

Análisis transcripcional de *Rhodococcus jostii*
durante la acumulación de triglicéridos.

Autor: Val Fernández Lanza

Director: Fernando de la Cruz Calahorra

Master de Biomedicina y Biotecnología 2010-2011

Facultad de Medicina - Universidad de Cantabria

Índice:		<i>página:</i>
1. Introducción	1
- <i>Biosíntesis, acumulación</i>	2
<i>y metabolización de triglicéridos en Rhodococcus</i>		
- <i>Estudios de transcriptómica en organismos prokariotas</i>	10
- <i>Objetivo: Análisis del transcriptoma de</i>	13
<i>Rhodococcus jostii RHA1 en medio mínimo suplementado.</i>		
2 Metodología	14
- <i>Condiciones de cultivo</i>	14
- <i>Extracción y análisis de lípidos</i>	14
- <i>Secuenciación</i>	15
- <i>Alineamiento frente al genoma de referencia</i>	15
- <i>Cuantificación de la expresión de cada uno de los genes</i>	16
- <i>Normalización y tratamiento estadístico</i>	17
- <i>Análisis funcional</i>	19
3 Resultados	20
- <i>Resultados del análisis estadístico</i>	22
- <i>Alineamiento y mapas de expresión</i>	25
- <i>Análisis funcional</i>	27
4 Discusión	35
5 Conclusiones	38
6 Referencias	39

Introducción.

El mundo se enfrenta actualmente a una grave crisis energética. Por un lado las fuentes accesibles de petróleo y otros recursos fósiles se están lenta pero continuamente empobreciendo. Por el otro, la demanda de recursos fósiles está aumentando debido a la industrialización mundial, en particular, los países con grandes poblaciones como China e India. Estas necesidades están creando gran expectativa sobre las energías renovables. Los macroproyectos de investigación que están financiando Europa, Estados Unidos o Japón, así como compañías estratégicas del sector energético, ilustran las necesidades ante las cuales nos encontraremos en un futuro cercano. Los triglicéridos como otros ácidos grasos de valor añadido, como los aceites Omega 3, se encuentran entre las sustancias más interesantes para la producción de energía, fertilizantes o complementos alimenticios. Un ejemplo destacado es el metanol que se está utilizando actualmente en la producción de biodiesel. El biodiesel es producido a partir de metanol sintético de la industria química y a partir de triglicéridos (TG) por transesterificación química. El biodiesel y el bioetanol constituyen en la actualidad alrededor del 90% del mercado de los biocombustibles.

Actualmente los TG se obtiene a partir de productos agroalimentarios. Esto produce una dicotomía, por una parte la producción de combustibles a partir de fuentes renovables debería paliar la dependencia de los combustibles fósiles, pero la gran demanda de biocombustibles genera perturbaciones en los mercados alimenticios. En este momento la agricultura mundial no puede sustentar la demanda alimenticia y energética al mismo tiempo. La obtención de TGs a partir del reciclaje de aceites de uso alimenticio (especialmente los aceites usados para freír) es insignificante frente a las cantidades obtenidas en los campos y más aun frente a las necesidades mundiales. Los principales cultivos para la producción de TGs son de colza en Europa, árboles de palma aceitera en el sudeste de Asia, y Soja en América del Norte.

La biotecnología nos abre una puerta hacia una posible solución. En la naturaleza podemos

encontrar diversos microorganismos capaces de sintetizar y almacenar triglicéridos. De esta forma podríamos utilizar las características especiales de algunos microorganismos capaces de crecer en condiciones especiales. Así podríamos no solo producir TGs, sino, aprovechar excedentes industriales tales como los sueros lácticos procedentes de la industria lechera o la xylosa proveniente de la industria papelera. La idea es sencilla, no solo producir TGs, sino también reciclar productos que actualmente se consideran vertidos contaminantes.

1. Biosíntesis, acumulación y metabolización de triglicéridos en Rhodococcus

De entre estos microorganismos destacan los del género *Rhodococcus* [1]. Los miembros del género *Rhodococcus* son bacterias aerobias que son comunes en muchos nichos ambientales como suelos y agua de mar. También como agentes patógenos de mamíferos y plantas. Una característica notable es son capaces de crecer en nichos ambientales sorprendentemente diversos y pueden degradar numerosos contaminantes tóxicos. Taxonómicamente, el *rhodococci* pertenecer a la agrupación más amplia de actinomicetos. Se diferencian de otros actinomicetos debido a la presencia de ácidos micólicos en sus paredes celulares y pertenecen al grupo diverso llamado Mycolata. Todos los *rhodococci* investigados hasta la fecha parecen tener genomas grandes y complejos, que reflejan su diversidad catabólica. Hay una evidencia creciente en los análisis bioquímicos y la secuenciación del genoma, que hay múltiples vías y genes homólogos dentro de muchas cepas, lo que indica la versatilidad del género *Rhodococcus*. A este respecto muchos de ellos también poseen una variedad de grandes plásmidos lineales y pequeños plásmidos circulares que contribuyen también a explicar el inmenso repertorio de las capacidades catabólicas. También parece tener la capacidad de adaptarse fácilmente para degradar muchos sustratos cuando se le presenta un nuevo reto catabólico, sin embargo, los mecanismos genéticos moleculares subyacentes a la flexibilidad del genoma de *Rhodococcus* aun no se han dilucidado.

Una característica notable es la presencia de plásmidos lineales muy grandes en muchas de las cepas aisladas. Edemas es contienen un sistema que promueve la alta frecuencia de recombinación ilegítima y la presencia de transposones (aunque se han identificado un pequeño número de secuencias de inserción). Ocurren eventos de recombinación que pueden servir para promover la incorporación de ADN en el genoma, sin la ayuda de elementos genéticos móviles.

Los miembros del género *Rhodococcus* son especialistas en la acumulación de triglicéridos (TGs) [2]. Algunos de ellos pueden ser considerados microorganismos oleaginosos, ya que son capaces de producir cantidades significativas de los lípidos en determinadas condiciones. La biosíntesis y acumulación de TGs por los miembros de *Rhodococcus* y otros actinomicetos parece ser un proceso relacionado con la fase estacionaria de crecimiento o como una respuesta al estrés.

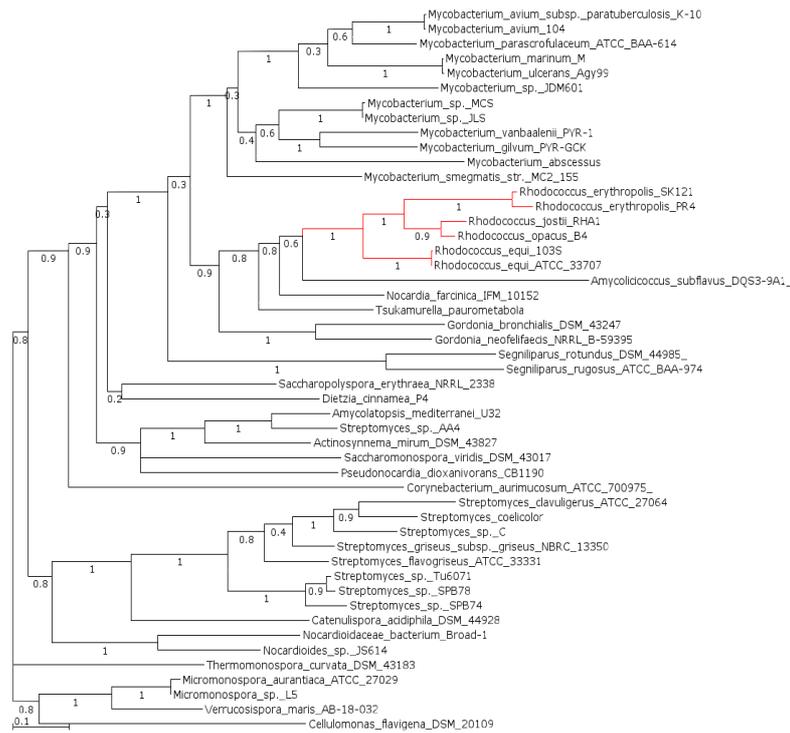


FIGURA 1: Árbol filogenético de los miembros cercanos del género *Rhodococcus*. Árbol construido a partir del RNA ribosomal 16S. Se alinearon las secuencias con Muscle [3] para posteriormente reconstruir el árbol con el software FastTree [4]

La estructura química de los TG de *rhodococcal* puede ser controlada por la composición de la fuente de carbono utilizada. Se ha observado que miembros del género *Rhodococcus* y géneros afines son capaces de biosintetizar y acumular TG con componentes inusuales, tales como los ácidos grasos aromáticos e isoprenoides. La baja especificidad de la ester sintetasa de ceras/diacilglicerol acetiltransferasa (WS/DGAT), las enzimas que catalizan la biosíntesis de TG en procariontes, pueden contribuir a la alta variabilidad de la composición de TG. La presencia de genes que codifican para WS/DGAT es muy redundante en los genomas *rhodococcal*. El enriquecimiento de los genes y las enzimas implicadas en el metabolismo de TG en los *rhodococci* sugiere el importante papel de estos lípidos en la fisiología de estos microorganismos.

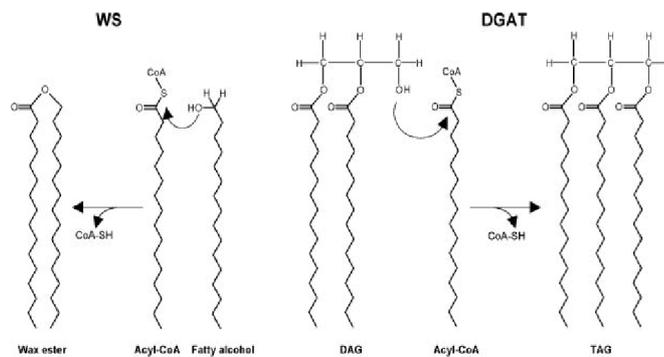


FIGURA 2: Reacciones de biosíntesis de los triglicéridos (TGs) y de las Ceras (WS) mediante las enzimas WS/DGAT.

Se estipula que la ruta más probable para la biosíntesis de TG es un proceso secuencial de reacciones dependientes de Acyl-CoA llamada ruta de *Kennedy*. Esta ruta está descrita para levaduras y plantas. Esta ruta está muy bien estudiada en eucariotas pero aún hay incógnitas sin resolver en los organismos procariotas. La ruta metabólica consiste en la acilación secuencial de las posiciones sn-1, 2 de glicerol-3-fosfato, lo que resulta en la formación de ácido fosfatídico. La eliminación del grupo fosfato catalizada por la enzima fosfatasa ácido-fosfatídico se produce antes de la etapa final de acilación. En la tercera reacción de acilación, un residuo acil es transferido a la plaza vacante de diacilglicerol, que es el último paso de la biosíntesis de TG. Las tres reacciones son catalizadas por aciltransferasas diferentes. Las especificidades de las diferentes aciltransferasas determinan la distribución de los diversos grupos acilo de los grupos hidroxilo de la espina dorsal del glicerol y, por tanto, la composición final de acilos en los TG biosintetizados. El ácido fosfatídico y el diacilglicerol generado en la ruta de *Kennedy* también se utilizan para la síntesis de fosfolípidos que ocurren en las membranas. Por lo tanto, el tercer paso de acilación de la espina dorsal del glicerol es la única reacción enzimática para la biosíntesis de TG. Esta reacción es catalizada por una enzima diacilglicerol aciltransferasa (DGAT). Kalscheuer y Steinb [5] identificaron la primera DGAT procariotas en *Acinetobacter baylyi* ADP1, que exhiben simultáneamente actividades DGAT y acil-CoA:aciltransferasa de alcoholes grasos (sintasa éster de cera, WS). La cepa ADP1 acumula principalmente ésteres de cera y TGs como componentes menores, con proporciones de hasta 6,9% y el 1,4% del peso seco celular, respectivamente. Curiosamente, las WS/DGAT de *A. baylyi* ADP1 representa una nueva clase de enzima de síntesis de TGs, que no presenta similitud de secuencia con cualquier aciltransferasa eucariota conocida [5].

En las proteínas WS/DGAT de la cepa ADP1 y de otras cepas cercanas, podemos encontrar un motivo HHxxxDG altamente conservado. Este motivo puede ser el responsable de la actividad catalítica en la formación del cuerpo central de ésteres de los TGs. Más tarde, se describieron varios WS/DGATs en bacterias capaces de almacenar ceras y TGs. Como podemos observar en el árbol filogenético, mientras que en las bacterias Gram-negativas sólo encontramos unos pocos WS/DGATs capaces de producir ceras y TG, en los actinomicetos, como los géneros *Mycobacterium*, *Nocardia* y *Rhodococcus* encontramos una gran redundancia de estas enzimas.

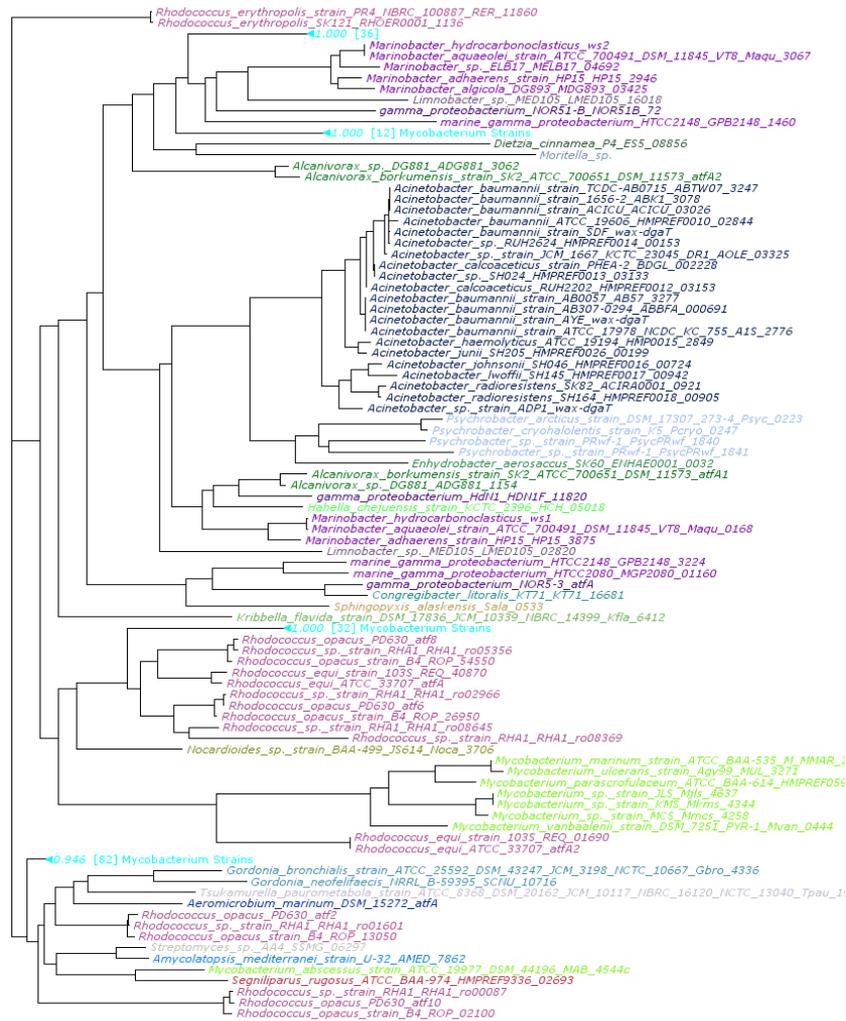


FIGURA 3: Árbol Filogenético de las WS/DGAT homologas. Se alinearon las secuencias con Muscle[3] y se reconstruyó el árbol mediante máxima verosimilitud con el programa FastTree[4].

Rhodococcus jostii RHA1 es capaz de acumular grandes cantidades de TG, además de compuestos de almacenamiento, tales como los PHA, glucógeno y polifosfatos [6]. Esta cepa posee todos los genes necesarios y enzimas para la biosíntesis de TG a través de la ruta de *Kennedy*, como ocurren en el genoma de *R. opacus* B y *R. erythropolis* PR4 y SK121. Un análisis bioinformático del genoma identificó 14 genes que codifican posibles enzimas WS/DGAT que participan en la biosíntesis de TG y ceras. Un total de 54 genes que codifican para posibles enzimas lipasas y enzimas esterasas, posiblemente implicados en la degradación de los TGs y ceras. Tres conjuntos de genes que codifican las PHA-sintetasas y PHA-despolimerasas. Seis genes que codifican las enzimas clave del metabolismo del glucógeno, un gen que codifica para una posible polifosfato quinasa, y tres hipotéticos genes exopolifosfatasa posiblemente implicados en la biosíntesis y la degradación de polifosfato [6]. Uno de estos WS/DGATs contiene el hipotético motivo sitio-activo de WS/DGATs

(HHxxxDG), mientras que en *atf4*, *atf10* y *atf14*, la segunda histidina del motivo se sustituye por lisina, serina y prolina, respectivamente. Algunos genes *atf* se encuentran en el cromosoma RHA1, mientras que *atf12*, *atf13* y *atf14* se encuentran en el plásmido pRHL1.

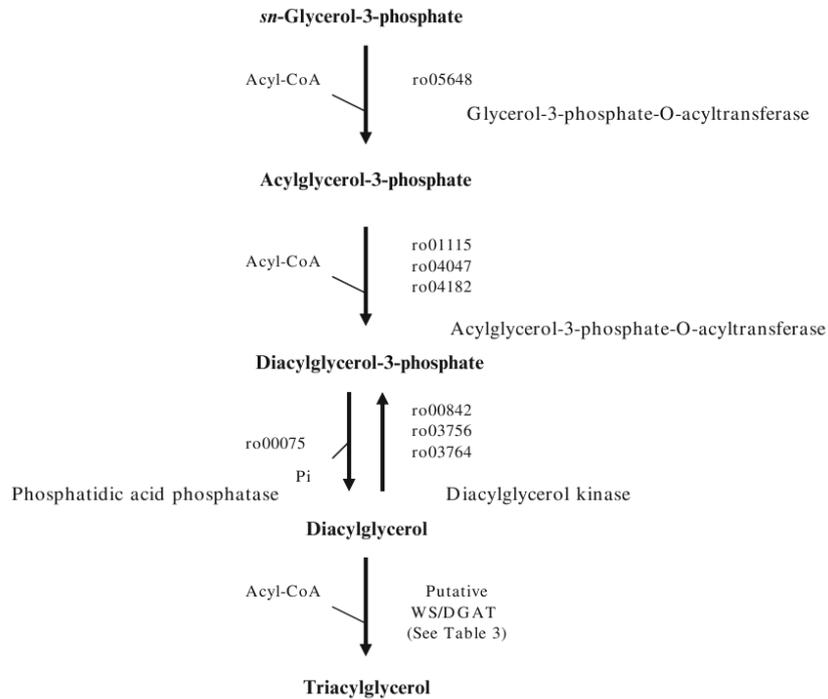


FIGURA 4: Ruta de síntesis de los TGs. En el diagrama se muestran los genes codificantes de las hipotéticas proteínas de la ruta de Kennedy en el genoma de *Rhodococcus jostii* RHA1.[1]

La biosíntesis de los TG requiere una red metabólica eficiente, capaz de producir los precursores y la energía necesaria para las reacciones específicas. En general, el metabolismo central de *rhodococci* posee una gran flexibilidad y diversidad de reacciones metabólicas, utiliza la energía necesaria para el proceso de biosíntesis de los TGs en determinadas condiciones a partir de una diversidad de fuentes de carbono. Las vías del metabolismo central *rhodococcal* son capaces de convertir de manera eficiente diversas fuentes de carbono a los intermediarios metabólicos claves, como el piruvato, acetil-CoA y glicerol-3-fosfato, el poder reductor necesario que es requerido por las vías de biosíntesis de lípidos y para producir la energía necesaria en forma de ATP. Sin embargo, muchas bacterias que no son capaces de acumular TGs también son capaces de producir estos intermediarios metabólicos, el poder reductor y el ATP. Por lo tanto, un microorganismo oleaginoso también debe ser capaz de mantener un flujo de carbono hacia las vías de alta producción de lípidos. Dado que la acumulación de TGs es un proceso que necesita una gran cantidad de carbono y energía, la células de *rhodococcal* son capaces de detener el crecimiento celular y la replicación y cambiar su metabolismo y flujo de carbono a la vía de la biosíntesis de lípidos. Estos cambios en el metabolismo celular depende de los estímulos del ambiente, como se mencionó anteriormente.

Diversas vías pueden contribuir a la producción del reservorio de acetil-CoA en *rhodococci*. La conversión de la acetil-CoA a partir del piruvato derivado de la glicólisis podría ser la principal vía de flujo de carbono para la biosíntesis de ácidos grasos. El intermediario acetil-CoA podría producirse por la reacción catalizada por la enzima citrato liasa. La citrato liasa, que convierte el ácido cítrico en acetil-CoA y oxalacetato, es una de las enzimas clave del ciclo de Krebs reductivo. La presencia de citrato liasa y la sintasa 2-oxoglutarato en las bases de datos del genoma de *R. jostii* RHA1 y *R. opacus* B4 sugiere que estos microorganismos son capaces de conducir el ciclo del TCA en el sentido reductivo. Esto permite al metabolismo la incorporación de CO₂ para la síntesis de productos intermedios, que pueden alimentar las vías de biosíntesis de lípidos en condiciones restrictivas. Por otra parte, el acetato libre podría ser activado a acetil-CoA mediante una acetil-CoA-sintetasa en una reacción dependiente de ATP. Esta enzima, junto con la acetato quinasa y la fosfo-transacetilasa, que fueron detectadas en el genoma de *R. jostii* RHA1 y *R. opacus* B4, pueden estar implicadas en el mantenimiento del reservorio intracelular del acetil-CoA y acetil-P en estos microorganismos. El otro metabolito intermediario necesario para la biosíntesis de ácidos grasos en las células de *rhodococci* es el propionil-CoA, que generalmente se utiliza para la síntesis de los ácidos grasos impares.

La síntesis de ácidos grasos requiere cantidades estequiométricas de ATP, acetil-CoA, NADH y NADPH para cada adición de C₂ en las reacciones catalizadas por la acetil-CoA carboxilasa y sintetasa de ácidos grasos. El ATP necesario puede ser generado, a nivel de sustrato, por la fosforilación en *rhodococci* a través de la glicólisis, entre otras posibles reacciones de generación de ATP. La fuente de poder reductor para la biosíntesis de ácidos grasos en *rhodococci* en realidad no se conoce, aunque la vía de las pentosas fosfato podría ser una posible fuente de NADPH. La enzima malica podría estar implicada en la generación de NADPH en fuentes de carbono, que probablemente tienen un bajo flujo a través de la vía de las pentosas fosfato.

Los motivos por los cuales los organismos procariotas acumulan ceras y TGs han incitado numerosos estudios. Son varias las utilidades que les pueden dar las células entre las cuales se encuentran la reserva de energía y carbono, fuente para los precursores de las membranas y el desarrollo celular, para la detoxificación de los ácidos grasos libres o como elemento de control en el balance del metabolismo central.

La acumulación de cantidades significativas de TGs por *rhodococci* es un proceso que exige grandes cantidades de carbono y energía, que compite con el crecimiento celular. Por lo tanto, la aparición de compuestos de almacenamiento, en los microorganismos que habitan en ambientes pobres en energía debe ser una característica importante de su fisiología. Se sabe que los TGs son un excelente material de reserva debido a sus propiedades extremadamente hidrofóbica, lo que permite su acumulación en grandes cantidades en las células sin necesidad de cambiar la osmolaridad del

citoplasma. Además, la oxidación de los TG produce el máximo rendimiento de la energía en comparación con los compuestos de almacenamiento, tales como hidratos de carbono y PHA, ya que los átomos de carbono de los restos acilo de los TG están en su forma más reducida. La represión del metabolismo en condiciones desfavorables para el crecimiento permite una lenta utilización de los lípidos almacenados, que pueden movilizarse de forma programada. La energía obtenida por la lenta movilización de los TG almacenados puede ser útil en los procesos bioquímicos y mecanismos fisiológicos ante la necesidad de adaptación al medio. Este proceso puede proporcionar a las células de autonomía energética y una independencia temporal del medio ambiente y contribuir para la supervivencia celular cuando no tienen acceso a los recursos energéticos.

Por otra parte, los TG pueden servir como precursores para la biosíntesis de ácidos micólicos por parte de algunos actinomicetos que producen estos ácidos, cuando crecen en medios estresantes. Los ácidos micólicos son ácidos grasos de cadena larga producidos por el alargamiento de los ácidos grasos normales, que son componentes clave para la integridad y la función de la membrana celular. Se ha observado que, en condiciones de desecación, no aparecen diferencias significativas en la membrana de *R. opacus* PD630 por lo que la adaptación de su fluidez y la permeabilidad puede ser el resultado de la variación del contenido de lípidos en respuesta al estrés hídrico [2]. Este suceso parece estar balanceado por la concentración de ácidos micólicos. Dado que los ácidos micólicos son producidos por el alargamiento de los ácidos grasos por el tipo II de la sintasa de ácidos grasos (FAS II), se ha estudió el efecto de la isoniazida, que es un inhibidor del sistema FAS II, sobre la supervivencia en estrés hídrico. Los resultados sugieren que el recambio de ácidos micólicos con los ácidos grasos preformados contenidos en los TGs contribuye a la adaptación de la envoltura celular bajo condiciones de escasez de agua en *R. opacus* PD630 [7]. Los TGs también pueden desempeñar un papel en la regulación de la composición de ácidos grasos de los lípidos de membrana, con el fin de adaptar su fluidez. Además, los TGs pueden servir como donantes de ácidos grasos para la biosíntesis de fosfolípidos en condiciones nutricionales desfavorables.

Existen estudios que sugieren que los TGs sirven como aceptores ácidos grasos inusuales, que de otro modo perturbarían la fluidez de la membrana durante la degradación de hidrocarburos en condiciones que normalmente se producen en el medio ambiente [8][9][2]. Así, las especies de *Rhodococcus* y actinomicetos cercanos, parecen poseer mecanismos metabólicos que permiten a las células mantener las condiciones fisiológicas de la membrana citoplasmática durante la degradación de hidrocarburos en medios de cultivo limitantes.

La biosíntesis de TGs por *rhodococci* también puede ser un mecanismo para equilibrar el metabolismo central frente a un eventual exceso de intermediarios, tales como la acetil-CoA reductasa, o el poder reductor, en condiciones fluctuantes que con frecuencia se producen en entornos

naturales. Algunos estudios previos revelan que la limitación de oxígeno promueve la acumulación de TGs por los miembros de *Mycobacterium* y los géneros *Rhodococcus* [10]. Cuando los receptores terminales de electrones no tienen el suministro suficiente durante el cultivo de células en condiciones de ventilación limitada, los TGs pueden servir como sumidero de equivalentes de reducción. Bajo condiciones limitantes de oxígeno, el exceso de poder reductor puede inhibir algunas enzimas clave del metabolismo central en las células. La biosíntesis de los ácidos grasos para la producción de TGs, que consume nucleótidos reducidos de pirimidina, puede evitar su acumulación en las células. Por lo tanto, la biosíntesis de TGs permite a las células equilibrar su metabolismo de acuerdo a los cambios de las condiciones ambientales [2].

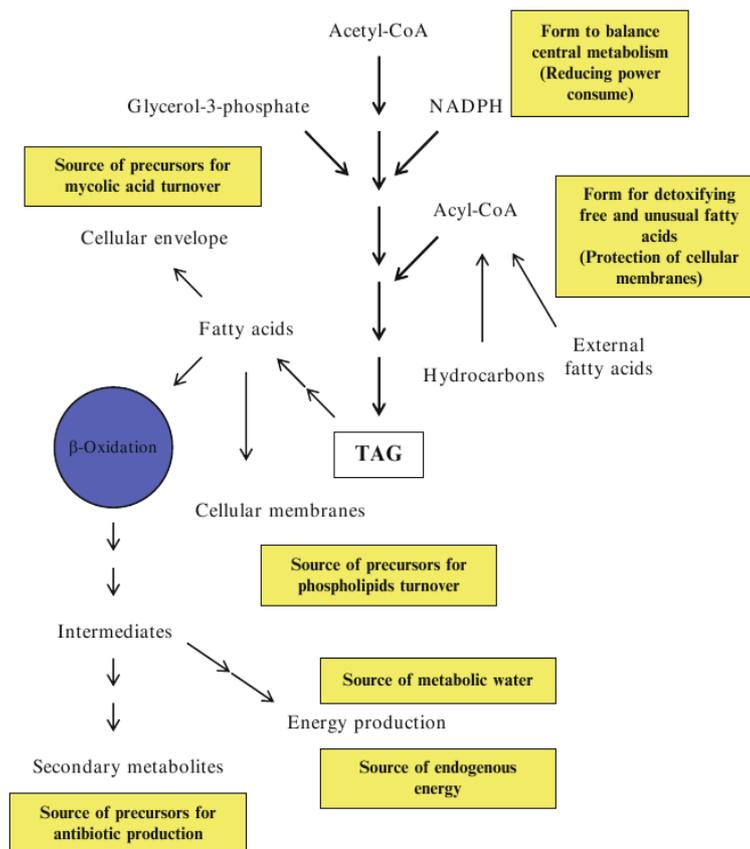


FIGURA 5: Esquema de las funciones fisiológicas de los TGs (TAG) en organismos capaces de producir y almacenar triglicéridos y ceras [1].

Existen evidencias de que los TGs pueden servir como fuente de productos intermedios para la síntesis de compuestos que no son esenciales para el crecimiento, sino para la supervivencia de las células en el medio ambiente. Algunos autores han demostrado que los TGs actúan como fuente de carbono para la biosíntesis de antibióticos a partir de la acetil-CoA reductasa o precursores de la malonil-CoA reductasa, como ha sido descrito en cepas de *Streptomyces* [11]. Se desconoce si este proceso también se produce en los *rhodococci* productores de antibióticos. Por otra parte, los TGs pueden servir como una fuente de intermediarios para la biosíntesis los polímeros extracelulares (EPS)

que se produce como respuesta a diversas condiciones de estrés, tales como la desecación, en los miembros de *Rhodococcus*. Se ha reportado que *opacus* R. PD630 es capaz de acumular progresivamente EPS en la superficie de las células durante la incubación en condiciones de desecación[7]. Dado que la biosíntesis de polisacáridos es fuertemente dependiente de carbono y energía, y no hay ninguna fuente externa de carbono disponible, las células que producen EPS de protección, ha de ser a partir de una fuente de energía y carbono endógenos, tales como los TGs.

2. Estudios de transcriptómica en organismos prokaryotas.

El estudio de los transcriptomas ha sido clave durante las últimas décadas. Desde los primeros pasos que se centraban en el estudio individual mediante Northern Blot hasta la última tecnología de secuenciación que nos permiten medir cuantitativamente la expresión completa de un genoma bacteriano. Durante la última década la tecnología predominante han sido los microarrays. La capacidad de los arrays para medir al mismo tiempo la expresión de miles de transcritos ha conducido a grandes avances en el conocimiento de los organismos prokaryotas. Sin embargo, la tecnología de microarrays tiene varias limitaciones. Por ejemplo, los niveles de “background” de la hibridación (es decir, la hibridación con una sonda que se produce con independencia del nivel de expresión del transcrito correspondiente) limitan la exactitud de las mediciones de expresión, en particular para las transcripciones con baja abundancia [12]. Con las nuevas técnicas de ultrasecuenciación, ahora podemos secuenciar todo el cDNA de un cultivo y cuantificar su abundancia. Esta técnica ofrece grandes ventajas frente a las técnicas de arrays [13]. Principalmente no estamos restringidos por el diseño de las sondas, sino que secuenciamos todos los transcritos que produce el organismo. Tenemos además un rango dinámico muy amplio debido a que no tenemos fenómenos de saturación. Además la cuantificación es directa y no depende de la conversión de una intensidad lumínica, sino que es el resultado de contar las pares de bases que secuenciamos. Existen diversos estudios que muestran las bondades de esta nueva tecnología [14][15] así como numerosas revisiones acerca de estos métodos [16][17][18][19].

La transcriptómica es una poderosa herramienta para la comprensión de la estructura de regulación de los genes y basada en el RNA presente en cualquier organismo en unas condiciones específicas. En contraste con el enfoque reduccionista que estudian los fenómenos biológicos usando un modelo constituido por un pequeño conjunto de genes, la investigación del transcriptoma permite una visión panorámica de los fenómenos seleccionados en todos los genes de forma simultánea. En los eucariotas, como los humanos, ratones, moscas y las levaduras, los transcriptomas se estudiaron inicialmente por secuenciación de millones de EST (etiquetas de secuencias expresadas) y, más recientemente, mediante la secuenciación del cDNA usando ultra secuenciación. El análisis de las

secuencias de cDNA ha alterado en muchos puntos de vista los tradicionales mecanismos de regulación basados en el RNA: por ejemplo, ahora se sabe que el splicing alternativo afecta a la mayoría de los genes humanos, y se ha observado que las transcripciones “*antisense*” superponen más del 10% de todos los genes codificadores de proteínas en los genomas de los metazoos.

Aunque durante décadas los estudios del transcriptoma han sido muy productivos en eucariotas, el estudio de los transcriptomas de bacterias y arqueas se ha pasado por alto hasta hace poco. Una de las razones es que los transcriptomas procariotas eran considerados como simples en comparación con transcriptomas eucariotas; los transcritos procariotas, a excepción de raros casos, no existen intrones y no hay “*splicing*” alternativo. Otra razón importante es que el enriquecimiento de mRNA es más difícil en los procariotas. Mas del 95% del mRNA en los organismos procariotas esta compuesto por RNA ribosomal. Si realizásemos una secuenciación sin enriquecer las proporciones del RNA total, obtendríamos que la mayoría de nuestros datos se pierden en rRNA, el cual no nos ofrece apenas información.

Con el enorme aumento de la capacidad de secuenciación a través de las nuevas tecnologías de secuenciación, en combinación con las nuevas técnicas de enriquecimiento de mRNA y los “*tilling arrays*”, se ha conseguido recientemente la posibilidad de investigar un transcriptoma procariota en su totalidad. Los primeros estudios de transcriptómica[20][21][22] por ultrasecuenciación han revelado cosas sorprendentes, como la gran cantidad de RNA no codificante, la estructura alternativa de los operones o las secuencias reguladoras no traducidas.

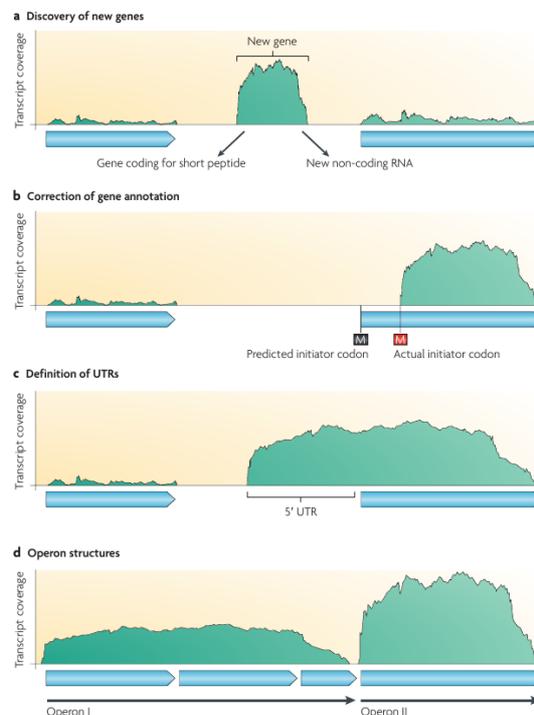


FIGURA 6: Resumen de las principales aplicaciones del RNA-Seq en la determinación de la estructura del transcriptoma en procariotas[18]

Una de las ventajas principales de la secuenciación de RNA es que no solo nos restringimos a inspeccionar la expresión de los genes entre condiciones, sino que podemos observar la topología del transcriptoma. Podemos identificar nuevos genes o zonas no codificantes, podemos corregir anotaciones en base a los alineamientos, definir regiones 5' UTR (*UnTranslate Region*) o dictaminar la estructura de los operones.

En los estudios de secuenciación de RNA (RNA-Seq), el proceso es básicamente el mismo, independientemente de la tecnología de secuenciación que vayamos a usar. Comienza con una la extracción del RNA total de la célula, para este propósito existen numerosos *kits* comerciales y protocolos. Posteriormente se hace un enriquecimiento de mRNA. Como ya hemos indicado anteriormente, el 95% de RNA en un organismo procarionta es rRNA.

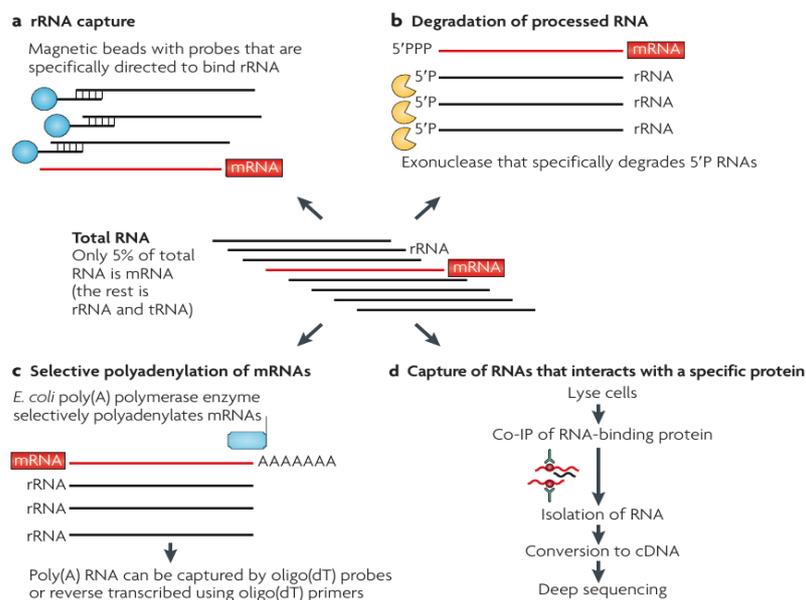


FIGURA 7: Esquema ilustrativo de los distintos métodos del enriquecimiento de mRNA presente en una muestra procarionta[18].

A diferencia de los organismos superiores, las bacterias y arqueas no contienen secuencias terminales poly-A, por las cuales podríamos diferenciar el rRNA y el mRNA. Por esta razón se han desarrollado diversos métodos.

Una vez que obtenemos el mRNA enriquecido el siguiente paso es convertirlo en cDNA mediante una RT-PCR, opcionalmente se puede amplificar esta muestra mediante una PCR con *random primers*. Este paso no siempre es necesario y depende del protocolo específico que se este llevando a cabo. Se sabe que la inclusión de este paso introduce una desviación en los datos posteriores [23]. Los protocolos de preparación de la muestra para la secuenciación son específicos de cada plataforma.

Llegados a este punto, en el cual tenemos millones de fragmentos de cDNA secuenciados, llamados *reads*, comienza una intensa labor bioinformática para dar sentido a los datos. Existe un continuo desarrollo de utilidades en este ámbito.

3. *Objetivo: Análisis del transcriptoma de Rhodococcus jostii RHA1 en medio mínimo suplementado.*

Mucho se ha discutido acerca de los mecanismo que utiliza *Rhodococcus jostii* durante la biosíntesis y acumulación de TGs. He este trabajo hemos aprovechado la técnica de RNA-Seq en donde hemos comparado el transcriptoma de *R.jostii* en condiciones óptimas para su cultivo frente a un cultivo realizado en un medio sin los nutrientes esenciales para el crecimiento, pero con una fuerte de carbono (gluconato) abundante. Trabajos anteriores [6] demostraron que *R. jostii* acumula mayor cantidad de triglicéridos cuando utiliza el gluconato como fuente de carbono. Hemos observado como se modifica la expresión de las rutas que transforman el gluconato en piruvato y como con pequeñas variaciones transcripcionales el la ruta de *Kennedy* son suficientes para un aumento representativo en la biosíntesis de los TGs. Hemos obtenido el patrón de expresión de todos los genes del cromosoma, así como de los tres plásmidos endógenos.

Con estos resultados pretendemos entender como poder implementar estos mecanismos para, en un futuro, poder maximizar la producción de triglicéridos por parte de las bacterias. Discriminar que genes son los posibles limitantes en el flujo metabólico es importante para la optimización del proceso. En un futuro, se podría utilizar la información obtenida para la implementación de una ruta de biosíntesis de TGs en bacterias que de forma nativa no producen estos ácidos grasos. De esta forma podríamos producir TGs a partir de una mayor variedad de sustratos que sirvan de medios de cultivo, e implementarlo en biorreactores industriales para una producción a gran escala.

Metodología

Condiciones de cultivo:

Primero se cultivo *R. jostii* RHA1 de forma aeróbica a 30°C en medio Streptomyces Fluka (Medio Rico, MR). Posteriormente se escoge un inóculo y se cultiva en medio con sales minerales M9 (Medio Mínimo Suplementado, MMS) de acuerdo a las especificaciones de [24]. Se uso Gluconato sódico (25 peso/volumen) como única fuente de carbono. Las condiciones limitantes de Nitrógeno en el medio mínimo suplementado (MMS) se obtuvieron añadiendo cloruro amónico 10mM para permitir la acumulación de lípidos como sugiere Alvarez HM [25].

De las células crecidas en medio rico (MR) se tomaron 25ml y se cosecharon en fase exponencial (D.O. 0.6) y en fase estacionaria. Para realizar el cultivo en MS, se escogió un inóculo de MR y se lavo con una solución de MMS, posteriormente se cultivo en 25ml de MMS durante varias horas.

Extracción y análisis de lípidos:

A partir de células congeladas en seco, se hizo una extracción mediante ácido hexa-2-isopropano (3:1 v/v). Una alícuota del extracto celular se analizó por una cromatografía de capa fina (TLC) sobre plataforma de gel de silicio (Merck) aplicando ácido n-hexa-dietileter-acético (80:20:1 v/v/v) como solvente del sistema. Se utilizó ácido oleico y trioleico como controles.

Extracción de RNA:

El RNA de cada una de las condiciones (MMS y MR) se extrajo a partir de 3ml de cultivo de cada una de ellas. El total del *pellet* aislado se agito con 6ml de RNA Protect QIAGENglass. Ambos RNAs se extrajeron mediante el procedimiento descrito anteriormente y se diferenciaron usando el mini kit RNA/DNA de QIAGEN. Para lisar el *pellet* se utilizó Zwittergent (Calbiochem) al 10%, lisozima (Sigma) en TE 15mg/ml y Proteinasa K (Roche) a 20 mg/ml. Se resuspendió el *pellet* en una solución de 280 µl de lisozima, 15 µl de Zwittergent 10% y 6 µl de proteinasa K. El RNA resultante fue tratado con Dnase I (QIAGEN)

y purificado con el mini kit RNeasy (QIAGEN) de acuerdo a las instrucciones del fabricante. Para enriquecer la muestra, primero se utilizó el kit Ambion MicrobExpress según las instrucciones del fabricante para eliminar el rRNA. Posteriormente se transformó el RNA a cDNA mediante retrotranscripción y se amplificó mediante una PCR con cebadores aleatorios (*random primers*). La calidad y cantidad de la muestra fue medida con un espectrofotómetro NanoDrop ND-1000 (NanoDrop Technology, Rockland, DE) y un Experion Automated de electroforesis usando el kit de análisis RNA Stdsens (Bio Rad).

Secuenciación:

La construcción de las librerías y la secuenciación se realizó en el servicio de secuenciación del grupo de genómica del instituto de biomedicina y biotecnología de Cantabria. La secuenciación se realizó bajo el protocolo de Illumina® con un equipo Genome Analyzer II. Se utilizaron 2 calles de una célula de flujo (*flowcell*) que reportaron 18.908.860 secuencias para la muestra de cultivo en MMS y 18.528.574 en MR. La longitud de secuenciación fue de 35 nucleótidos por cada una de las lecturas, lo que suma un total de 661Mb y 650Mb respectivamente, lo que ofrece un solapamiento teórico del $\sim 72x$.

Alineamiento frente al genoma de referencia:

Como las extracciones se realizaron a partir de cultivos de millones de células suponemos que tenemos una representación homogénea de todos los posibles estados de transcripción de las bacterias, de esta forma nuestra muestra representaría un promedio de todas las posibles condiciones dentro del cultivo. Con esta asunción y teniendo en cuenta que no tenemos replicas para nuestro estudio, con el fin de poder asignar unos valores estadísticos a nuestros resultados, así como de poder filtrar el posible ruido de fondo que podemos obtener durante el proceso de preparación de la muestra, generamos 3 subconjuntos de cada uno de los sets de secuencias en donde, aleatoriamente seleccionamos un 70% de las lecturas secuenciadas. El propósito es, que los genes que se expresan muy poco y que por cuestiones matemáticas podrían mostrar variaciones muy grandes, se puedan descartar más eficientemente. Al obtener un gen que se expresa poco podemos obtener resultados del tipo $N/0$, que daría como resultado una diferencia infinita. La cuestión aquí es intentar diferenciar si esa cantidad N es una señal robusta o simplemente ruido de fondo. Si la señal es débil debería tener una gran variación entre los 3 subconjuntos seleccionados y por lo tanto tener un valor estadístico asociado que lo permita identificar como un artefacto. Por lo tanto al final vamos a manejar 6 conjuntos de lecturas de aproximadamente 12 millones de secuencias cada uno de los conjuntos.

Para el proceso de alineamiento, se probaron varios programas: Bowtie [26], Soap2 [27] y Maq [28]. El programa que ofreció el mejor ratio de resultados/rendimiento fue Bowtie, que fue el utilizado finalmente en el proceso de alineamiento. Los genomas de referencia utilizados para el alineamiento fueron los archivos depositados en GenomeReviews (<http://www.ebi.ac.uk/GenomeReviews/files/cellular/>) CP000431_GR (Cromosoma RHA1), CP000432_GR (plásmido pRHL1), CP000433_GR (plásmido pRHL2) y CP000434_GR (plásmido pRHL3). Los parámetros variaron en los valores por defecto en los términos $-k 1$ para que solo nos reporte una coincidencia por cada secuencia alineada y $-best$ para que nos reporte siempre la mejor coincidencia. Posteriormente al alineamiento los resultados se pasaron a forma SAM [29] el cual actualmente se considera un estándar. Para poder visualizar los resultados utilizamos el software de visualización Artemis[30].

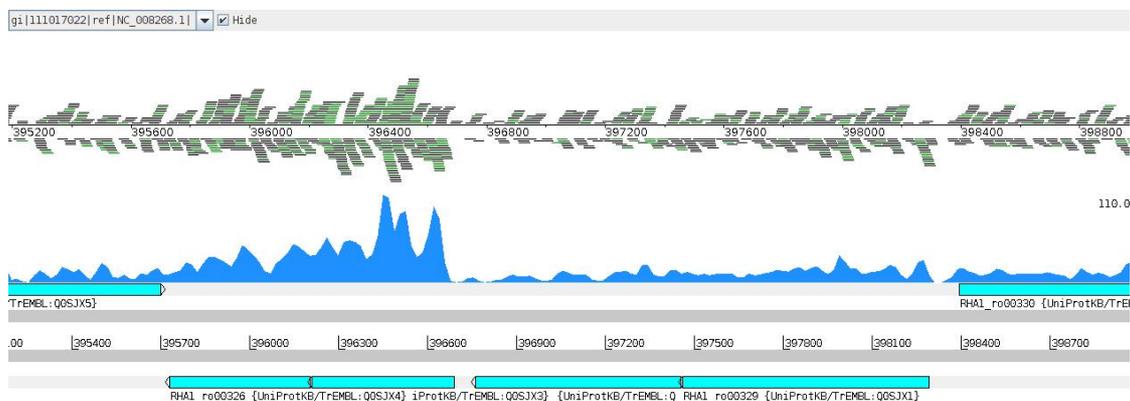


FIGURA 8: Visualización de Artemis. Ejemplo de visualización de los datos de transcriptómica de uno de los conjuntos de secuencias sobre el genoma de referencia del cromosoma de *R. jostii* RHA1. En la parte superior podemos observar cada una de las secuencias alineadas frente al cromosoma. En azul se representa la suma de las bases secuenciadas para cada una de las posiciones, componiendo entre todas ellas un perfil de expresión para el cromosoma.

Cuantificación de la expresión de cada uno de los genes:

Para la cuantificación de las lecturas por cada uno de los genes desarrollamos un *script* en el lenguaje de programación Perl. Con este programa sumamos, por cada una de las bases de DNA del gen, las bases alineadas. Como esta medida de expresión es dependiente de la longitud del gen es conveniente normalizar estos datos. Para ello vamos a aplicar una variante al método empleado en el trabajo de Mortazavi *et. al.*[31] en donde se define el parámetro RPKM (*reads per KB per million reads*). Este parámetro nos solo normaliza los genes entre sí, sino que además normaliza las muestras entre ellas utilizando el número total de lecturas como parámetro de normalización. Con es nuestro caso no vamos a utilizar este parámetro para normalizar (ver el apartado *Normalización y tratamiento estadístico*) hemos realizado una variante. En nuestro caso nosotros contamos cuentas por kilobase del gen.

$$RPK = \frac{\text{Cuentas}}{\frac{\text{longitud gen}}{1\text{kB}}}$$

Con todos los datos, construimos una matriz de expresión.

Gen	mm	mr	mm	mr	mm	mr
RHA1_ro03132	4047	1577	5081	1683	4117	1578
RHA1_ro00993	8430	3202	9474	3469	8564	3196
RHA1_ro04342	491	190	565	190	496	190
RHA1_ro02378	5724	1330	6427	1330	5727	1330
RHA1_ro01671	403	245	441	280	406	245
RHA1_ro03949	4050	4062	4644	4272	4074	4062
RHA1_ro02879	1533	751	1752	889	1602	749
RHA1_ro07232	1206	782	1385	785	1280	785
RHA1_ro02046	9034	4943	10128	5310	9222	4951
RHA1_ro01230	1332	329	1453	328	1332	328
RHA1_ro05572	15756	9644	18231	10395	16245	9660
RHA1_ro04815	962	648	1033	648	979	648
RHA1_ro06442	51861	32538	56706	34249	52682	32552
RHA1_ro04654	1803	680	1839	682	1804	682
RHA1_ro02743	1293	805	1574	875	1329	805
RHA1_ro04774	2703	1785	3277	1925	2752	1785
RHA1_ro02457	6867	1784	7685	1927	7091	1787
RHA1_ro06959	3262	707	3650	707	3294	707
RHA1_ro00880	15612	13936	17892	14991	15986	13956
RHA1_ro04674	5636	4111	6473	4379	5863	4134
RHA1_ro06610	18270	10842	19608	11447	18414	10837
RHA1_ro06177	48536	33691	53525	35184	49312	33695
RHA1_ro04718	38491	18803	43184	20219	39302	18823
RHA1_ro02929	3167	3370	3468	3485	3209	3380
RHA1_ro07275	1890	908	2302	770	2189	893

FIGURA 9: Fragmento de la matriz de expresión obtenida de los datos de transcriptómica.

Esta matriz será la base para nuestro tratamiento estadístico.

Normalización y tratamiento estadístico:

El primero de los procesos estadísticos que se debe abordar en un estudio como este es la normalización de los datos de las distintas condiciones. Básicamente hay tres formas de tratar este problema. Cada uno de ellos se basa en unas premisas biológicas.

Si asumimos que la cantidad total de RNA por célula es independiente e invariante de las condiciones de cultivo, podemos asumir que el numero total de lecturas en cada una de las condiciones, ha de ser siempre el mismo. Con esta premisa, podemos realizar una normalización dividiendo cada una de las condiciones por la suma total de la expresión de los genes o por la relación entre las sumas de cada una de las condiciones. Este es el método empleado cuando tratamos las expresiones como RPKM[31]. Este tipo de normalización da buenos resultados en estudio eucariotas en donde el numero de genes es muy grande y el numero de genes invariantes frente a los modificados es pequeño. En el caso de procariotas, afirmar que la cantidad total de RNA es constante es muy arriesgado. En un estudio como el nuestro, en donde pasamos de una condición en donde el crecimiento es óptimo a otra en donde la bacteria esta estresada no nos ha parecido los mas indicado hacer este tipo de normalización.

Desde otro punto de vista se puede imponer un control interno, es decir, podemos escoger un gen como referencia y asumir que es invariante frente a las distintas condiciones. Esto es un mecanismo que se aplica normalmente en numerosas técnicas molecular. Es muy común usar genes ribosómicos en PRC cuantitativas o proteínas como la beta-actina en los Western-Blot. En primer lugar, es muy difícil encontrar un gen que nos sirva para todas las condiciones, como igual de difícil es demostrar que es invariante. De nuevo, como en el caso anterior, es plausible hacerlo en organismo superiores, pero en bacterias es difícil establecer un gen de referencia. Aún así sería más fácil realizarlo en organismo modelo como *E. Coli*, o *Bacillus Subtilis* los cuales están mucho mejor caracterizados. En nuestro caso, con *R. jostii* un organismo poco estudiado y además teniendo unas condiciones tan distintas, de nuevo hemos descartado esta aproximación.

Por último, se puede realizar una aproximación basándonos en parámetros estadísticos de las muestras. Como siempre, debemos de basarnos en una hipótesis. En este caso, supongamos que los resultados responden a una distribución estadística concreta (distribución Gaussiana, binomial, Loretziana etc ..). Normalmente este tipo de eventos responden a distribuciones Gaussianas o Normales. Toda distribución estadística tiene unos parámetros que la definen (media, mediana, desviación estándar, etc...). Para este tipo de normalizaciones lo que asumimos es que alguno de estos parámetros es invariable y por lo tanto los podemos tomar como referencia para normalizar. Durante mucho tiempo, se tomó el valor estadístico de la mediana como referencia para normalizar. Incluso hay métodos que utilizan la mediana y la media al mismo tiempo. La mediana es un dato más fiable distribuciones con un número de elemento muy grande, ya que es insensible a los valores extremos, pero es altamente sensible a distribuciones bimodales. Como lo lógico es pensar que nuestra distribución de valores es Gaussiana la media y la mediana deben de ser muy parecidas y en cualquier caso la mediana será más acertada. En cualquier caso, lo que estamos asumiendo aquí, es que los genes que se activan se compensan con los que se reprimen o bien que el número de genes que va a variar de una condición a otra es despreciable frente al total. Nosotros nos hemos guiado por una normalización por cuantiles. Este es otro parámetro de posición al igual que la mediana. Básicamente asumimos que las distribuciones de los cuantiles centrales son invariantes. Hemos escogido este método basándonos en el trabajo de Bullard *et. al.* [32] que recomiendan este método para los experimentos de RNA-Seq.

Para la normalización utilizamos el entorno estadístico R, que es de libre distribución junto con el paquete de análisis *limma* que contiene los programas necesarios para la normalización. Después realizamos el análisis de expresión diferencial con el paquete DESeq

[33]. De esta forma obtenemos una lista con la expresión de los genes normalizados, los valores de variación y los valores de confianza en base el *p-value* y *p-value ajustado*.

id	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldChange	pval	padj	resVarA	resVarB
2711	RHA1_ro03916	20554.2505226317	3983.33425588669	37125.1667893747	9.32012339524139	3.2203490579418	0	0.0.0893192696473622	0.102271786779777
2415	RHA1_ro06364	18412.9975674604	3764.96614297444	33061.0289919464	8.78122876447095	3.13442283137486	0	0.0.0837986062641786	0.0983851043836088
7056	RHA1_ro03318	46480.5828138274	9759.19513377917	83201.9704938757	8.52549512058546	3.09178362149998	0	0.0.0517321909842633	0.0474674555392368
909	RHA1_ro01380	39489.8611091201	8321.27369863946	70658.4485196008	8.49130206246594	3.08598579465364	0	0.0.0330105067818655	0.031651168999663
6184	RHA1_ro03321	34666.2200198113	7796.18253188157	61536.257507741	7.89312682920079	2.9805969310809	0	0.0.0172162925417904	0.0170989065722609
2860	RHA1_ro03319	28600.9576858815	6721.17766281326	50480.7377089497	7.51069830935254	2.90894704896519	0	0.0.0194796494966074	0.0203692164213219
2441	RHA1_ro05024	53540.6425767476	13096.5526135211	93984.732539974	7.17629557284737	2.84323931231674	0	0.0.0030426662667826	0.02069076347599105
4392	RHA1_ro03811	63661.4259944574	15735.9520859198	111586.899902995	7.09120740160618	2.82603129324571	0	0.0.00695399180389286	0.0085959304835322
2699	RHA1_ro03320	22102.4623967254	5466.47363893089	38738.4510745199	7.0865322516443	2.82508389837513	0	0.0.104776770174772	0.117283336331383
4415	RHA1_ro04433	7448.04995514711	1873.0554480248	13023.0443614918	6.9528340309063	2.79760115178425	0	0.0.0213957967008494	0.0216274249286021
504	RHA1_ro01994	57958.4595228997	14990.9079867522	100926.011058967	6.73248152467867	2.75113836867889	0	0.0.0722311298994151	0.0623574753454101
1876	RHA1_ro04379	5283.42336358115	1381.70224011643	9185.14448704587	6.64770181328783	2.73285670095616	0	0.0.0286262865490547	0.0354757996967353
5900	RHA1_ro02813	47207.7912742384	12401.3202820348	82014.262266442	6.61334925647006	2.72538109331036	0	0.0.0071733836698333	0.00657011386519035
6050	RHA1_ro04380	9012.89625996058	2377.88739054155	15647.9091213796	6.58059300184789	2.71621759646457	0	0.0.0046647725085259	0.00591886440743995
2447	RHA1_ro02930	23773.2152935626	6281.38979764938	41265.0407894798	6.5694125215598	2.71576436139571	0	0.0.0439901017304237	0.0482848604181318
4117	RHA1_ro06986	3387.548015759605	902.146878072208	5872.9491534399	6.50997004610797	2.70265090522948	0	0.0.0729515282142355	0.0749531456523086
5807	RHA1_ro05788	11077.8274962350	3067.71056898088	19087.9444230891	6.22221164518763	2.63742746820395	0	0.0.0860533605081893	0.106967742711195
6347	RHA1_ro03236	7586.63442594949	2198.62050679149	12974.6483451075	5.9012677451092	2.56102492320697	0	0.0.273107355212893	0.347131893123588
2805	RHA1_ro06098	18617.1595937828	4878.10628585495	28356.2129219106	5.81293514645855	2.53927177858434	0	0.0.0354258733150028	0.041970797823379
3001	RHA1_ro05414	11773.4146489822	3498.08094053694	20047.7483573874	5.72943258474926	2.51839228852800	0	0.0.0483218754334077	0.058254515386238
2138	RHA1_ro01635	23344.0502748142	6988.38501448973	39699.7155347387	5.68081401531616	2.50609787122933	0	0.0.0918883075199718	0.101140550742926
5352	RHA1_ro03209	13016.0815682901	3905.85762784814	22126.3055087321	5.66490323430509	2.50205131269531	0	0.0.049103568680801	0.0600724556649167
7115	RHA1_ro01301	13868.6347644795	4262.10081985876	23475.1687091002	5.50788676788996	2.46149890059722	0	0.0.345754411924093	0.420119082079949
6889	RHA1_ro02029	4161.45500868245	1295.10257689287	7027.80744047202	5.42644850366423	2.44000828333882	0	0.0.027902932071912	0.0309878039131214
1836	RHA1_ro03812	1841.90382447431	582.006089312405	3101.80155963621	5.32950018323821	2.41400023899285	0	0.0.0922578320577976	0.081100993893909
3877	RHA1_ro00480	5353.27286959669	1694.44928800708	9012.0964511863	5.31859909886643	2.4110462392277	0	0.0.0213075910380994	0.025922802720246
3164	RHA1_ro05599	14470.0343044238	4583.88430597888	24356.1843028687	5.31343783504753	2.40964559789025	0	0.0.0334331545514019	0.0402306288951071
4766	RHA1_ro03322	6518.97292242685	2069.54042422544	10968.4054206283	5.2999232545715	2.40597146891156	0	0.0.10687481188157	0.13479657905636
1434	RHA1_ro03276	20428.4258821755	6503.76870401144	34353.0830603395	5.28202717897255	2.4010917249834	0	0.0.0859173613330312	0.097442230357181
5192	RHA1_ro04432	12784.8292356119	4092.8915551236	21476.7869161002	5.2473334807058	2.39158448139241	0	0.0.0948890797166764	0.0115788508993843
5091	RHA1_ro04918	164692.333855655	53029.0165556358	276355.651155674	5.21140441791399	2.38167221582871	0	0.0.038071350581723	0.0243570967134033
239	RHA1_ro06097	32245.72209141	10508.7273544723	53982.7168283476	5.13694142092035	2.36090962201472	0	0.0.0115908312430714	0.0117429786160743
2742	RHA1_ro06023	19947.4536631952	6510.522670807	33384.3846556097	5.12775799175681	2.35832811452857	0	0.0.0220443457854638	0.025113331406209
836	RHA1_ro06594	17120.1048044752	5611.4610967079	28628.7486242424	5.10183503635294	2.35101625133491	0	0.0.30999332827561	0.363768996042645

FIGURA 10: Fragmento de la tabla de expresiones del experimento. Columnas por orden de izquierda a derecha: **id**: locus_tag del gen, **baseMean**: Media de expresión del gen entre condiciones, **baseMeanA**: Media de expresión en la condición A, **baseMeanB**: media de expresión en la condición B, **foldChange**: Variación entre las condiciones, **log2foldChange**: Variación en logaritmo en base 2, **pval**: valor del p-value, **padj**: valor de p-value ajustado, **resVarA**: valor de los residuos en la condición A, **resVarB**: valor de los residuos en la condición B

Análisis funcional:

Para el análisis de los resultados nos hemos basado en las rutas metabólicas anotadas en la base de datos KEGG [34]. Para realizar una representación gráfica de los resultados hemos tomado los esquemas de las rutas metabólicas como base y hemos añadido los datos de expresión de los genes que codifican las isoenzimas. En el genoma de *Rhodococcus* podemos encontrar una gran cantidad de enzimas que realizan la misma reacción catalítica (isoenzimas). Ante el desconocimiento de si actúan todas al mismo tiempo y en todas las posibles rutas o no, debemos de entender que todas ellas son igual de importantes. Para la representación de las expresiones hemos desarrollado un método que nos representa las diferencias de expresión entre las distintas isoenzimas y al mismo tiempo la variación de cada una de ellas entre las condiciones.

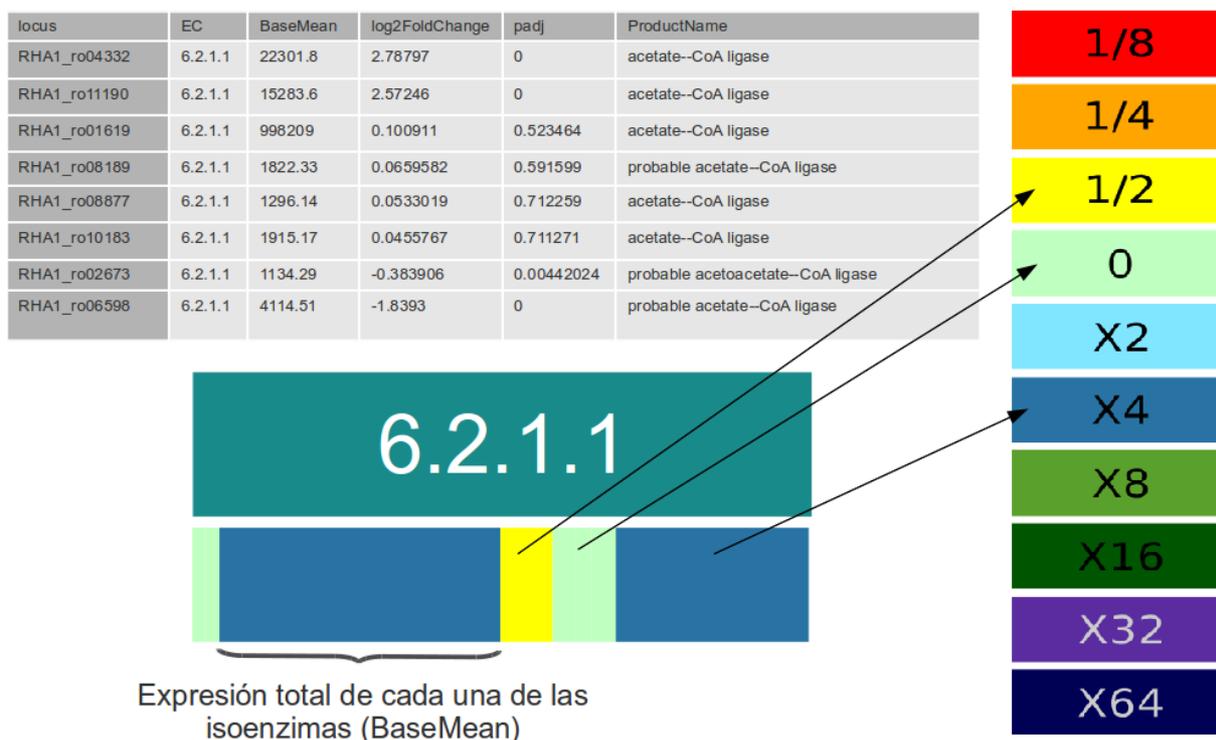


FIGURA 11: Ejemplo de representación de un conjunto de isoenzimas. Arriba una tabla con los datos de las isoenzimas con código enzimático (EC) 6.2.1.1. A la derecha la leyenda con el código de colores relativa a las diferencia de expresión. Abajo un ejemplo de la representación gráfica, donde podemos ver el EC y la expresión de cada una de las isoenzimas, donde la anchura de las franjas representa la expresión media relativa entre las isoenzimas y el color la variación entre las condiciones para cada una de ellas.

Para construir automáticamente estas figuras hemos escrito un programa en Perl y utilizando el lenguaje de gráficos vectoriales SVG (*standard vector graphics*). Utilizando el valor medio (*baseMean*) de cada isoenzima, calculamos el valor relativo que debe ocupar en el gráfico y lo coloreamos en función del valor de cambio (*foldChange*).

Resultados

Como control de que nuestros cultivos son correctos, hemos comprobado la acumulación de triglicéridos en cada uno de ellos mediante la técnica descrita en el apartado Metodología.

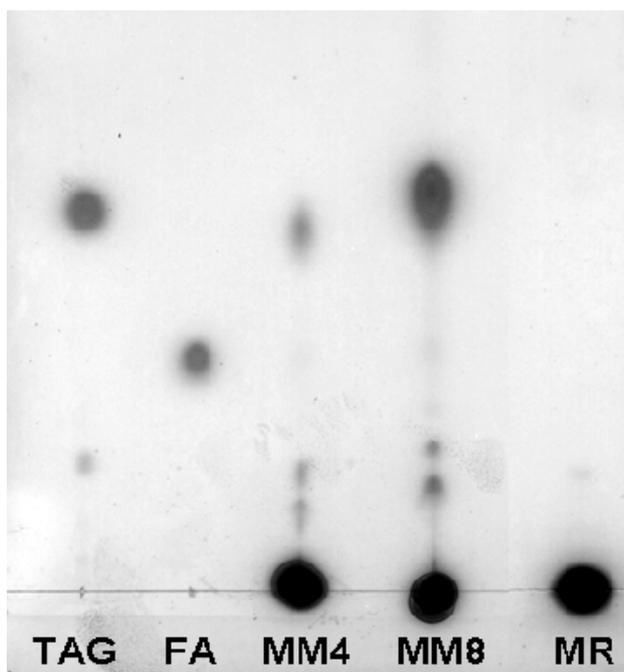


Figura 12: TLC de los cultivos realizados en medio rico MR y en distintos medios mínimos suplementados (MM4 y MM8). Como controles ácidos grasos (FA) y triglicéridos (TAG)

Realizamos los 6 alineamientos (3 para cada una de las condiciones) sobre las secuencias de referencia, obteniendo las siguientes estadísticas.

	MR	MMS
Cromosoma	77,29%	78,35%
Plásmido pRHL1	0,78%	0,80%
Plásmido pRHL2	0,28%	0,29%
Plásmido pRHL1	0,13%	0,13%
No alineados	21,52%	20,43%

Tabla 1: Porcentajes de las secuencias alineadas.

Resultados del análisis estadístico:

Una vez realizado el alineamiento y construido la matriz de expresión normalizamos los datos por cuantiles según el protocolo establecido

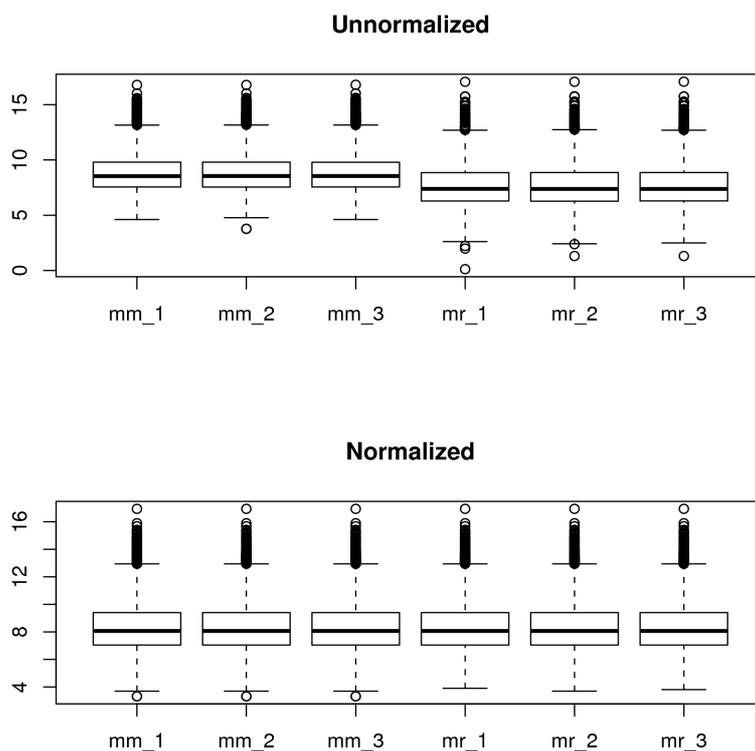


FIGURA 13: Representación en de los datos de expresión de los 6 conjuntos de secuencias; 3 conjuntos de medio rico (mr) y 3 de medio mínimo suplementado (mm). Arriba, datos sin normalizar en escala logarítmica. Abajo, datos normalizados por cuantiles, representados igualmente en escala logarítmica

Como podemos observar, las medias en los datos crudos no coinciden ni tampoco las medianas, lo que indica que se debe normalizar los datos porque existe una desviación sistemática de los datos entre los distintos experimentos. De la misma forma, vemos que la utilización de pseudo-replicas genera una cierta dispersión entre los valores de los conjuntos de datos. Este era nuestro objetivo, cuanto mayor es la robustez de la señal, menor sera la dispersión que podamos ejercer sobre ella.

Mediante las siguientes figuras podemos estudiar si nuestros datos son coherentes con los datos depositados en la literatura.

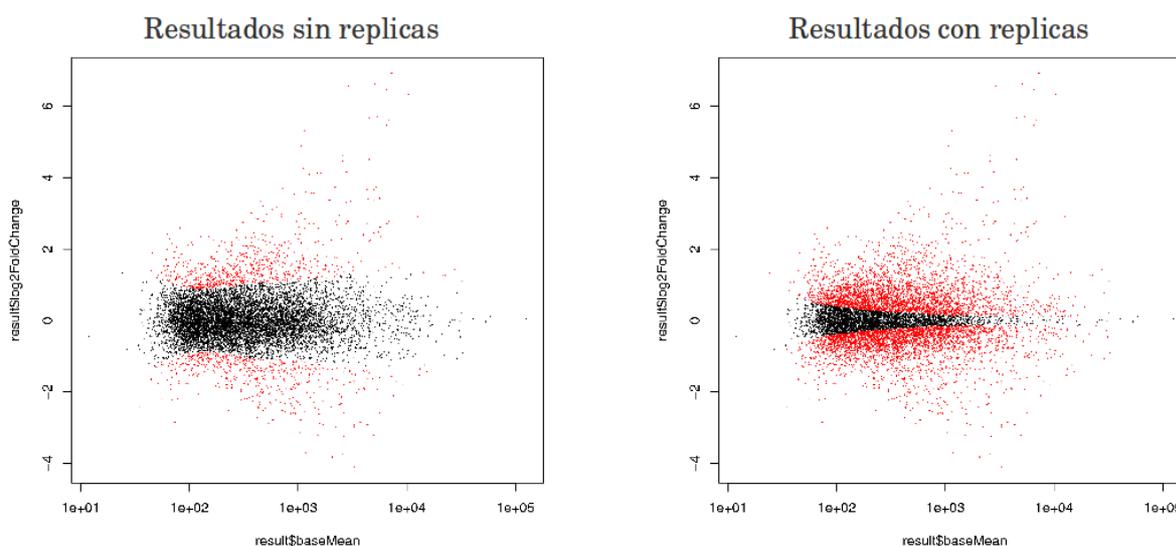


FIGURA 14: Comparación entre el tratamiento con pseudo-replicas y sin replicas. Representación la expresión media del gen (eje X) frente a su variación entre condiciones(eje Y) en color rojo aquellos valores con un $p\text{-value} < 0.001$.

Aunque el paquete de datos DESeq esta preparado para funcionar sin replicas, aquí podemos observar las diferencias. Bien es cierto que no tenemos replicas verdaderas sino pseudo-replicas pero se puede observar que los datos que tienen una variación destacable tienen buenos valores de confianza. El programa DESeq asigna a cada valor una distribución binomial negativa (BN). Una distribución BN solo tiene dos parámetros, la media y la varianza. En el caso de no tener replicas, la varianza se asigna proporcional a la media y de esta forma se puede obtener un parámetro estadístico de confianza. Como podemos observar en la figura 14 existe una tendencia en cuanto a que, cuanto mayor es la expresión mejor es su valor de confianza (menor $p\text{-value}$). Pensamos que es mejor el uso de pseudo-replicas ya que nosotros hemos realizado anteriormente una normalización por el tamaño del gen y por lo tanto pueden existir señales de baja intensidad pero robustas que proporcionan una variación pequeña, que no depende de la media. De esta forma somos capaces de rescatar mas información.

Hemos utilizado como control interno, la presunción de que solo un pequeño porcentaje de los genes tiene una variación destacable, mientras que la mayoría de ellos no deberían de cambiar.

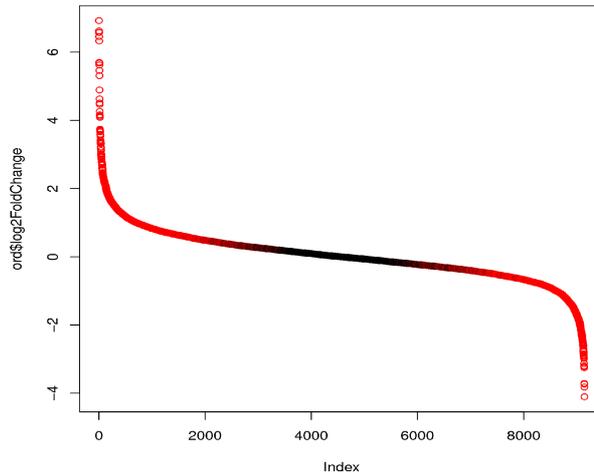


FIGURA 15: Para representar los datos, se han ordenado los genes por su variación entre condiciones y se representado frente a ella misma. En rojo los valores con $p\text{-value} < 0.001$

Como podemos ver, la mayoría de los genes (parte central de la figura) no tienen cambios destacados mientras que existe un conjunto de genes (parte izquierda y derecha de la gráfica) en el cual se encuentran los genes con activaciones o represiones destacables. Quizá el conjunto de genes representativos sea mas grande de lo normal, pero tenemos que recordar que nos encontramos ante dos condiciones muy distintas es las cuales hemos cambiado varios parámetros (fuente de carbono, concentraciones de varias sales minerales, etc ...).

Por ultimo, podemos representar la variación entre las condiciones, enfrentando a cada gen con sí mismo en cada condición.

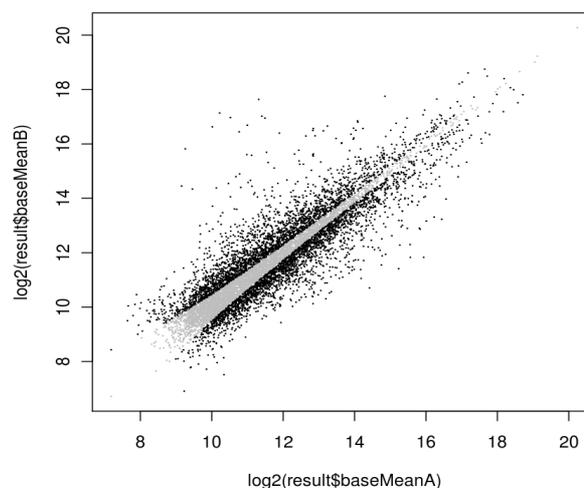


FIGURA 16: Representación de la cada la media de expresión de cada gen en medio rico (eje X) frente a la media en medio mínimo suplementado (eje Y). En negro los valores con $p\text{-value} < 0.001$

En la figura 16 vemos como los genes que se encuentran en la diagonal son los que no tienen variaciones representativas. Los valores que se desvían de la diagonal hacia el cuadrante superior son activaciones y los que se desplazan hacia el cuadrante inferior represiones. De igual forma que en la gráfica anterior podemos ver que es un pequeño porcentaje el que muestra cambios representativos.

Además hemos estudiado la expresión de los plásmidos. Queríamos comprobar si algunos de los plásmidos varían su expresión total en función de las condiciones. En la siguiente gráfica, se muestran las expresiones totales de los plásmidos, normalizadas frente a la expresión total del cromosoma.

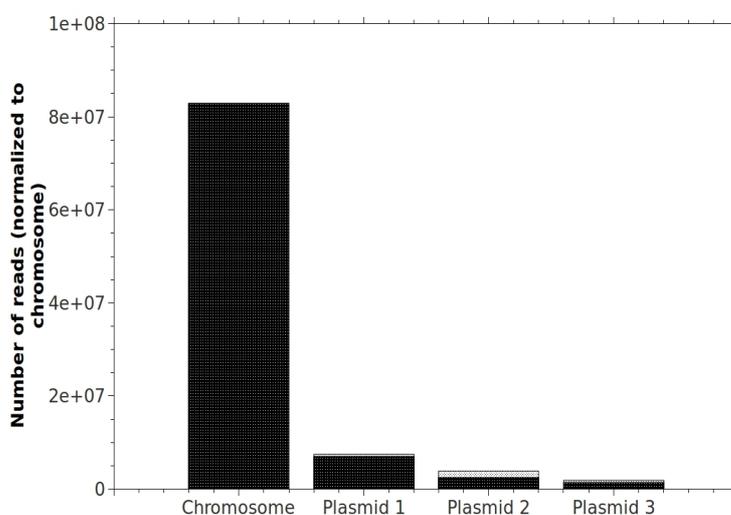


FIGURA 17: Representación de la expresión total de los genes contenidos en los plásmidos referenciados a la expresión total del cromosoma. En blanco sombreado la expresión en el medio mínimo suplementado, en negro la expresión en el medio rico.

Se puede apreciar que no existen diferencias significativas, lo que indica, que ninguno de los plásmidos está implicado, en exclusiva, en ninguno de los procesos de producción o almacenamiento de triglicéridos.

Alineamiento y mapas de expresión:

Una vez que obtenemos los datos y hemos comprobado que son coherentes con lo esperado, podemos representar la expresión diferencial del genoma de *Rhodococcus jostii*.

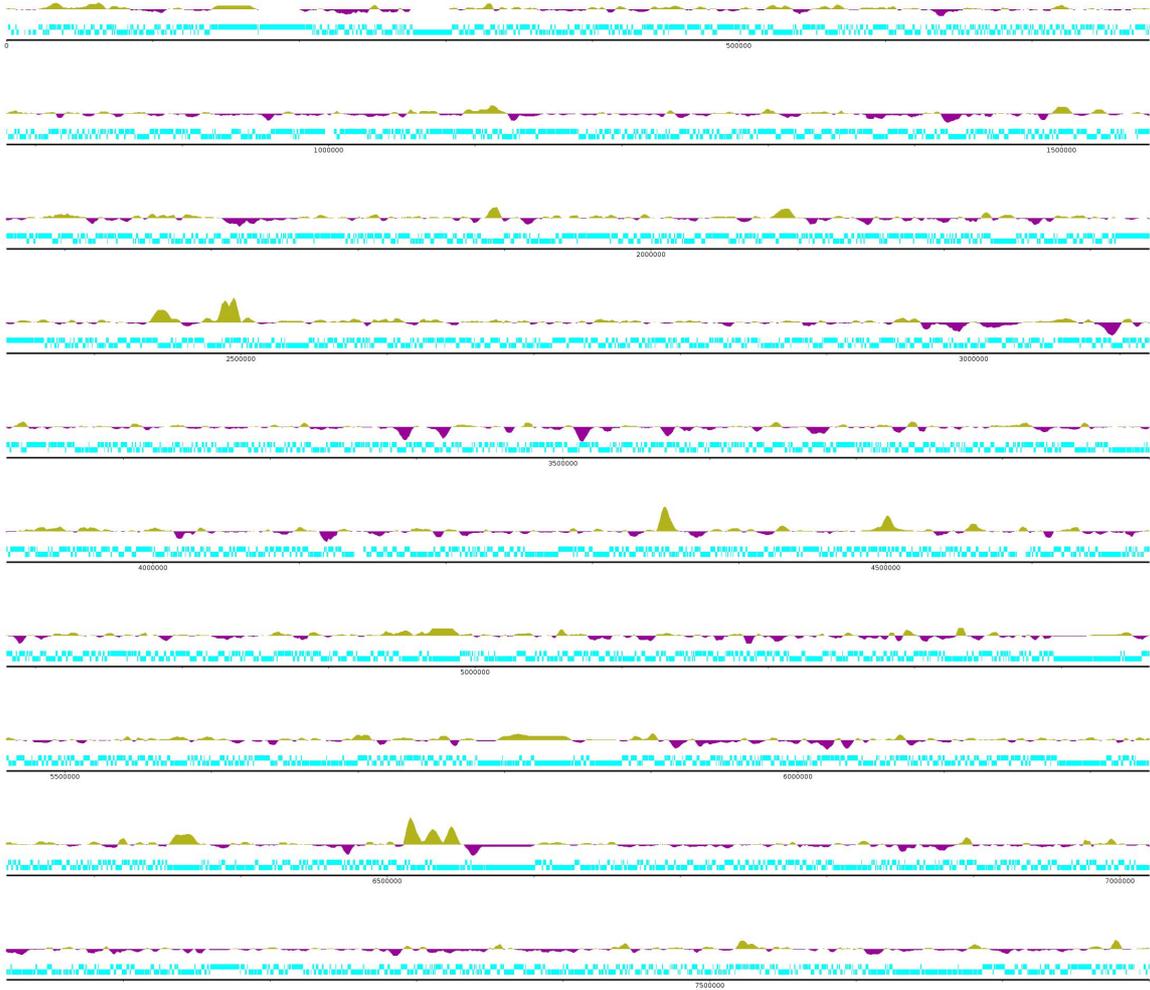


FIGURA 18: Mapa de expresión diferencial del cromosoma de *R. jostii*. En verde se observan los elementos activados. En morado los elementos reprimidos. Bajo las expresiones en azul claro cada segmento representa un gen y si esta en la hebra directa o la complementaria. Resaltar que el cromosoma de *R. jostii* es lineal.

A primera vista se observa como se expresan los genes agrupados en operones y como las variaciones están distribuidas homogéneamente a lo largo del cromosoma.

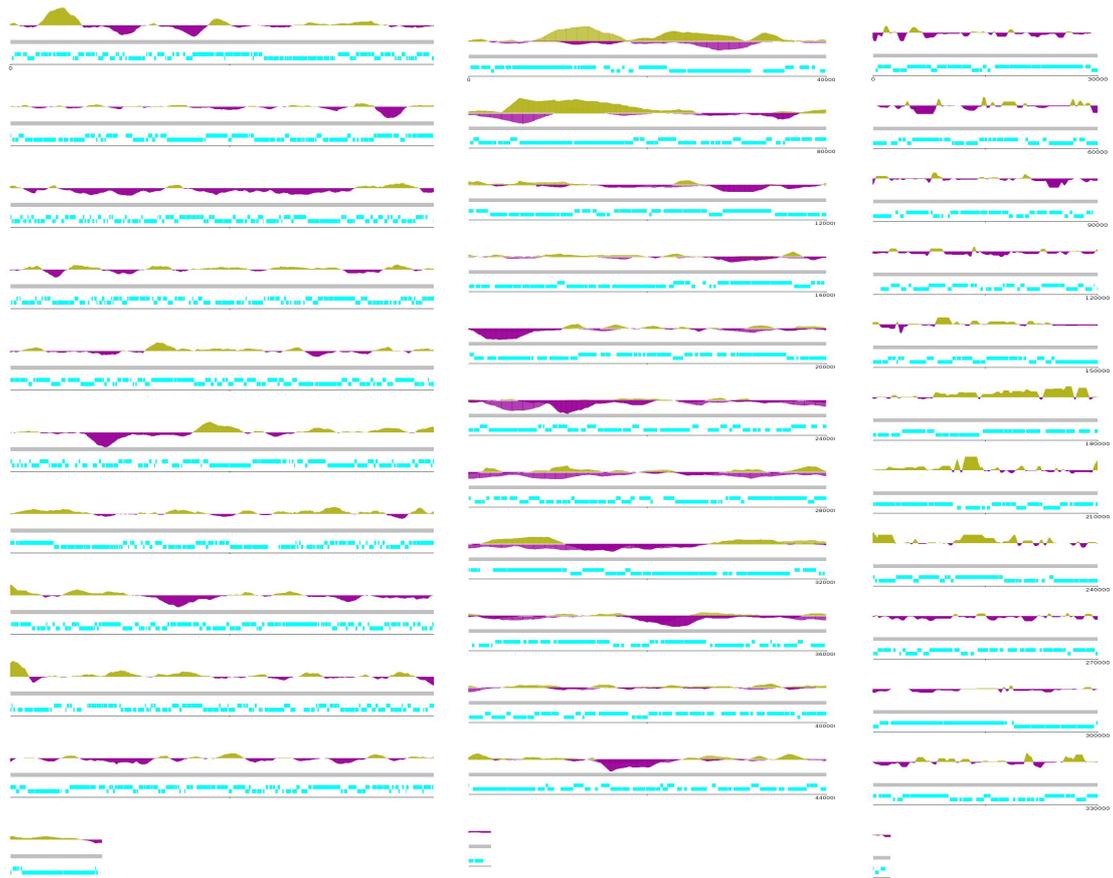


FIGURA 19: Mapa de expresión diferencial del de los plásmidos endógenos de *R. jostii*. En verde se observan los elementos activados. En morado los elemento reprimidos. Bajo las expresiones en azul claro cada segmento representa un gen y si esta en la hebra directa o la complementaria. La anotación de los plásmidos, al igual que en el caso del cromosoma, indica que son lineales. De izquierda a derecha vemos pRHL1 (1,1Mb), pRHL2(442 Kb) y pRHL3 (332 Kb),

Vemos igualmente que la expresión esta distribuida homogéneamente a lo largo de los plásmidos. Esto corrobora la teoría anterior de que ninguno de ellos parece especializado en la adaptación a un medio en concreto, ya que los tres se ven alterados, tanto en activación como en represión. Estas ilustraciones fueron realizadas con el modulo DNAPlotter [35] del programa Artemis.

Análisis funcional

A continuación expondremos algunos de las rutas metabólicas mas interesantes para el estudio basadas en la información contenida en KEGG. Para ello hemos construido unos gráficos basados en los esquemas oficiales de la base de datos y hemos superpuesto la representación que se explico anteriormente en el apartado de metodología.

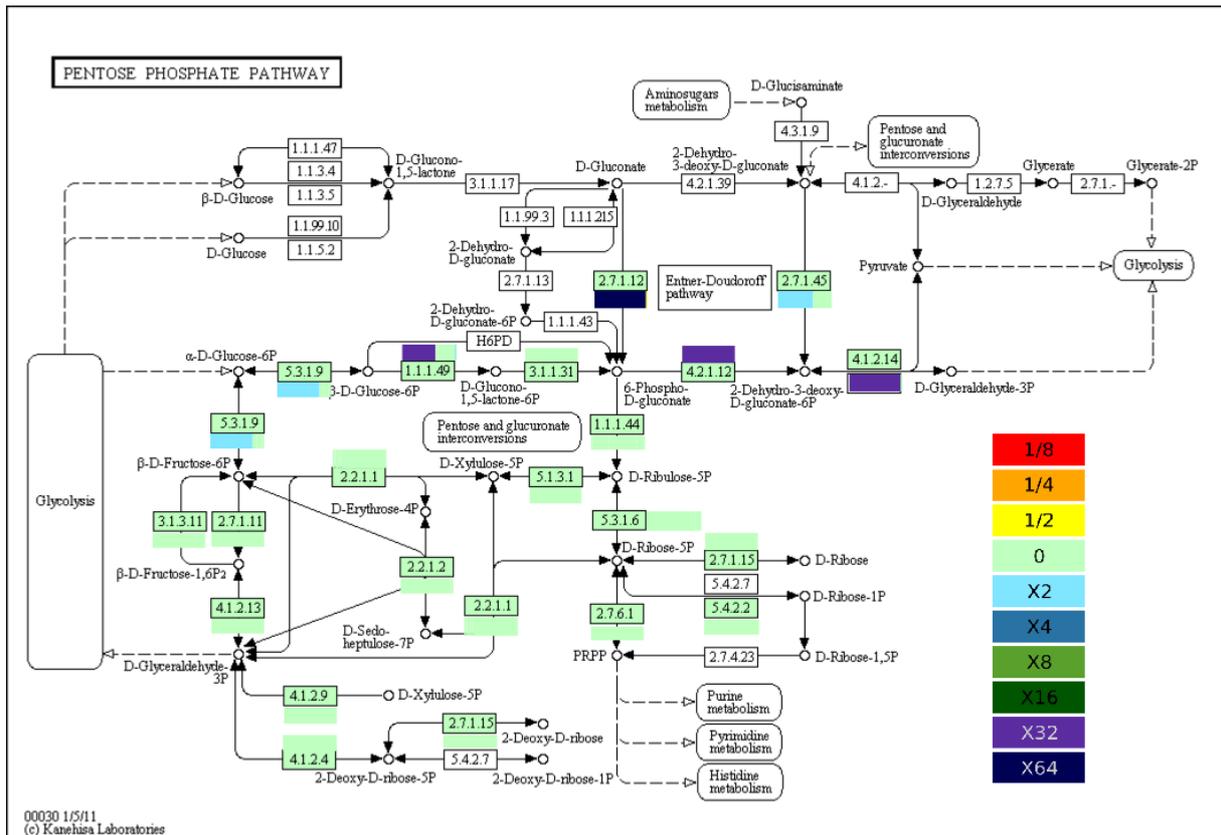


FIGURA 20: Ruta de las pentosas fosfato. Las enzimas con fondo en color verde son aquellas para las cuales existe una anotación en el genoma de *R. jostii* mientras que las que contienen un fondo blanco, o bien no se ha anotado correctamente el genoma o realmente el organismo no contiene estas funciones.

Podemos observar una activación del módulo de Entner-Doudoroff, que esta íntimamente ligado al metabolismo del gluconato. Debemos de atenernos al dato inicial que hemos intercambiado las fuentes de carbono; glucosa por gluconato. Tenemos resultados de experimentos de proteómica que indican que esta activación se conserva aun cuando la fuente de carbono en el medio mínimo suplementado es glucosa. De cualquier forma la ruta desde la α -D-Glucosa-6P hasta el D-Gliceraldehído-3P esta fuertemente activada.

Siguiendo el curso de la biosíntesis de los triglicéridos, vamos a revisar la ruta de la Glucólisis/Gluconeogénesis, para seguir investigando la fuente de producción del D-Gliceraldehído-3P precursor esencial en el proceso.

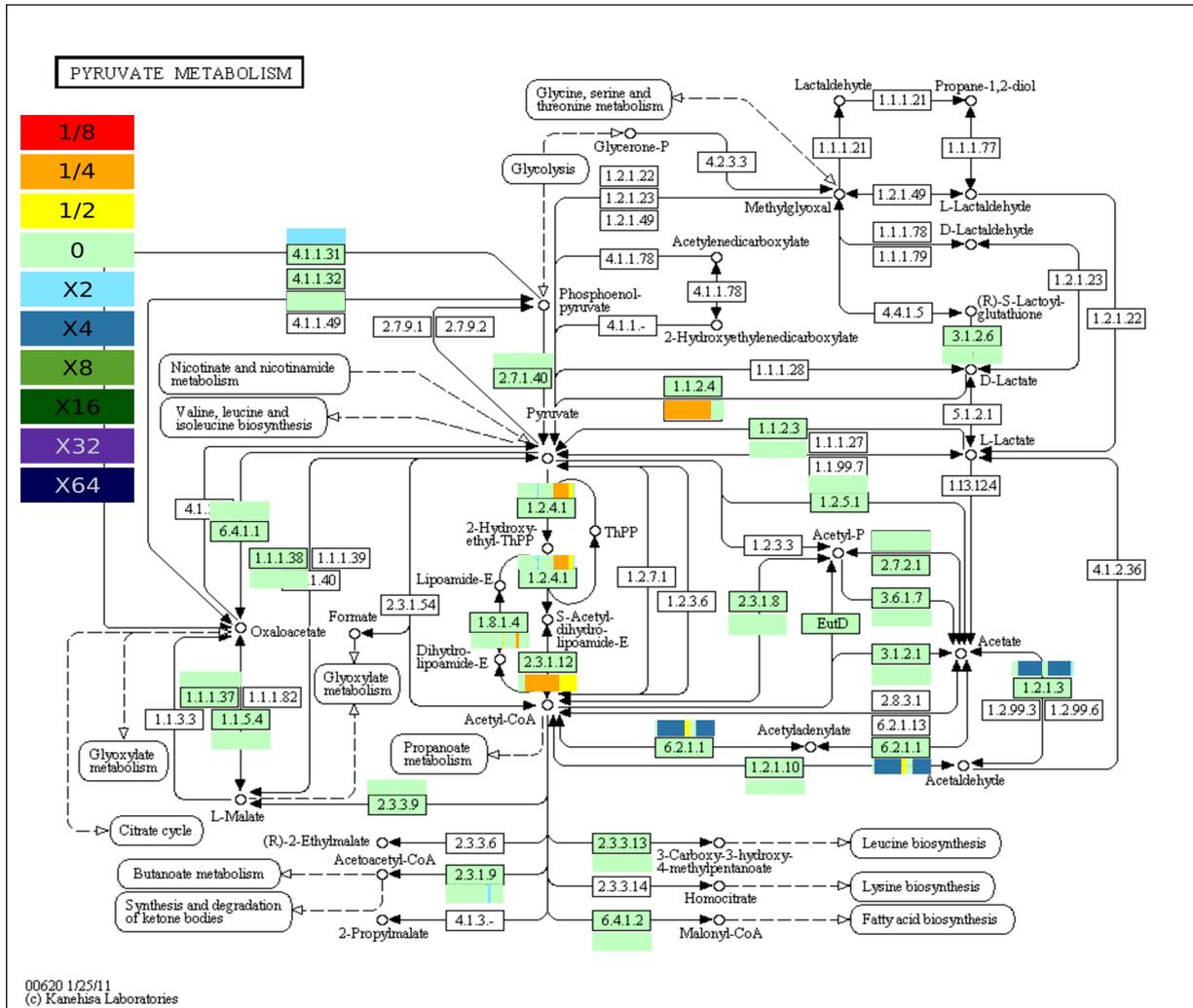


FIGURA 22: Metabolismo del piruvato. En esta ruta se encuentran presentes las posibles rutas de la síntesis de acetil-CoA a partir del piruvato.

Nuevamente vemos la activación de las rutas de conversión del piruvato en acetil-CoA por las vías en donde está implicado el acetato. También están repetidos los datos de la oxidación directa del piruvato hacia acetil-CoA donde vemos represiones en los genes implicados. Además vemos que las vías del piruvato hacia el Oxalacetato y posteriormente Malato para terminar igualmente en acetil-CoA, no presentan variaciones destacables. Por último, la conversión de acetil-CoA en malonil-CoA (principal precursor de los ácidos grasos y por ende de los triglicéridos) tampoco ha variado.

Como indicamos en la introducción la biosíntesis de los triglicéridos requiere una gran cantidad de poder reductor. La principal fuente de producción es la oxidación fosforilativa.

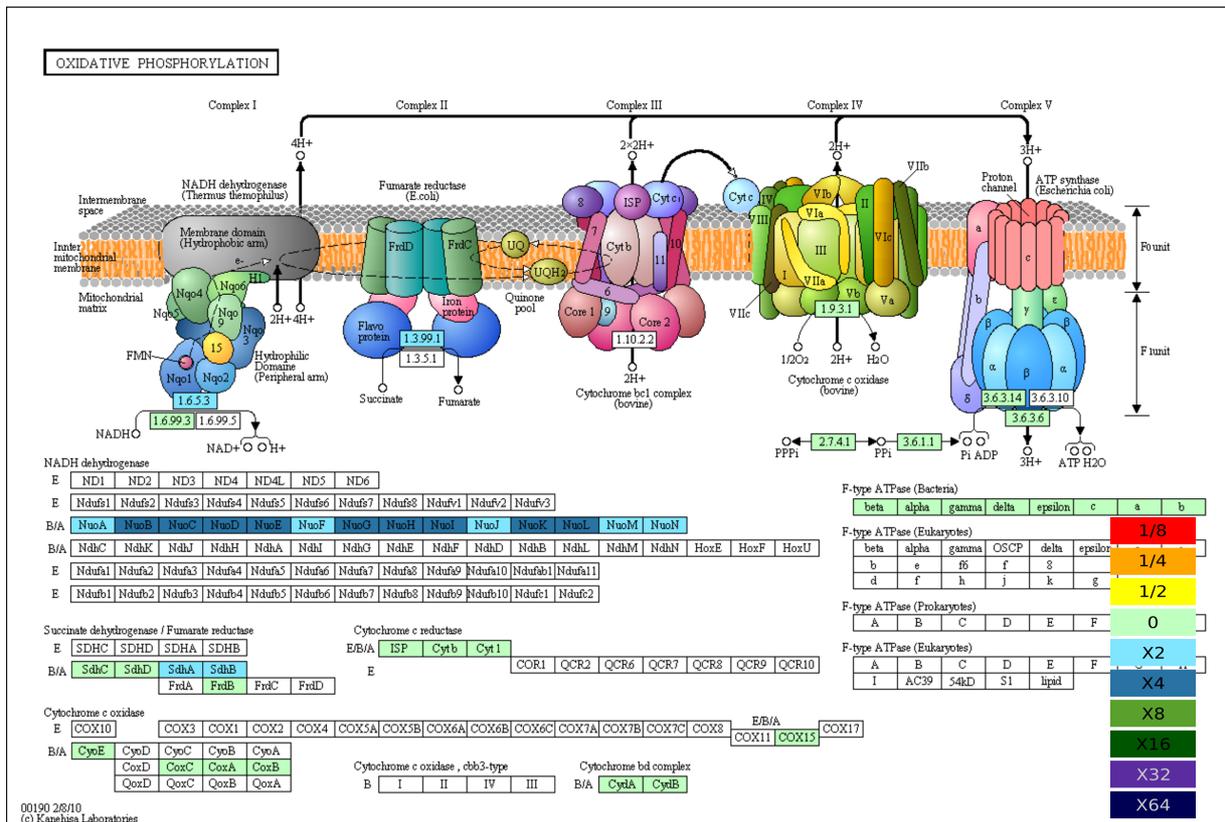


FIGURA 23: Esquema ilustrativo de los complejos existentes en las membranas para la producción de NADH y ATP. Esta es la principal fuente energía para la célula.

Como se podía esperar observamos una gran activación de las deshidrogenasas de NADH. He aquí la fuente de poder reductor que necesitamos para lo triglicéridos. También podemos observar como la F-ATPasa no esta alterada transcripcionalmente, lo que podría indicar que la cantidad de ATP en el proceso no es un factor limitante.

Para ver los últimos pasos en la formación de los triglicéridos debemos examinar la ruta de los glicerolípidos.

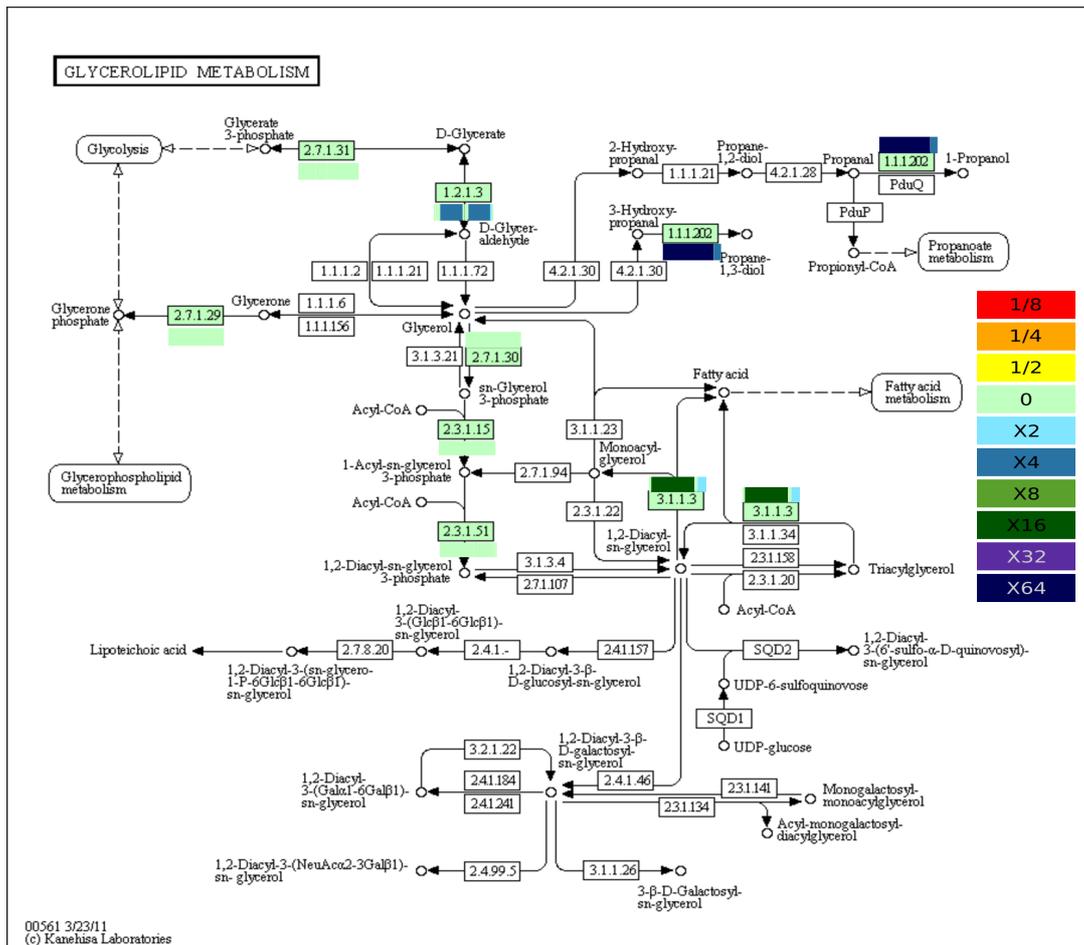


FIGURA 24: Ruta de los Glicerolípidos. Vía que representa, entre otras cosas, la incorporación del glicerol a los ácidos grasos y los triglicéridos.

Es esta vía se reconocen algunos de los genes con mayor activación. Aún no esta claro el motivo de la activación tan drástica de la enzima 1.1.1.202 encargada de la generación de propanol. Quizá se deba a la ventaja de producir propionil-CoA que puede oxidarse a malonil-CoA, metabolito secundario necesario en la formación de ácidos grasos y triglicéridos.

Como resumen final hemos construido dos gráficos en los que se muestra un resumen de los resultados. Por una parte tenemos las rutas de formación del acetil-CoA como eje de la figura (Fig 25), en la segunda (Fig 26) representamos el flujo metabólico desde la fuente de carbono hacia la formación de los triglicéridos.

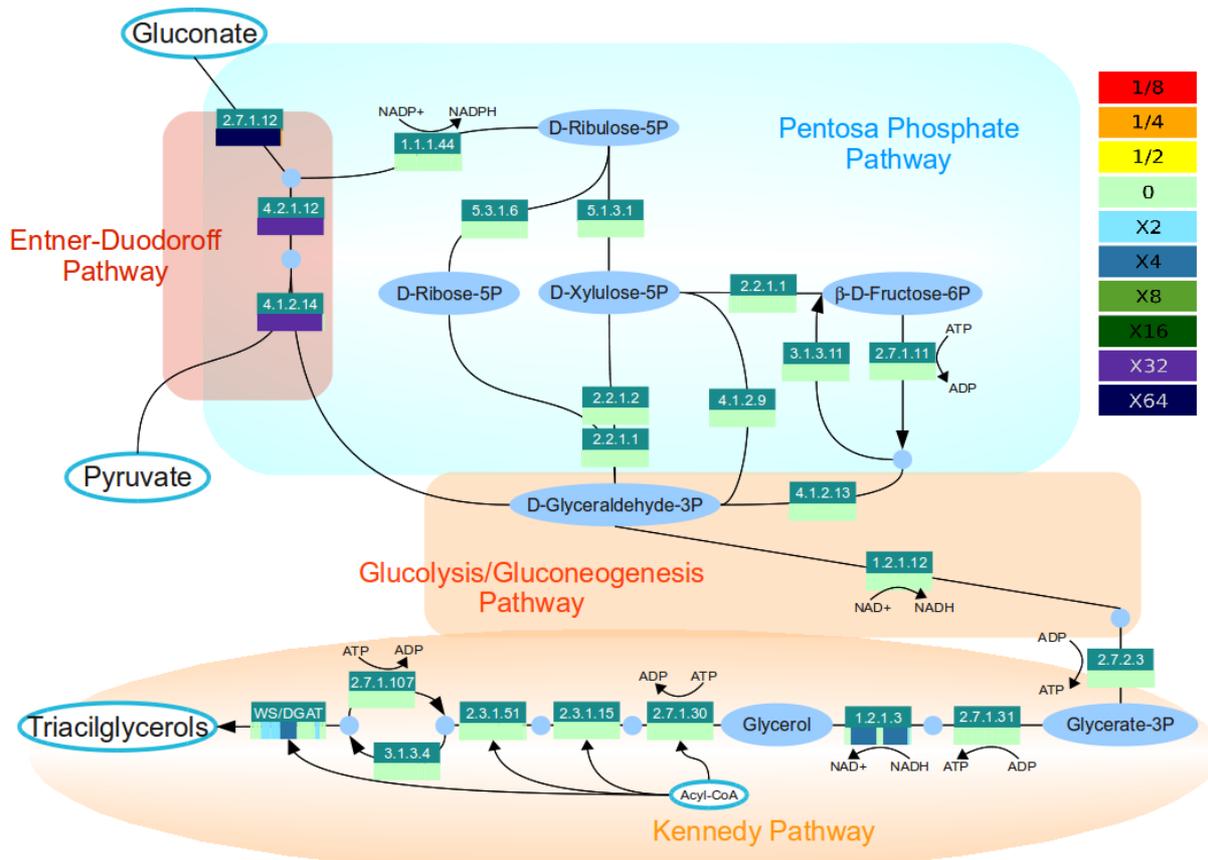


FIGURA 26: Resumen de la ruta directa entre la fuente de carbono y la formación de los triglicéridos a través de las pentosas fosfato, la glicólisis y la ruta de Kennedy.

Para terminar, vemos como realmente la ruta de Entner-Duodoroff esta muy activada en comparación con las demás rutas de síntesis de gliceraldehído. Estudio preliminares de proteómica indican que esta situación no es exclusiva en el uso del gluconato como fuente de carbono, sino que también podría ocurrir en el caso de la glucosa. Con las activaciones de la aldehído-deshidrogenasa (1.2.1.3) y las WS/DGAT para ser suficiente para una acumulación singular de triglicéridos. Esto sugiere que la mayoría de las enzimas de la ruta de Kennedy no son limitantes en el proceso. Las WS/DGAT mapeadas en este esquema son las anotadas por Hernández *et. al.* [6]. Parece que dos de ellas si son realmente activas. En la siguiente tabla podemos ver sus expresiones.

Gene ID	Enzyme Name	Expression (log2)
ro02966	WS/DGAT	2.12
ro05356	WS/DGAT	2.61
ro05649	WS/DGAT	1.13
ro06332	WS/DGAT	0.88

Tabla 2: Expresión de las hipotéticas WS/DGAT.

Discusión

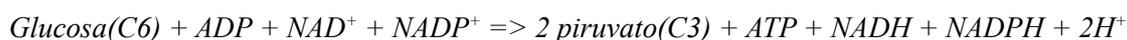
La perspectiva de poder producir TGs a partir de sobrantes industriales, tales como el suero láctico o los excedentes de las papeleras, es un aliciente para el estudio de bacterias tan versátiles como *Rhodococcus jostii*. Estas bacterias dotadas con un metabolismo muy complejo, capaz de realizar reacciones que le permiten aprovechar multitud de sustratos para su crecimiento. Es importante recalcar que el único objetivo de una bacteria es crecer y duplicarse, por lo tanto, todo lo que haga tiene que ir orientado hacia este objetivo. Es por esto que si *R. jostii* almacena TGs es por que le ofrece una ventaja adaptativa. En la introducción ya discutimos los motivos por los cuales almacenaba ácidos grasos. En vista de que se observa que se produce una mayor acumulación de TGs cuando no tiene una fuente de nitrógeno accesible pero si suficiente carbono, parece indicar que efectivamente la acumulación de estos ácidos grasos es una forma de guardar el excedente de carbono para utilizarlo cuando exista presencia de nitritos. Queda por comprobar si el efecto de la acumulación es debido a un exceso en la producción o a una falta de dilución por la baja tasa de crecimiento. El estudio transcripcional deja claro que los canales de producción de los TGs se encuentran activados, aunque esta es una prueba indirecta.

En este aspecto, es tan interesante observar los genes mas sobre expresados, como realizar una lectura acerca de los enzimas que no son limitantes. Hay tres aspectos interesantes del estudio que merecen ser comentados detenidamente. La activación del ciclo de las pentosas fosfato, en particular la vía Entner-Duodoroff, el posible aumento de la producción de piruvato a través del acetato y la dinámica transcripcional de la ruta de *Kennedy*.

La ruta de las pentosas fosfato es quizá lo mas destacable y sorprendente del trabajo. En un principio la activación de la ruta Entner-Duodoroff no es relevante, si bien, la mayor parte de la ruta comprende la conversión de gluconato en piruvato. La diferencia inicial de fuente de carbono entre glucosa y gluconato parece suficiente explicación para este efecto. Sin embargo estudios posteriores que se están actualmente analizando parecen indicar que esta activación no es exclusiva del uso de gluconato. La ruta de Entner-Duodoroff comprende la conversión de glucosa a piruvato de forma alternativa a la ruta de la glucólisis o de la pentosas fosfato. Esta conversión incluye el paso intermedio de la formación de 6-fosfo-D-gluconato al cual se puede llegar desde el gluconato mediante la enzima

2.7.1.12 (RHA1_ro02362) que es una glucokinasa y que en nuestro caso esta activada fuertemente. Además su expresión en el medio mínimo suplementado esta muy por encima de la media de expresión (203714 frente a la media de 8000). Según la anotación existente, *R. jostii* es capaz de realizar tanto la ruta de las pentosas fosfato como la glicólisis, así que el uso de la vía Entner-Duodoroff tiene que ofrecerle un beneficio. Este paso intermedio en la conversión al 6-fosfo-D-gluconato puede ser la explicación del porque *R. jostii* acumula más TGs [6] en gluconato que en glucosa

En términos químicos la reacción ofrece el siguiente balance.



En este caso la producción de ATP y NADH es la mitad que en el caso de lo glucólisis pero se produce directamente D-gliceraldehído-3-fosfato y piruvato, lo cual le puede ofrecer una ventaja en la producción de los TGs ahorrando pasos intermedios.

Es estos momentos estamos tratando de verificar estos resultados mediante qPCR para asegurarnos de que la activación de la ruta Entner-Duodoroff es independiente de la fuente de carbono.

La biosíntesis de acetyl-CoA es muy importante durante todo el proceso de la formación de TGs. Es por esto que le hemos prestado un especial interés a todos los genes implicados en este cometido. Tanto en la glucólisis como en la vía de Entner-Duodoroff el paso del D-gliceraldehído-3-fosfato a piruvato (EC 4.1.2.4. aldolasa deoxiribosa-fosfato), involucra la generación de acetato. Podría suceder que es por este exceso de producción de D-gliceraldehído-3-fosfato se deba de reciclar el acetato en acetyl-CoA. El excedente de acetato (en forma acetaldehído) es perjudicial para el metabolismo. Esto explicaría porque existe una activación de las rutas que implican la conversión del acetato en acetyl-CoA. No podemos asegurar si de esta forma se genera mayor cantidad de acetyl-CoA o es un efecto de estar metabolizando mayor cantidad de carbono. En cualquier caso, parece que *R. jostii* escoge esta activación como una forma de equilibrar el metabolismo central.

El paso fundamental de la producción de TGs se produce en la ruta de *Kennedy*, nuestros resultados establecen que la activación de las WS/DGAT y la de la aldehído-deshidrogenasa (EC:1.2.1.3) parecen suficientes para la acumulación de los TGs. Resultados de nuestro laboratorio, externos a este trabajo, en donde la sobreexpresión de WS/DGAT es suficiente para la producción de TGs, indican que esta hipótesis es correcta. De igual forma observamos que la actividad de el resto de enzimas parece no ser limitante en el proceso. En el caso de la aldehído-deshidrogenasa su actividad

no es exclusiva de la ruta de *Kennedy* por lo que su modificación no es tan representativa. Sin embargo parece que la activación o represión de las WS/DGAT puede ser el interruptor final en la acumulación o no de los TGs.

Además de los datos específicos, en general, se observa en el estudio que el metabolismo del DNA como la maquinaria de degradación de RNA están reprimidos. La degradación de algunas fuentes de carbono alternativas están reprimidas. El ciclo citrato esta moderadamente reprimido así como el catabolismo de aminoácidos.

Todo parece indicar que *R. jostii* realiza una parada de su maquinaria de crecimiento y desvía todos sus esfuerzos en mantener los niveles mínimos de metabolitos necesarios para la supervivencia y el aprovechamiento del excedente de carbono, almacenándolo para su uso posterior cuando las condiciones sean mas favorables.

Conclusiones

En vista de los resultados obtenidos podemos asegurar que.

1. En un cultivo de *Rhodococcus jostii* en un medio con gluconato como fuente de carbono, y bajas concentraciones de nitrógeno y azufre, existe una gran activación de la ruta de Entner-Duodoroff.
2. Que el exceso necesario de la producción de acetyl-CoA se produce a través de la metabolización del acetato resultante de la oxidación del gliceraldehído-3P.
3. Que el complejo I de las NADH deshidrogenasas esta fuertemente activado.
4. Solo las WS/DGAT y la aldehído-deshidrogenasa están significativamente activadas en la ruta de *Kennedy*.

Referencias

- [1] Alvarez H. *Biology of Rhodococcus*. Springer Verlag, 2010.
- [2] Alvarez HM & Steinbüchel A. *Triacylglycerols in prokaryotic microorganisms*. Appl. Microbiol. Biotechnol. (2002) **60**: pp. 367-376.
- [3] Edgar R. *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. Nucleic acids research (2004) **32**: p. 1792.
- [4] Price M, Dehal P & Arkin A. *FastTree 2--approximately maximum-likelihood trees for large alignments*. PLoS One (2010) **5**: p. e9490.
- [5] Kalscheuer R & Steinbüchel A. *A Novel Bifunctional Wax Ester Synthase/Acyl-CoA: Diacylglycerol Acyltransferase Mediates Wax Ester and Triacylglycerol Biosynthesis in Acinetobacter calcoaceticus ADP1*. Journal of Biological Chemistry (2003) **278**: p. 8075.
- [6] Hernández MA, Mohn WW, Martínez E, Rost E, Alvarez AF & Alvarez HM. *Biosynthesis of storage compounds by Rhodococcus jostii RHA1 and global identification of genes involved in their metabolism*. BMC Genomics (2008) **9**: p. 600.
- [7] Alvarez HM, Silva RA, Cesari AC, Zamit AL, Peressutti SR, Reichelt R, Keller U, Malkus U, Rasch C, Maskow T, Mayer F & Steinbüchel A. *Physiological and morphological responses of the soil bacterium Rhodococcus opacus strain PD630 to water stress*. FEMS Microbiol Ecol (2004) **50**: pp. 75-86.
- [8] Silva RA, Grossi V & Alvarez HM. *Biodegradation of phytane (2,6,10,14-tetramethylhexadecane) and accumulation of related isoprenoid wax esters by Mycobacterium ratisbonense strain SD4 under nitrogen-starved conditions*. FEMS Microbiol Lett (2007) **272**: pp. 220-228.
- [9] Silva RA, Grossi V, Olivera NL & Alvarez HM. *Characterization of indigenous Rhodococcus sp. 602, a strain able to accumulate triacylglycerides from naphthyl compounds under nitrogen-starved conditions*. Res. Microbiol. (2010) **161**: pp. 198-207.
- [10] Daniel J, Deb C, Dubey VS, Sirakova TD, Abomoelak B, Morbidoni HR & Kolattukudy PE. *Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in Mycobacterium tuberculosis as it goes into a dormancy-like state in culture*. J Bacteriol (2004) **186**: pp. 5017-5030.
- [11] Olukoshi ER & Packter NM. *Importance of stored triacylglycerols in Streptomyces: possible carbon source for antibiotics*. Microbiology (1994) **140 (Pt 4)**: pp. 931-943.
- [12] Bloom JS, Khan Z, Kruglyak L, Singh M & Caudy AA. *Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays*. BMC Genomics (2009) **10**: p. 221.
- [13] Wang Z, Gerstein M & Snyder M. *RNA-Seq: a revolutionary tool for transcriptomics*. Nat. Rev. Genet. (2009) **10**: pp. 57-63.
- [14] Marioni JC, Mason CE, Mane SM, Stephens M & Gilad Y. *RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays*. Genome Res. (2008) **18**: pp. 1509-1517.
- [15] Wilhelm BT & Landry J. *RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing*. Methods (2009) **48**: pp. 249-257.
- [16] van Vliet AHM. *Next generation sequencing of microbial transcriptomes: challenges and opportunities*. FEMS Microbiol. Lett. (2010) **302**: pp. 1-7.
- [17] de Magalhaes JP, Finch CE & Janssens G. *Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions*. Ageing Res. Rev. (2010) **9**: pp. 315-323.
- [18] Sorek R & Cossart P. *Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity*. Nat. Rev. Genet. (2010) **11**: pp. 9-16.
- [19] Linnarsson S. *Recent advances in DNA sequencing methods - general principles of sample preparation*. Exp. Cell Res. (2010) **316**: pp. 1339-1343.
- [20] Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF & Vogel J. *The primary transcriptome of the major human pathogen Helicobacter pylori*. Nature (2010) **464**: pp. 250-255.
- [21] Filiatrault MJ, Stodghill PV, Bronstein PA, Moll S, Lindeberg M, Grills G, Schweitzer P, Wang W, Schroth GP, Luo S, Khrebtukova I, Yang Y, Thannhauser T, Butcher BG, Cartinhour S & Schneider

- DJ. *Transcriptome analysis of Pseudomonas syringae identifies new genes, noncoding RNAs, and antisense activity.* J. Bacteriol. (2010) **192**: pp. 2359-2372.
- [22] Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR & Dougan G. *A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi.* PLoS Genet. (2009) **5**: p. e1000569.
- [23] Hansen KD, Brenner SE & Dudoit S. *Biases in Illumina transcriptome sequencing caused by random hexamer priming.* Nucleic acids research (2010) : .
- [24] Sambrook J, Fritsch E & Maniatis T. *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- [25] Alvarez H, Kalscheuer R & Steinbüchel A. *Accumulation and mobilization of storage lipids by Rhodococcus opacus PD630 and Rhodococcus ruber NCIMB 40126.* Applied microbiology and biotechnology (2000) **54**: pp. 218-223.
- [26] Langmead B, Trapnell C, Pop M & Salzberg S. *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol (2009) **10**: p. R25.
- [27] Li R, Yu C, Li Y, Lam T, Yiu S, Kristiansen K & Wang J. *SOAP2: an improved ultrafast tool for short read alignment.* Bioinformatics (2009) **25**: p. 1966.
- [28] Li H, Ruan J & Durbin R. *Mapping short DNA sequencing reads and calling variants using mapping quality scores.* Genome research (2008) **18**: p. 1851.
- [29] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G & Durbin R. *The sequence alignment/map format and SAMtools.* Bioinformatics (2009) **25**: p. 2078.
- [30] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M & Barrell B. *Artemis: sequence visualization and annotation.* Bioinformatics (2000) **16**: p. 944.
- [31] Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B. *Mapping and quantifying mammalian transcriptomes by RNA-Seq.* Nat. Methods (2008) **5**: pp. 621-628.
- [32] Bullard JH, Purdom E, Hansen KD & Dudoit S. *Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments.* BMC Bioinformatics (2010) **11**: p. 94.
- [33] Anders S. *Analysing RNA-Seq data with the DESeq package.* EMBL internal document (2010) : pp. 1-17.
- [34] Kanehisa M & Goto S. *KEGG: Kyoto encyclopedia of genes and genomes.* Nucleic acids research (2000) **28**: p. 27.
- [35] Carver T, Thomson N, Bleasby A, Berriman M & Parkhill J. *DNAPlotter: circular and linear interactive genome visualization.* Bioinformatics (2009) **25**: p. 119.