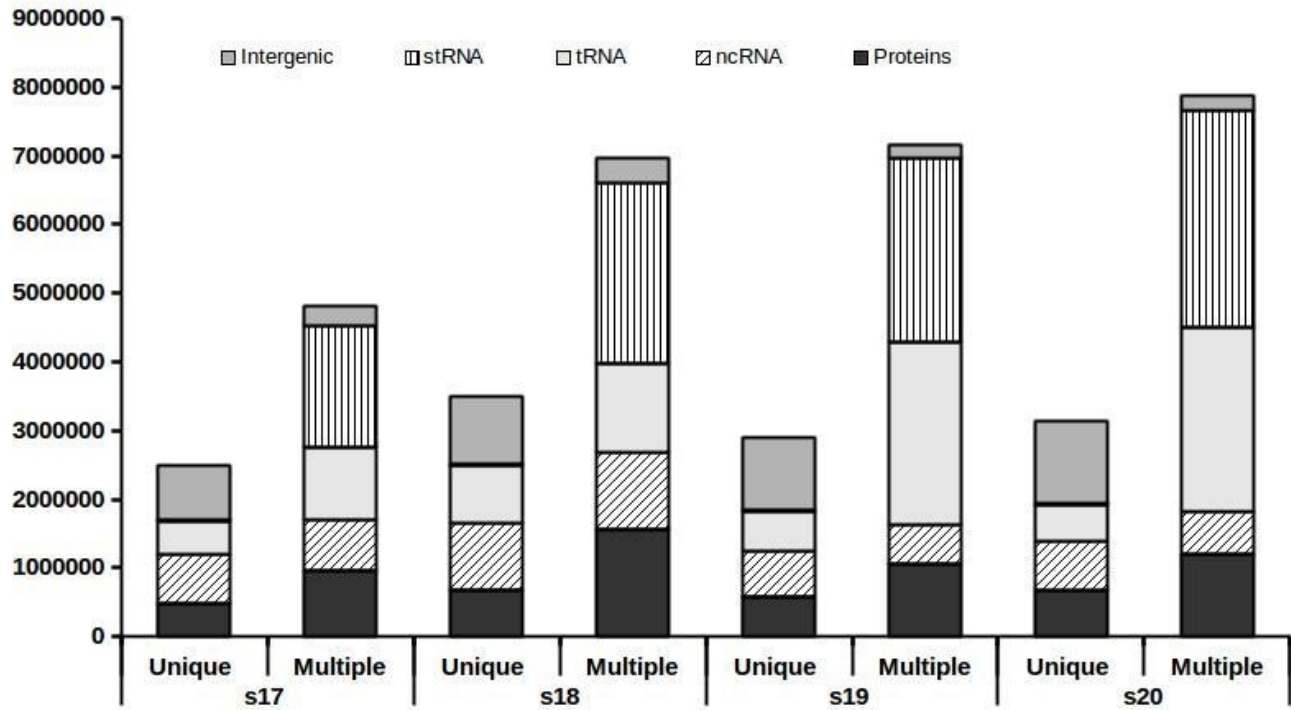


Fig. 6: Distribución de lecturas de acuerdo a los genes canónicos presentes en el genoma. Las lecturas se dividieron en las que se encuentran de forma única y múltiple en el genoma. de *S. pombe*

Lecturas sobre Proteínas, densidad de las lecturas.

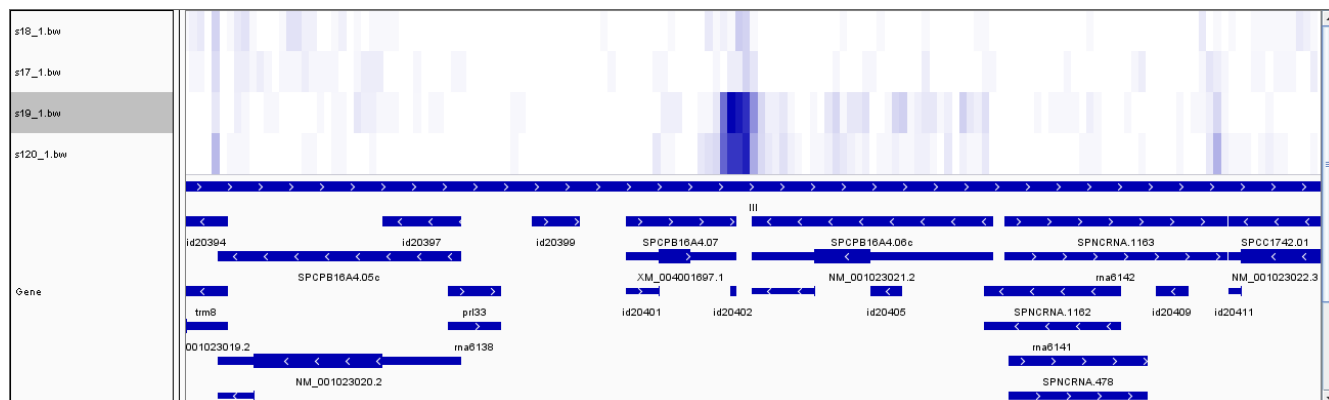


Fig. 9: Análisis de lecturas alineadas a genes anotados en el Genoma de *S. pombe*.

Análisis de lecturas alineadas sobre ncRNAs.



Fig. 10: Exploración de genes anotados como ncRNA y en los archivos formato BAM obtenidos.

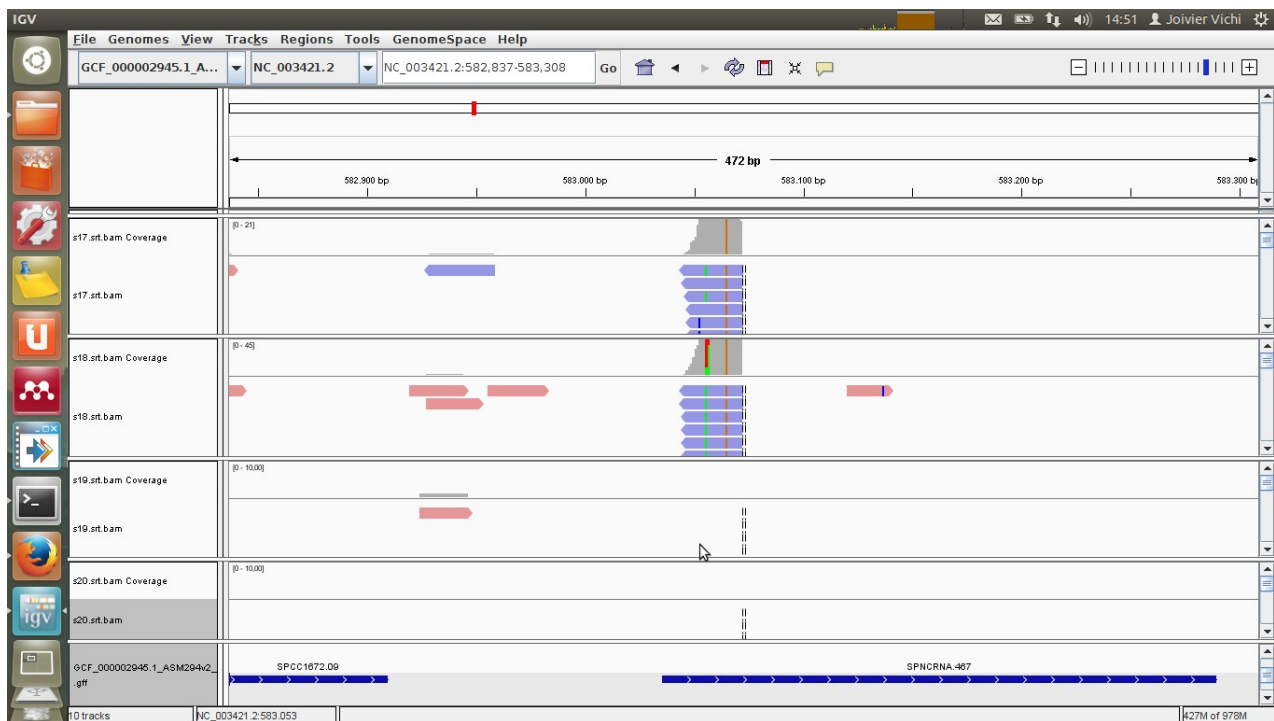


Fig. 11: Expresión diferencial de regiones génicas en el interior de ncRNA.

Generación de fragmentos de ncRNAs.

Lecturas asociadas a ncRNAs

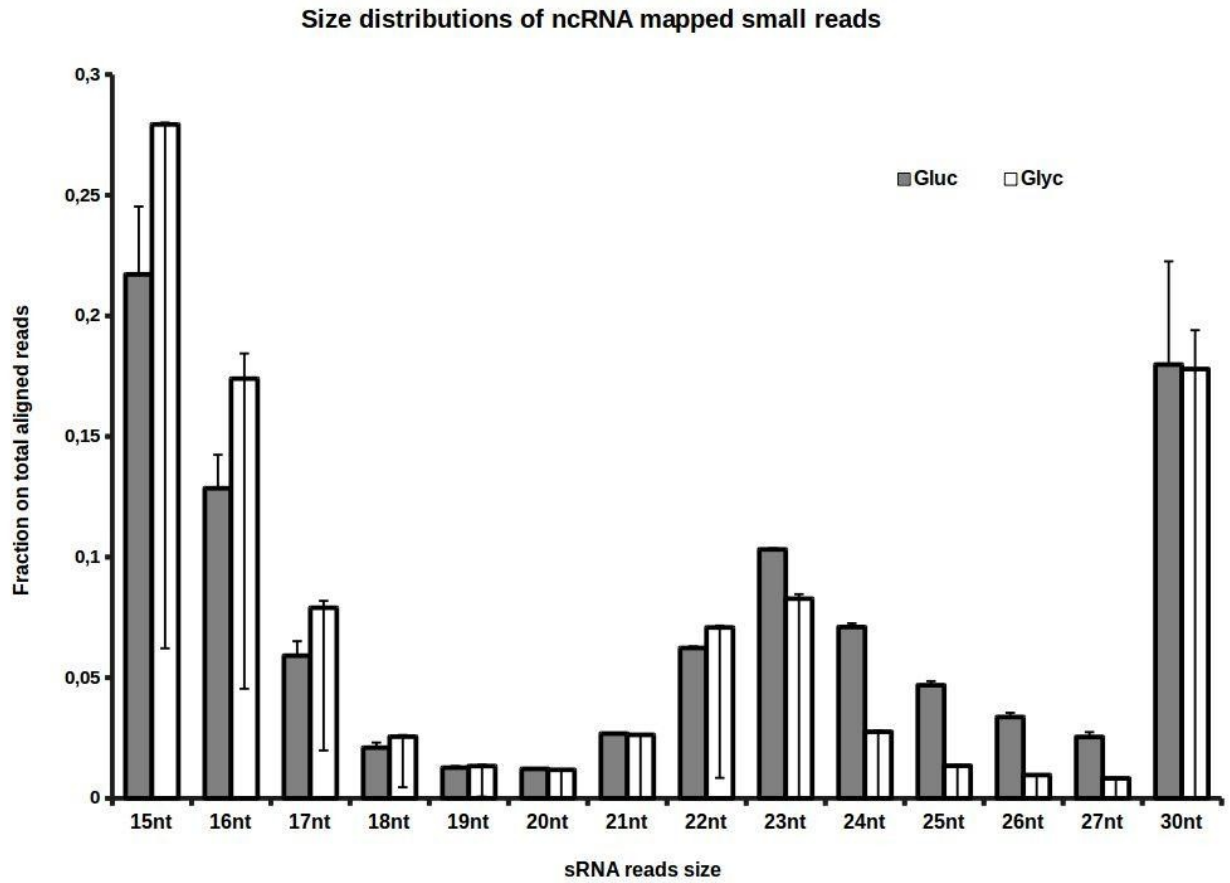


Fig. 12: Distribución de lecturas pequeñas alineadas con genes ncRNA anotados en el genoma de *S. pombe* de acuerdo a su extensión.

Análisis Global de ncRNFs enriquecidos.

Análisis de genes potencialmente precursores de ncRNAs por mecanismos de degradación y por origen de transcripción.

Label	Gluc	Glyc
CUT	156	148
DUT	61	38
STAB	131	106
XUT	44	36

The chi-square statistic is 3.3285. The p-value is .343695. The result is not significant at $p < .05$.

Biotype	Gluc	Glyc
Antisense *	119	69
Bidirectional	114	115
Intergenic	150	129
Overlapping	9	15

* The chi-square statistic is 10.7792. The p-value is .012982. The result is significant at $p < 0.05$.

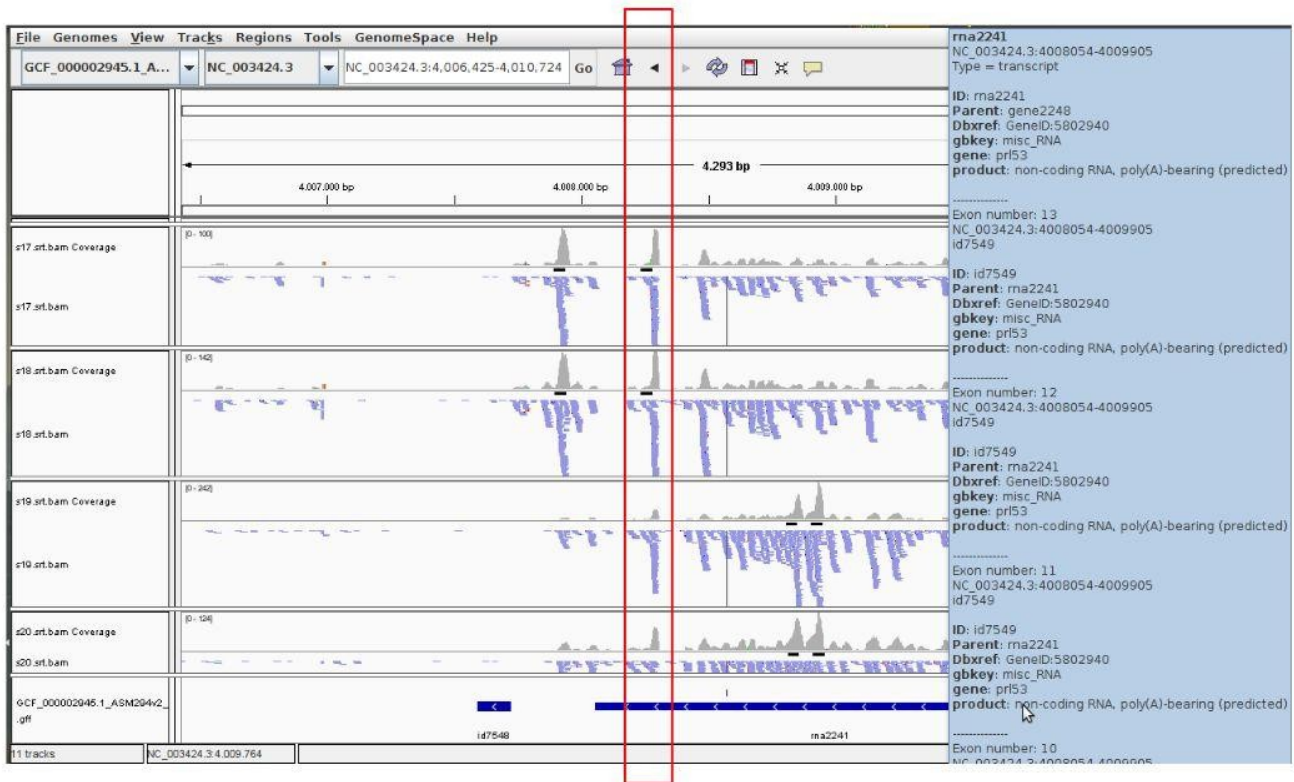


Fig. 14: Visualización del alineamiento de candidato a microRNA en prf53. el ncRNAs se observa expresado diferencialmente. s17 y 18 corresponden a las réplicas biológicas presentes en el crecimiento en glucosa. s19 y 120 réplicas biológicas en crecimiento en glicerol/acetato.

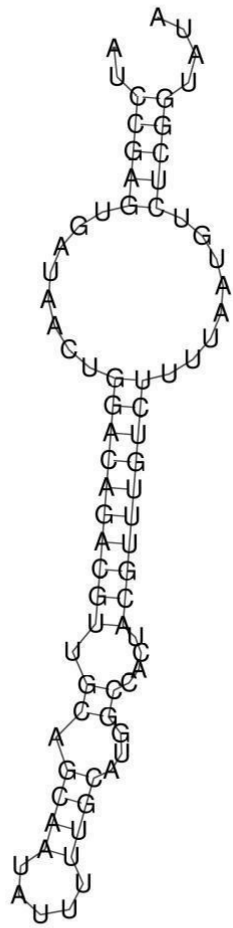
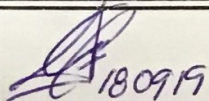
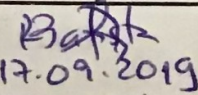
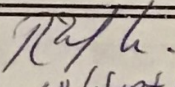
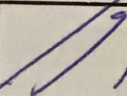
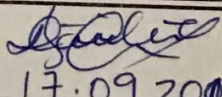
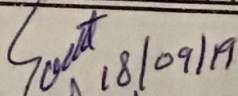
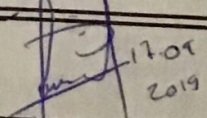


Fig. 15: Estructura generada del ncRFs de pr153 a partir del análisis con mirDeep.

DR. VICTOR BARBA LÓPEZ
COORDINADOR DEL POSGRADO EN CIENCIAS
PRESENTE

Atendiendo a la solicitud para emitir DICTAMEN sobre la revisión de la TESIS titulada **"Análisis comparativo del transcriptoma de *Escherichia coli* y *Schizosacchomyces pombe* en dos escenarios metabólicos"** que presenta el alumno Joivier Vichi Lozada (5920130701) para obtener el título de **Doctor en Ciencias**.

Nos permitimos informarle que nuestro voto es:

NOMBRE	DICTAMEN	FIRMA
Dr. José Luis Puente IBT-UNAM	Aprobado	 18/09/19
Dr. Ramón Alberto Batista García CIDC-UAEM	Aprobado	 17.09.2019
Dra. Rosa M. Gutiérrez Ríos IBT-UNAM	Aprobado	 18/Sep/2019
Dr. Ricardo A. Grande Cano IBT-UNAM	APROBADO	 11/09/19
Dra. Sonia Dávila Ramos CIDC-UAEM	Aprobado	 17.09.2019
Dr. Enrique Morette Sánchez IBT-UNAM	APROBADO	 18/09/19
Dr. Armando Hernández Mendoza CIDC-UAEM	Aprobado	 17.09.2019