



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

CENTRO DE CIENCIAS GENÓMICAS

ANÁLISIS GLOBAL DE LA EXPRESIÓN DEL GENOMA DE *Rhizobium etli* CFN42
EN LA VIDA LIBRE Y DURANTE LA SIMBIOSIS CON *Phaseolus vulgaris* L.

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

BIÓL. EMMANUEL SALAZAR BUSTAMANTE

TUTOR PRINCIPAL:

DR. SERGIO MANUEL ENCARNACIÓN GUEVARA

CENTRO DE CIENCIAS GENÓMICAS

COMITÉ TUTOR:

DR. JAIME MORA CELIS

CENTRO DE CIENCIAS GENÓMICAS

DR. JUAN ENRIQUE MORETT SÁNCHEZ

INSTITUTO DE BIOTECNOLOGÍA

MÉXICO, D.F. JULIO 2013



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

AGRADECIMIENTOS

Al Doctorado en Ciencias Biomédicas, UNAM.

Al CONACyT por la beca otorgada para realizar los estudios de doctorado (No. de apoyo 170199).

A la DGAPA-PAPIIT por los apoyos otorgados para la realización de este trabajo (proyectos IN-206113 e IN-212710).

De manera especial, quiero agradecer al director de tesis al Dr. Sergio M. Encarnación Guevara, su apoyo, consejo, confianza y su compromiso en mi formación académica.

Agradezco a los tutores del proyecto inicial: Dr. Enrique Morett Sánchez y en especial al Dr. Emérito de la UNAM Jaime Mora Celis, porque todos sus consejos, opiniones y críticas ayudaron significativamente para la realización del trabajo.

A la Química Yolanda Pérez-Tejada Domínguez y a su equipo agradezco su total apoyo en los experimentos realizados en invernadero.

Un agradecimiento especial es para los ingenieros O. Alejandro Aguilar Vera y M. Gabriela Guerrero Ruiz, por su disponibilidad en el apoyo informático.

Agradezco al Biól. Andrés Andrade Domínguez por su asistencia en los ensayos de fluorescencia en nódulos de frijol.

También agradezco al Dr. Jorge Ramírez Salcido, al M. en C. Simón Guzmán León, al Dr. José Luis Santillán Torres y la Biól. Lorena Chávez González, integrantes de la unidad de microarreglos de la IFC-UNAM su apoyo en la impresión de los microarreglos. Así como a Lina Riego Ruiz, Gerardo Coello Coutiño, y demás participantes del programa "GenArise" utilizado para analizar los datos de microarreglos.

También agradezco a los integrantes de la unidad de síntesis de oligonucleótidos, IBT-UNAM, por la calidad y prontitud de los mismos.

Agradezco al Dr. Humberto Peralta Díaz y al Dr. Javier Izquierdo Sánchez por su apoyo en la redacción de la tesis y en el formato inicial del artículo, respectivamente.

A la M. en C. Magdalena Hernández Ortiz y a Gabriel Martínez Batallar por su auxilio en diversos asuntos del laboratorio.

A los miembros del Jurado: Dra. Laura Camarena Mejía, Dr. Ismael Hernández Lucas, Dr. Víctor Manuel González Zúñiga y Dr. Jorge Luis Folch Mallol, agradezco sus valiosos y atinados comentarios.

DEDICATORIA

A mi esposa Maricela Tejeda Aburto, por su amor y apoyo todos estos años

A nuestra pequeña Samantha Dinora, la alegría y motivación de la familia

A mis padres Rosa Refugio Bustamante Rodríguez y Jorge Salazar del Río, por el modelo de vida en todos los sentidos

A mis hermanos Georgina Cecilia y Jorge Alberto Salazar Bustamante, por la fraternidad

A mis suegros y cuñado, por integrarme en la familia

A mis tíos del Sur y los del Norte

A los primos y sobrinos del Sur

A los primos y sobrinos del Norte, en memoria de Gerardo Salazar Ponce

A la familia de Veracruz

A los compadres y ahijados

A todos los compañeros y amigos del CCG en especial los que conforman el programa de Genómica Funcional de Procariotes

A los compañeros y amigos de los miércoles y a veces sábados deportivos

“La expresión génica es como la vida, ambas presentan una fina y compleja regulación”

(Emmanuel S.B.)

ÍNDICE	Página
1) RESUMEN.....	6
2) ABSTRACT.....	7
3) INTRODUCCIÓN.....	8
4) SIMBIOSIS Y FIJACIÓN DE NITRÓGENO.....	10
4.1 <i>Rhizobium etli</i> CFN42.....	15
5) REGULACIÓN EN PROCARIOTES.....	16
5.1 Factores sigma.....	18
5.2 Sigma 54 Y NifA.....	20
6) MICROARREGLOS Y GENÓMICA FUNCIONAL.....	22
7) ANTECEDENTES.....	24
8) OBJETIVOS.....	28
9) MATERIAL Y METODOS ADICIONALES A LOS ARTICULOS.....	29
10) RESULTADOS.....	32
10.1 Artículo: “Characterization of the NifA-RpoN regulon in <i>Rhizobium etli</i> in free life and in symbiosis with <i>Phaseolus vulgaris</i> ”.....	33
11) RESULTADOS ADICIONALES.....	44

11.1	Artículo: “Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling”	45
11.2	Análisis de la comparación de la vida libre (VL) contra Simbiosis 18 ddi	60
11.3	Análisis de la expresión global en diferentes etapas de la fijación de nitrógeno y una condición de vida libre.....	78
	11.3.1 Agrupamiento y características globales por grupo o perfil de expresión.....	80
	11.3.2 Descripción individual de cada gen de acuerdo al perfil...	85
12)	DISCUSIÓN Y CONCLUSIONES.....	178
13)	PERSPECTIVAS.....	215
14)	REFERENCIAS.....	218

1. RESUMEN

Rhizobium etli CFN42, es una α proteobacteria que fija nitrógeno atmosférico en asociación con *Phaseolus vulgaris* (frijol). Su genoma está compuesto por un cromosoma y seis mega-plásmidos, el total de marcos abiertos de lectura (ORFs) es de 6034. *R. etli* presenta dos estilos de vida diferentes, en vida libre y en simbiosis con frijol. Para analizar la expresión del genoma de *R. etli* en vida libre y en simbiosis, construimos el microarreglo de *R. etli* CFN42, el cual está constituido por 6050 oligonucleótidos representando a los 6034 ORFs del genoma y 16 controles. El regulador maestro de los genes que participan en la fijación biológica de nitrógeno está compuesto por las proteínas NifA y RpoN. Para definir los genes dependientes de RpoN-NifA, utilizamos condiciones microaeróbicas (vida libre) y bacteroides (simbiosis) de 11 días después de la inoculación (ddi). El análisis comparativo de estas cepas, mutante y silvestre, en simbiosis y microaerobiosis definió un total de 79 genes. Por otro lado, en este estudio también se comparó una condición de vida libre en fase exponencial contra bacteroides de 18 ddi. Los datos, nos indicaron, que los genes que codifican para proteínas ribosomales son más expresados en vida libre en comparación con simbiosis, y los genes involucrados en el metabolismo de Nitrógeno presentan una expresión mayor en simbiosis, además, otras vías fueron enriquecidas específicamente, en vida libre (17) y simbiosis (18) indicando el contraste que existe entre el metabolismo en vida libre y en el bacteroide. Finalmente, en este trabajo de tesis presentamos el análisis de expresión realizado en 5 diferentes tiempos de simbiosis (11, 18, 25, 28 y 35 ddi) y la vida libre. En conjunto, con estos y otros datos proponemos el modelo de funcionamiento del bacteroide de *R. etli* CFN42 durante la simbiosis con *P. vulgaris* y la vida libre.

2. ABSTRACT

Rhizobium etli CFN42, is an α proteobacteria that fixes atmospheric nitrogen in association with *Phaseolus vulgaris* (bean). Its genome consists in a chromosome and six mega-plasmids, the total number of open reading frames (ORFs) is 6034. *R. etli*, presents two different lifestyles, in free life and in symbiosis with bean. To analyze the expression of the genome of *R. etli* in free living and in symbiosis conditions, we built the microarray of *R. etli* CFN42, the microarray consists of 6050 oligonucleotides representing, in fragments of 70 bases, 6034 ORFs of the genome and 16 controls. The master regulator of genes involved in biological nitrogen fixation is composed by the NifA and RpoN proteins. To define the RpoN-NifA-dependent genes, we use microaerobic conditions (free-living) and bacteroids (symbiosis) 11 days after inoculation (dpi). The comparative analysis of the mutant and wild type strains, in symbiosis and microaerobic defined a total of 79 genes. On the other hand, also using microarray data we compared a free-living in exponential phase condition against bacteroids of 18 dpi. These data, indicated us genes that encode for ribosomal proteins are more expressed in free-living compared with symbiosis, and genes involved in the metabolism of nitrogen presented a higher expression in symbiosis, in addition, other pathways were differentially enriched in free life (17) and symbiosis (18) indicating the contrast that exists between metabolism in free-living as well as in the bacteroid. Finally, in this thesis work we also present the analysis carried out in 5 different times of nitrogen fixation (11, 18, 25, 28 and 35 dpi) and free-living. Together, with these and other extra data we suggest the functioning model of the rhizobia for *R. etli* CFN42 during the symbiosis with *P. vulgaris* and free life.

3. INTRODUCCIÓN

Las bacterias de la familia Rhizobiaceae son α -proteobacterias que poseen la capacidad para infectar y establecer una relación simbiótica en raíces de plantas leguminosas. Esta simbiosis debe su importancia agronómica a varios miembros de la familia Leguminosae las cuales son utilizadas como alimento o bien como forraje de animales entre estas destacan el frijol, la soya, la alfalfa, el chícharo, el haba, el trébol, entre otras. En esta asociación todas ellas fijan nitrógeno atmosférico y reciben de la planta compuestos de carbono y otros nutrientes. El frijol común (*Phaseolus vulgaris* L.) es la leguminosa de grano más importante para el consumo humano y prácticamente se cultiva en todo el mundo. Latinoamérica es la región que produce más del 45% del total de la producción mundial. Los principales países productores son Brasil, Estados Unidos, México, Argentina en América, India, Birmania, China en Asia, Tanzania, Kenia, y Uganda en África, adicionalmente para los habitantes de estas zonas el frijol representa la fuente principal de proteínas, vitaminas, minerales y fibras a la cual tienen acceso, especialmente las poblaciones más pobres de África y América Latina (Voysest, 2000 y Bitocchi *et al.*, 2012).

La producción del frijol común en las regiones productoras de Latinoamérica y África es inferior a su potencial, debido a diversos factores limitantes que incluyen la sequía, plagas, enfermedades, y a la baja fertilidad de los suelos. La deficiencia de nutrientes en la mayoría de los suelos dedicados a la producción de este cultivo, limita los niveles de productividad. La baja inversión en la aplicación de insumos y prácticas recomendadas para el manejo en la mayoría de las regiones productoras, contrasta con la de los países desarrollados donde la productividad del mismo es significativamente mayor. Aunque el nitrógeno es abundante en la atmósfera (alrededor del 80%), las plantas no pueden utilizarlo en su forma elemental (N_2) y tienen que obtenerlo del suelo en forma de nitratos. En los suelos, dicho elemento está contenido en la materia orgánica y en muchas regiones su bajo contenido no cubre las necesidades de los cultivos. La Fijación Biológica de Nitrógeno (FBN) es un proceso clave en la biosfera señalado incluso como el segundo proceso más importante para la vida solo por debajo de la fotosíntesis, en el cual,

bacterias de la familia Rhizobiaceae, inducen en las raíces de las leguminosas la formación de estructuras especializadas llamadas nódulos mediante el establecimiento de un proceso simbiótico. Estos microorganismos sintetizan la enzima nitrogenasa y por medio de ésta el nitrógeno atmosférico (N_2) se reduce a amoníaco (NH_3), y posteriormente a amonio (NH_4) compuesto estable que ya es asimilable por la planta, también se ha reportado la alanina es excretada por el bacteroide, siendo esta otra forma para acarrear el amonio hacia la planta (Waters et al., 1998). Se estima que este proceso contribuye en el 60-80% de la FBN, sin embargo es un proceso altamente demandante de energía, ya que se utilizan 16 moléculas de ATP por N_2 reducido. Las leguminosas en asociación con las rizobios, fijan alrededor de 227 Kg por hectárea y al año producen alrededor de 21.5 teragramos (Tg) de nitrógeno a nivel mundial, de este nitrógeno fijado alrededor del 70% le corresponde a los cultivos de soya de EUA, Brasil y Argentina (Herridge *et al.*, 2008). La inoculación con cepas mejoradas de *Rhizobium* que presenten una mayor capacidad de nodulación (competencia) y fijación de N_2 , es una alternativa para suplir las necesidades del nutriente y aumentar la productividad de este cultivo en zonas con suelos de bajo contenido de materia orgánica. El uso de organismos simbióticos en los cultivos de frijol incrementa su productividad. Un estudio reciente determinó el origen del frijol común en Mesoamérica, lo que brinda a la región y a los investigadores la oportunidad de maximizar y explorar la diversidad genética de la planta y de sus simbioses (Bitocchi *et al.*, 2012). Una de las bacterias del género *Rhizobium* preponderante en la región Mesoamericana es *R. etli*; sin embargo, el efecto benéfico de esta bacteria sobre el cultivo de frijol se ve afectado por las cepas nativas presentes en los suelos donde se cultiva esta leguminosa (Thies *et al.*, 1991; Cuadrado *et al.*, 2009). El proceso de adaptación de las bacterias a su medio, ya sea alrededor de la raíz o en el nódulo, implica señales físicas y químicas, que pueden ser detectadas de diferentes maneras, dichas señales o estímulos pueden penetrar la célula o ser detectadas en la superficie celular a través de receptores de membrana, el resultado al final es la modificación de la expresión genética, también conocida como regulación transcripcional.

4. SIMBIOSIS Y FIJACIÓN DE NITRÓGENO

La simbiosis es la interacción entre 2 o más individuos de especies diferentes donde cada uno obtiene un beneficio, el sistema simbiótico entre rizobiáceas y leguminosas es uno de los eventos analizados en cientos de trabajos y por consecuencia a mayor detalle. En el siglo XVI Leonhard Fuchsius dibujó leguminosas con nódulos en sus raíces (Fuchsius L., 1542); más tarde, en el siglo XVII, Malpighi reportó sus observaciones de nódulos presentes en las raíces de frijoles o judías (*Phaseolus vulgaris*) y habas (*Vicia faba*) (Malpighi M., 1675). Posteriormente, en 1888 Hellriegel y Wilfarth, reportaron en distintas leguminosas, que para la formación de nódulos se requiere de una infección previa, por este trabajo ambos son reconocidos como los descubridores de la fijación simbiótica de nitrógeno (Hellriegel H. y Wilfarth H., 1888). En el mismo año, Beijerinck comprobó la necesidad de una infección bacteriana para la formación del nódulo esto al inocular plantas de *Vicia faba* con cultivos de bacterias provenientes de nódulos de la misma leguminosa (Beijerinck M.W., 1888).

Las rizobiáceas son un grupo muy heterogéneo de bacterias que se han dividido en cuatro familias: Rhizobiaceae, Phyllobacteriaceae, Hyphomicrobiaceae y Bradyrhizobiaceae (Madigan *et al.*, 2000). Dentro de estas familias sólo determinados géneros son capaces de efectuar el proceso de fijación de nitrógeno: *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium* y *Allorhizobium*. A diferencia de las cianobacterias y otros organismos, por ejemplo del genero *Frankia*, *Klebsiella* entre otras, las rizobiáceas no pueden generar su ambiente microaeróbico para poder fijar nitrógeno en vida libre, por lo que requieren de un hospedero que ofrezca las condiciones para llevar a cabo el proceso de fijación de nitrógeno. Para que el proceso simbiótico entre la bacteria y la planta se realice se requieren una serie de pasos que involucran desde señales químicas a modificaciones estructurales tanto de la planta como de la bacteria. Este se puede resumir en cuatro etapas que son:

- i) Intercambio de señales de naturaleza química entre la planta y el microorganismo (flavonoides y factores de nodulación).

- ii) Activación del ciclo celular en células del córtex e iniciación del nuevo órgano en la planta (deformación de la raíz).
- iii) Invasión de la bacteria, formación del canal de infección y penetración a los tejidos recién formados (formación del nódulo).
- iv) Diferenciación de la bacteria a su forma especializada (bacterioide con capacidad de fijar nitrógeno).

En cada etapa tanto la bacteria como la planta tienen que seguir un proceso, si alguno incumple con uno de los elementos que interviene en la interacción mutualista, se presenta lo que se conoce como sanciones por parte del hospedero, dichas sanciones pueden abortar la nodulación o bien que el nódulo inicie una senescencia temprana (Kiers *et al.*, 2003, Marco *et al.*, 2009).

La interacción entre la bacteria fijadora y la planta leguminosa, que conduce a la formación de la estructura donde se llevará a cabo la fijación de nitrógeno, inicia con un intercambio específico de señales. Una parte de los compuestos producidos en la fotosíntesis de la planta son secretados en los exudados radiculares, estos incluyen carbohidratos, ácidos orgánicos, vitaminas, aminoácidos y derivados fenólicos. Entre los compuestos secretados por las plantas se encuentran los flavonoides, cuya composición varía dependiendo de la especie, los cuales desencadenan una serie de respuestas específicas en los rizobios, es decir, inducen la transcripción de los genes *nod*. La inducción de los genes *nod* a su vez resulta en la secreción de lipoquitoligosacáridos los cuales se unen y activan en la planta receptores tipo cinasas que contienen motivos LysM, al final se inicia una compleja vía de señalización que involucra oscilaciones de calcio en las puntas de las raíces. La excepción del uso o esencialidad de los factores de nodulación se presenta en algunas bacterias fotosintéticas del género *Bradyrhizobium* en las cuales se observa una estrategia de nodulación independiente a los factores clásicos de nodulación, posiblemente promovida por proteínas (Giraud *et al.*, 2007). No obstante, el resultado final es el enrollamiento de los pelos radiculares con lo que se capturan los rizobios, formando los hilos de infección. Lo que al final promoverá la morfogénesis del nódulo (Prell y Poole, 2006, Oldroyd *et al.*, 2011).

Dependiendo del sistema simbiótico podemos encontrar dos tipos de nódulos: determinados e indeterminados. Si las divisiones mitóticas se originan en el córtex interno se originan nódulos indeterminados (p ej. *S. meliloti-Medicago truncatula*) y si lo hacen en el córtex externo son nódulos determinados (p ej. *R. etli-Phaseolus vulgaris*, *B. japonicum-Glycine max*). Ambos tipos de nódulos, además de presentar una estructura anatómica distinta, también difieren en la forma en que se comporta la bacteria dentro del nódulo en formación. A pesar de ello, la inducción del ciclo celular en ambos sistemas sigue en la planta una regulación similar.

Los nódulos indeterminados se presentan en los géneros *Medicago*, *Pisum*, *Trifolium* y *Vicia* entre otros. En este tipo de nódulos son las células del córtex interior las que se reintroducen en el ciclo celular, además, tienen la característica de poseer un meristemo permanente, lo que les otorga una forma cilíndrica con simetría radial en la organización de los tejidos. Así en la zona más exterior se hallan las células vacuoladas del córtex, y hacia el interior se encuentran la endodermis y el parénquima, en donde también aparecen los haces vasculares. Todo ello cubre una zona central en donde *Rhizobium* se alberga y realiza la fijación de nitrógeno, en el nódulo indeterminado se pueden identificar cuatro zonas o etapas de interacción entre bacterias y células de la planta (Hirsch, 1992; Timmers *et al.*, 1999 y Oldroyd *et al.*, 2011) (Fig. 1):

- i) Zona I o meristemática: en el ápice del nódulo, corresponde a la zona de células en proliferación.
- ii) Zona II o de invasión: inmediatamente por debajo de la zona meristemática, es la región en la que se produce la invasión bacteriana a través de los canales de infección. Las células de esta región son más grandes y vacuoladas que las meristemáticas. Los rizobios en esta zona aún poseen una forma cilíndrica y pueden dividirse (bacteroides tipo 1).
- iii) Zona de prefijación: en esta región las células vegetales, que aún no han finalizado su diferenciación, están repletas de bacteroides de tipo 2 que son más alargados que los de tipo 1.

- iv) Interzona II y III: en esta franja, las células vegetales finalizan su proceso de diferenciación. Las células de esta región presentan numerosos amiloplastos así como transcritos de leghemoglobina, proteína encargada de regular la presencia del oxígeno en el nódulo. En esta zona ya se localizan bacteroides de tipo 3 los cuales presentan su tamaño final definitivo, así como una heterogeneidad citoplasmática característica.
- v) Zona III o de fijación: región completamente diferenciada en la que se realiza la fijación de nitrógeno. Se subdivide en dos regiones: la zona de fijación y la zona ineficiente. En la primera, las gotas de infección han culminado su proceso de diferenciación, resultando en la formación del simbiosoma, compuesto por la membrana vegetal original modificada en su composición y un bacteroide de tipo 4 con una estructura normalmente en forma de “Y” o de “T” y con una heterogeneidad citoplasmática notable, indicativa de su maduración en forma fijadora de nitrógeno. La zona ineficiente está compuesta por células con bacteroides de tipo 5 que presentan un citoplasma homogéneo lo que significa el inicio de la senescencia.
- vi) Zona IV o de senescencia: región en la base del nódulo, comprendida por células vegetales y bacterianas en degradación que depende de la edad del nódulo.
- vii) Zona V bacterias no diferenciadas a bacteroides, no hay actividad de la nitrogenasa y las bacterias presentan un fenotipo vegetativo, se observa después de la 6 semanas el tamaño de la zona V aumenta con la edad del nódulo.

Los nódulos determinados se inducen en plantas del género *Phaseolus*, *Glycine*, *Vigna* y *Lotus*, entre otras. Este tipo de nódulos no presentan un meristemo permanente, por lo que su crecimiento se basa en la expansión y no en la división celular, razón por la que presentan una morfología esférica en lugar de cilíndrica (Hirsch, 1992). La causa de la ausencia de un meristemo permanente se debe a que las primeras divisiones celulares en respuesta a la presencia de *Rhizobium* son anticlinales y se producen en la hipodermis (Newcomb *et al.*, 1979; Rolfe y Gresshoff, 1988). Posteriormente, se genera otro foco de división celular en el periciclo, a continuación, estos dos meristemas convergen

generando el primordio nodular, en el cual se localizan células no vacuoladas procedentes de las divisiones de la hipodermis conformando el tejido central del nódulo, y células con un alto grado de vacuolización procedentes de las divisiones en el periciclo, formando el parénquima nodular que rodea al tejido central. Gran parte de la actividad mitótica en la región central del nódulo se pierde transcurridos 12 a 18 días tras la inoculación (Newcomb *et al.*, 1979). Algunas células de este tejido central son invadidas a través de los canales de infección y pueden ser identificadas por su gran tamaño y densidad, debidos a la elevada presencia de simbiosomas en su interior. Estos simbiosomas, a diferencia de los existentes en nódulos indeterminados, pueden presentar más de un bacteroide en su interior. El resto de células, no infectadas, presentan un tamaño inferior y con una elevada vacuolización. Por otro lado, en el parénquima se encuentran varias capas de células separadas por exiguos espacios intercelulares y con un elevado contenido de proteínas ricas en prolina en su pared, que pueden contribuir a limitar la difusión del oxígeno al tejido central (Tjepkema, 1983 y Oldroyd *et al.*, 2011) (Fig. 1).

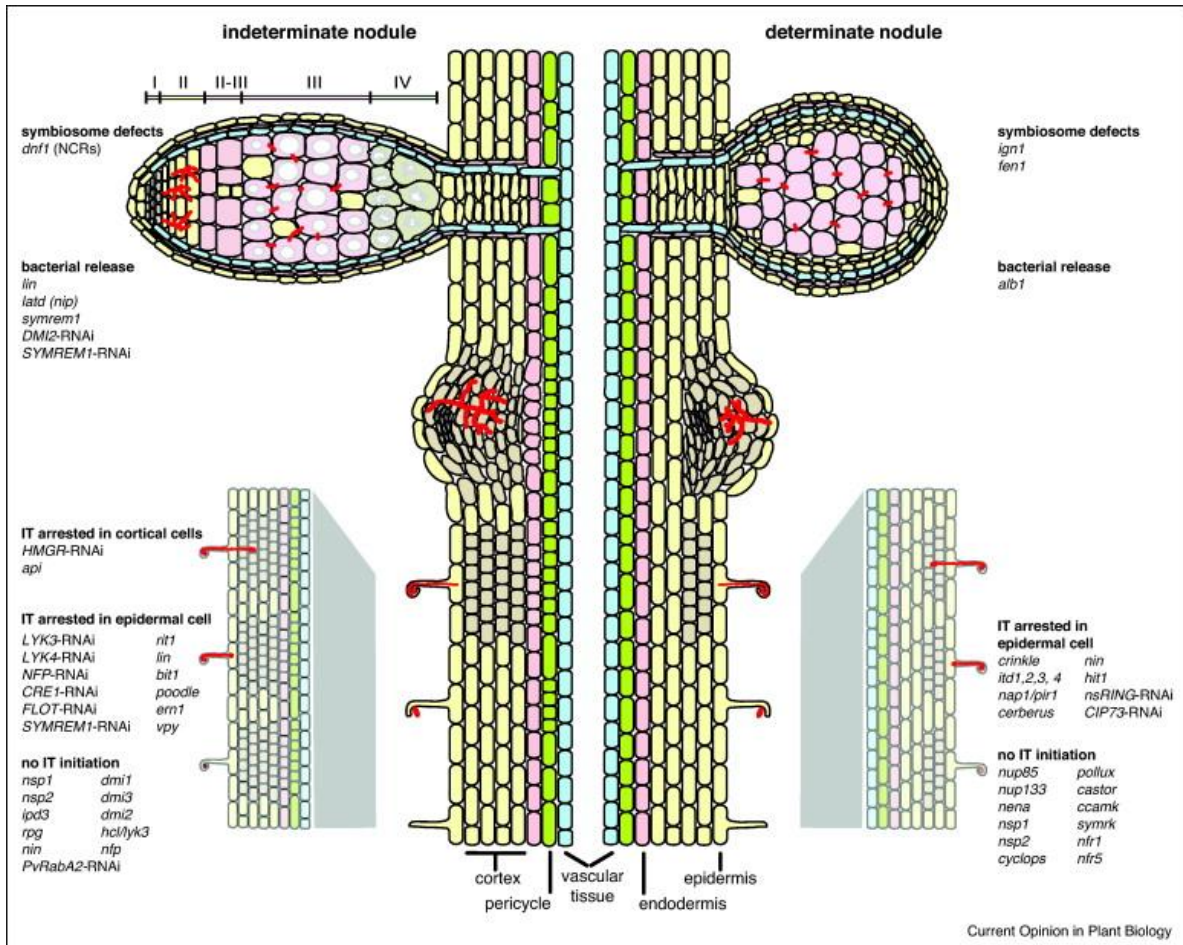


Figura 1. Progresión de la formación de un nódulo determinado y uno indeterminado, tomada de Popp y Ott, 2011.

4.1. *R. etli* CFN42

Entre las bacterias de importancia agrícola, debido a que interacciona con cultivos de consumo humano, se encuentra clasificada *R. etli* la cual nodula al frijol (*Phaseolus vulgaris* L). *R. etli* es un bacilo aeróbico, gram negativo, que pertenece a las α proteobacterias, tiene un tamaño de 0.6 a 1.2 μ . *R. etli* es resistente al ácido nalidixico, su temperatura óptima de crecimiento es de 29°C, se cultiva en extracto de levadura y peptona (PY) o bien en medio mínimo (MM) principalmente con succinato o malato como fuentes de carbono (Segovia *et al.*, 1993). Es un rhizobio de crecimiento rápido, originario de Centroamérica y México. El genoma de la cepa CFN42 está compuesto por un cromosoma y seis mega-plásmidos, que en total contienen 6,530234 pares de bases y un total de 6034 marcos abiertos de lectura (ORFs) (González *et al.*, 2006). Una descripción

detallada, que incluye algunas de las funciones principales que contiene cada replicón del genoma de *R. etli* CFN42, se presenta en la tabla 1. Entre la familia Rhizobiaceae, *R. etli* es una de las bacterias más ampliamente estudiadas, otros modelos simbióticos pertenecientes a esta familia, que han sido utilizadas en una amplia diversidad de estudios son *Mesorhizobium loti*, *Sinorhizobium meliloti*, *Bradyrhizobium japonicum*, *R. leguminosarum* y *Agrobacterium tumefaciens*. El objetivo principal para estudiar a nivel genético estas bacterias, exceptuando *A. tumefaciens* que es un patógeno de plantas, básicamente tienen la finalidad de conocer la regulación genética y fisiología de la bacteria durante este importante proceso biológico, con el propósito de lograr aumentar la fijación de nitrógeno, la competitividad de la nodulación y la resistencia a condiciones no favorables. Definiendo además los componentes involucrados en la interacción entre los rizobios y sus plantas hospederas.

Tabla 1) Características generales del genoma de *R. etli* CFN42, tamaño de cada replicón, número de genes y funciones que los caracterizan (Modificada de González *et al.*, 2006).

Molécula	Tamaño (pb)	Genes	Principales Funciones
Cromosoma	4,381,611	4,067	Funciones celulares esenciales (genes "housekeeping")
Plásmido F	642,519	573	Fijación de nitrógeno, asimilación de nitritos
Plásmido E	505,334	459	Síntesis de cobalamina, división celular
Plásmido D	371,255	354	Simbiosis (genes de nodulación y fijación de nitrógeno)
Plásmido C	250,948	234	Transportadores ABC (azúcares y aminoácidos)
Plásmido B	184,338	165	Síntesis de tiamina y lipopolisacáridos
Plásmido A	194,229	182	Transportadores tipo IV (genes vir), conjugación
Totales	6,530,234	6,034	

5. REGULACIÓN EN PROCARIOTES

La regulación de la expresión genética es la base para la diferenciación celular, la morfogénesis y para la adaptabilidad de cualquier organismo. La regulación en bacterias es esencial para la correcta activación de los genes de acuerdo a los sustratos y a las condiciones fisiológicas a las que están expuestas. El primer paso de este proceso de expresión es la transcripción de DNA a RNA el cual es catalizado por una enzima con

multisubunidades llamada RNA polimerasa (RNAP)($\alpha_2\beta\beta'$). Las fases en las que se regula la transcripción son al inicio, durante la elongación o en la fase de terminación, pero es en la fase inicial donde se presenta la mayor regulación. En la mayoría de las bacterias este proceso es altamente o firmemente regulado por un amplio repertorio de reguladores transcripcionales, particularmente por los llamados factores sigma (σ), componente de la RNA polimerasa, los cuales coordinan la regulación de la transcripción en las etapas iniciales. Entre los elementos que forman parte de la regulación transcripcional están los promotores, que son de los elementos intrínsecos del DNA, estas secuencias indican el sitio de anclaje para el factor sigma y para la RNAP, para sigma 70 el promotor está conformado por dos hexanucleótidos en las posiciones -10 y -35, separados entre 15 y 21 pares de bases. Las secuencias de los promotores varían dependiendo de cada factor sigma.

Un regulón es un grupo de genes o unidades transcripcionales que son controladas por una única proteína reguladora, en caso de que la proteína fuese un factor sigma se conoce como "sigmulón". Aunque la mayoría de los regulones están constituidos por uno o unos cuantos genes, existen regulones de varios cientos de genes, estos se conocen como reguladores maestros, algunas veces los regulones poseen elementos en común. Por otro lado, existen grupos de genes afectados por algún factor o perturbación ambiental (ej. calor, radiación, pH, entre otros) a esto se le conoce como estimulón, el tipo de respuesta al estímulo puede ser compleja ya que muchas veces involucra varios factores transcripcionales (regulones y sigmulones) que incluso son diferentes de acuerdo a la etapa en la que se encuentre la perturbación, ya sea el inicio, durante o al final de la misma. En la delimitación del estimulón también participan diversos reguladores transcripcionales que pueden funcionar incrementando la afinidad de la $E\sigma^{70}$ (holoenzima, formada por la RNA polimerasa y el factor sigma) al promotor y con ello activando la transcripción (regulación positiva) o bien evitando la apertura del complejo o la unión del sigma al promotor, afectando la transcripción (regulación negativa), (Fig. 2).

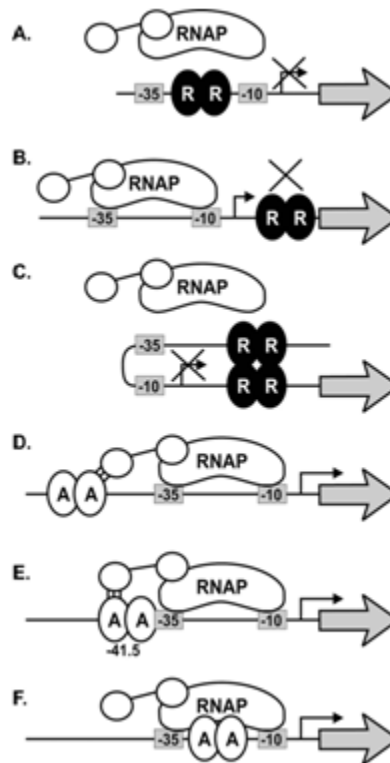


Figura 2. Ejemplos de mecanismos de regulación, negativa: A, Al unirse en la región promotora, el regulador negativo evita que se una la polimerasa (represión por impedimento estérico); B, El regulador negativo al unirse alrededor del inicio de la transcripción evita que ésta inicie (represión por bloqueo de la elongación); C, Los reguladores negativos se unen en diferentes regiones del DNA en la región de inicio y cadena arriba de sitio de unión de la polimerasa, esto pliega el DNA y con ello evita el acceso a la polimerasa (represión por bucle de DNA). Los mecanismos de regulación positiva: D, activación clase I (el regulador se une cadena arriba y contacta la subunidad α de la polimerasa); E, activación clase II, los activadores se unen cadena arriba del sitio de contacto de la polimerasa se une a la subunidad α , lo que permite el reclutamiento de esta al promotor; F, los reguladores modifican la polimerasa lo que permite a activación de la transcripción (cambio conformacional), tomada de Rodionov, 2007.

5.1. Factores sigma (σ)

La subunidad σ tiene tres funciones principales: asegurar el reconocimiento de una secuencia promotora específica, posicionar la holoenzima en un promotor diana o blanco, y facilitar la apertura de la doble hélice de ADN cerca del inicio de transcripción. Los factores sigmas están clasificados en 2 grupos, los del tipo sigma 70 y los sigma 54; a su vez los factores sigma 70 se clasifican en 4 subgrupos (Paget y Helmann, 2003). El primer

grupo corresponde a los factores sigmas esenciales relacionados con sigma 70 de *E. coli*. El segundo grupo incluye también a los relacionados con sigma 70 pero son usualmente dispensables. Las proteínas que pertenecen al tercer grupo son homólogos distantes de sigma 70 y son conocidos por activar regulones o sigmulones específicos como los de choque térmico o estados de desarrollo morfológicos como esporulación o biogénesis de flagelo. El cuarto grupo es el más grande e incluye el grupo de factores sigma extracitoplasmáticos (ECF), este cuarto grupo incluye proteínas que responde a señales del medio ambiente que incluye la detección de proteínas periplasmáticas mal plegadas (Wosten, 1998).

La distribución de los factores sigma en bacterias es amplia, siendo desde 1 en *Mycoplasma genitalium* a 63 en *Streptomyces coelicolor* (Wigneshweraraj *et al.*, 2008). *R. etli* cuenta con 23 factores sigma, un sigma 70 ó RpoD, 2 RpoH, 18 ECF y 2 RpoN o sigma 54, 8 se localizan en plásmidos y 15 en el cromosoma (González *et al.*, 2006). El “housekeeping” o gen esencial para *R. etli* es el sigma 70 (*sigA*), el cual codifica para una proteína de 685 amino ácidos y de un peso de 77.18 kDa. La estructura primaria de SigA es muy similar a RpoD de *E. coli* a excepción de la región 1, donde el SigA de *R. etli* presenta una extensión. El sigma 70 de *E. coli* no funciona en los promotores de *R. etli*, aunque el sigma 70 de *R. etli* sí reconoce los promotores de *E. coli*. Adicionalmente, las cajas -10 [TATNNT] y -35 [CTTGAC] de *R. etli* difieren del consenso de *E. coli* [TATAAT] y [TTGACA] (Ramírez-Romero *et al.*, 2006). Otro factor sigma caracterizado en *R. etli* controla la respuesta a choque térmico, RpoH, el cual es un factor sigma alternativo; *R. etli* cuenta con dos factores sigma 32. Ambos RpoH o sigma 32 de *R. etli*, son capaces de complementar a la mutante de RpoH de *E. coli*, aunque en *R. etli* se propone que tienen blancos de regulación diferentes, esto basado en que RpoH1 está involucrado en la respuesta al choque térmico y al estrés oxidativo, mientras que RpoH2 responde al estrés osmótico. Además, un análisis de sus regiones regulatorias sugiere que son regulados por diferentes sigmas (Martínez-Salazar *et al.*, 2009).

5.2. Sigma 54 y NifA

El factor transcripcional sigma 54 (σ^{54} o σ^N), es codificado por el gen *rpoN*, es un polipéptido de un peso molecular de 54 kDa. Es un sigma ampliamente distribuido en los genomas bacterianos, un poco más del 60% de las bacterias secuenciadas cuentan con al menos un σ^{54} ; la principal función asignada para la holoenzima $E\sigma^{54}$ es sobre la regulación del metabolismo nitrogenado, es por ello que en un inicio se le asignó el nombre de $E\sigma^N$ (Hunt y Magasanit, 1985). No obstante, a la fecha se le involucra en la regulación de otras funciones fisiológicas, por ejemplo la síntesis de flagelo en *Caulobacter crescentus* y *Rhodobacter sphaeroides*, en la regulación del transporte de compuestos de carbono como los ácidos dicarboxílicos en *S. meliloti*, en la regulación del transporte de fructosa, de electrones, quimiotaxis y en la respuesta de estrés por calor, entre otros. Incluso se ha propuesto que σ^{54} puede regular factores sigma alternativos como *rpoH* en *E. coli*, y *rpoS* de *Borrelia burgdorferi* (Pallen, 1999 y Smith *et al.*, 2006), a pesar de esto y de la amplia distribución de σ^{54} , solo en *Myxococcus xanthus* es vital para la sobrevivencia de la bacteria (Keseler y Kaiser, 1997). En los organismos diazótrofos, como *R. etli*, representantes de la Proteobacterias, la transcripción de casi todos los genes de la fijación dependen del factor sigma alternativo σ^{54} , el cual reconoce secuencias localizadas en las posiciones -24 y -12 (Morett y Buck, 1989, Barrios *et al.*, 1999, Dombrecht *et al.*, 2002, Leang *et al.*, 2009). Para el inicio de la transcripción a través de σ^{54} se requiere una clase específica de activador, el cual se une a elementos de tipo enlace cadena arriba de las secuencias promotoras dependientes de σ^{54} y a través de la hidrólisis de ATP cataliza los cambios conformacionales en σ^{54} , los cuales permiten a la holoenzima transitar de un complejo cerrado a uno abierto (Fig. 3). Los activadores de los que depende σ^{54} , son conocidos como EBPs por sus siglas en ingles “enhancer binding proteins” (proteínas potenciadoras de unión) (Morett y Segovia, 1993) y pertenecen a la superfamilia de ATPasas AAA+ que transforman la energía en una función mecánica para remodelar sus sustratos. La activación de la transcripción de los genes de fijación dependientes de σ^{54} requiere de NifA, la cual es una EBP conservada que regula la síntesis de los genes como la nitrogenasa en las proteobacterias. Las secuencias de ADN responsables de la activación a

distancia de los promotores de numerosos genes se denominan UAS (del inglés "Upstream Activation Sequences") que denominamos secuencias "enhancers", potenciadoras o activadoras. Este tipo de secuencias son una característica común de los genes eucarióticos y virales. En los organismos procarióticos existen, aunque no son frecuentes, secuencias de ADN que guardan una elevada similitud con las secuencias potenciadoras de los organismos eucarióticos. Normalmente se trata de secuencias palindrómicas situadas entre 100 y 200 pb cadena arriba de promotores característicos (Kustu *et al*, 1989; Thóny y Hennecke, 1989) que son capaces de estimular la transcripción aún cuando se localizan varios cientos de pb cadena arriba y, en algunos casos cadena abajo del inicio de transcripción, con independencia de la orientación (Belitsky y Sonenshein, 1999; Buck *et al*, 1986; Reitzer y Magasanik, 1986), esto sin afectar en la selección del inicio de transcripción (Molina-López y Santero, 1999).

R. etli cuenta con dos factores σ^{54} , uno localizado en cromosoma que se expresa en vida libre y está involucrado en la síntesis de flagelo, asimilación de carbono y nitrógeno, entre otras funciones, y otro en el plásmido simbiótico que participa de manera activa en la simbiosis, está en operón con el gen *prxS* (una peroxiredoxina). Una mutante en el gen *rpoN* disminuye la fijación de nitrógeno hasta en un 90%, la mutante en *prxS* tiene una disminución de alrededor del 40% (Dombrecht *et al.*, 2005). El gen *nifA* se localiza en el plásmido simbiótico, es vecino del operón *fixABCX* y tiene la misma orientación para la expresión. Los blancos reportados para RpoN-NifA son los genes *nifHDK*, *fixABC*, *melA*, *bacS*, *iscN*, todos a excepción de *melA* y *bacS* han sido descritos como participantes activos de la fijación de nitrógeno (Hawkins y Johnston, 1988; Valderrama *et al.*, 1996; Dombrecht *et al.*, 2002; Janh *et al.*, 2003; Moris *et al.*, 2004; Dombrecht *et al.*, 2005). La mutante *nifA* no fija nitrógeno, lo que indica que esta EBP es única en el genoma de *R. etli* (Valderrama *et al.*, 1996).

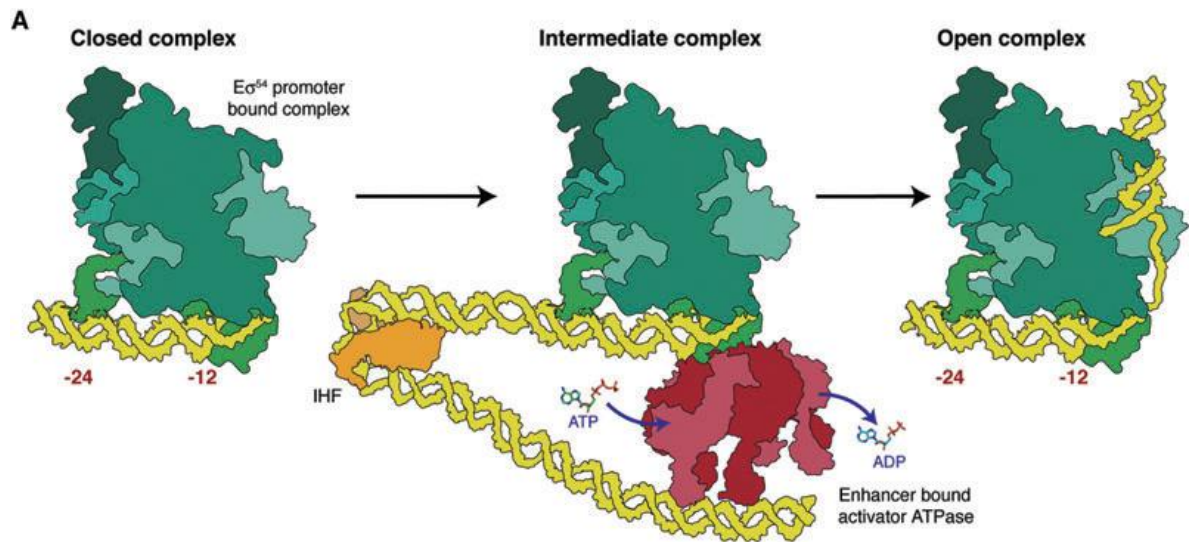


Figura 3. Representación de la participación de $E\sigma^{54}$, IHF y una EBP en la apertura de un complejo cerrado a uno abierto, tomada de Wigneshweraraj *et al.*, 2008.

6. MICROARREGLOS Y GENÓMICA FUNCIONAL

Desde que la secuencia del genoma de *Haemophilus influenzae* fue publicada en 1995, se han reportado a la fecha 3000 genomas completos de microorganismos (2847 de bacterias y 153 de arqueas), además de 173 genomas eucariotas (<http://www.genomesonline.org/>; www.ncbi.nlm.nih.gov/Genomes). La genómica comparativa presumiblemente puede reconstruir redes metabólicas entre los diferentes organismos, es decir, se puede inferir una reacción a partir de los datos obtenidos en un organismo diferente, pero para el caso de redes de regulación transcripcional la extrapolación se obstaculiza debido al bajo nivel de conservación evolutivo entre los reguladores transcripcionales (Herrgard *et al.*, 2004). Para enfrentar la dificultad de entender de manera global la regulación de los genes, así como entender la posible función de nuevos genes se requieren de tecnologías especiales que proporcionen información que faciliten el análisis de la regulación de la expresión llevada a cabo por los factores sigma, por factores transcripcionales o por condiciones ambientales determinadas. Una de las metodologías que destaca son los microarreglos de ADN, esta metodología se utiliza primordialmente para la medición de la expresión de RNA, aunque también se han utilizado para la comparación de genomas, detección de genes de virulencia, entre otras aplicaciones. Para la medición de los RNAs mensajeros se han

utilizado diferentes técnicas, iniciando la historia con el “Northern blot” en 1977. Seguido por los ensayos de protección a nucleasas (NPA’s), ensayos de RT-PCR, la expresión diferenciada (differential display), los análisis de expresión en serie (SAGE), alcanzando a los microarreglos (Affymetrix, oligonucleótidos y los “tiling arrays”) y finalmente el RNA seq (Güell *et al.*, 2011). Cada técnica tiene sus virtudes y defectos, que pasan por la sensibilidad, la estabilidad del mensajero, el manejo de los datos, la interpretación de los resultados, entre otras, incluso se considera el costo de cada metodología. En esta tesis nos enfocamos al uso de microarreglos para la medición global de los mensajeros en diferentes condiciones y en diferentes cepas. Los microarreglos, como las otras metodologías tienen defectos y ventajas, a continuación señalaré las ventajas que tiene esta metodología sobre otras técnicas de medición de mensajeros. Los microarreglos o ‘chips de DNA’ son portaobjetos, “slides” de vidrio o siliconas, que contienen miles de moléculas distintas de DNA en un orden determinado. Son colecciones de segmentos de DNA fijados a una superficie sólida y se utilizan en la cuantificación de DNA o RNA que provienen de muestras biológicas. El primer microarreglo que se publicó fue para 45 genes de *Arabidopsis thaliana* en 1995 (Schena *et al.*, 1995), desde ese primer reporte se han publicado miles de trabajos de este tipo en bacterias, arqueas y eucariotes. Los microarreglos duales, nos permiten analizar en un solo experimento la expresión de 2 condiciones o cepas de manera global, a diferencia del RT-PCR o RNAseq, no requerimos que nuestra muestra esté completamente libre de DNA ya que la enzima retrotranscriptasa utilizada para etiquetar el cDNA usa como templado el RNA. Otra ventaja es que al hibridar 2 condiciones o cepas en un mismo portaobjetos, estamos controlando las variantes de la hibridación, lavados, entre otros, facilitando con ello la comparación entre estas. Ni el RNA ni cDNA requieren de separaciones posteriores a su obtención, con lo que evitamos la posible degradación. Respecto al uso de la técnica de RNA seq, una de las ventajas de los microarreglos, en nuestro principal modelo de estudio la simbiosis, es que podemos etiquetar a partir de RNA total (planta y bacteria) y la especificidad la aportan los oligonucleótidos impresos en el portaobjetos, a diferencia del RNA seq donde las secuencias obtenidas representan en una mayor proporción al RNA de la planta; además,

si se quiere obtener un RNA que provenga mayoritariamente de la bacteria se necesitan métodos adicionales de purificación que pueden degradar la muestra y disminuir el rendimiento, obteniendo resultados probablemente sesgados.

7. ANTECEDENTES

Nuestro modelo de estudio, la bacteria *R. etli* CFN42, fue secuenciada en México, específicamente en el Centro de Ciencias Genómicas por el grupo del Dr. Guillermo Dávila (González *et al.*, 2006). Esta bacteria de importancia agronómica, tiene un genoma de 6.5 Mb y una anotación de 6034 ORF's, de los cuales 354 pertenecen al plásmido D o simbiótico (pSim), en este plásmido se localizan los genes que participan en la nodulación y en la fijación de nitrógeno; sin embargo, también se ha establecido que otros replicones, como el plásmido F, son necesarios para el completo establecimiento de la fijación de nitrógeno. El gene *nifA* de *R. etli* CFN42 está situado en el plásmido simbiótico al igual que uno de los dos genes *rpoN* presentes en esta bacteria (González *et al.*, 2003). Similar a nuestro modelo (*R. etli* CFN42), el genoma de *R. etli* CNPAF512 tiene dos genes *rpoN* que codifican para el factor sigma alternativo σ^{54} , estos se regulan diferencialmente durante la simbiosis y el crecimiento en vida libre (Michiels *et al.*, 1998). Durante el crecimiento en vida libre, el producto del gen *rpoN1*, codificado en el cromosoma, es requerido para el crecimiento en varias fuentes de carbono y nitrógeno (Michiels *et al.*, 1998). En cambio RpoN2 codificado en el pSim, es indispensable para la activación de los genes que participan en la fijación de nitrógeno; una mutante en *rpoN2* disminuye hasta un 90% la fijación de nitrógeno (Michiels *et al.*, 1998). En *R. etli* CNPAF512, la proteína NifA (una EBP), controla la activación de la transcripción de *rpoN2* bajo condiciones microaeróbicas en vida libre y durante la simbiosis (Dombrecht *et al.*, 2005). Diversas estrategias han sido utilizadas para identificar los genes regulados por RpoN-NifA, por ejemplo, un análisis *in silico* del plásmido simbiótico de *R. etli* CFN42 y otros rizobios secuenciados, identificó probables promotores para RpoN, en total en el pSim se detectaron 24 sitios que controlan 52 genes (Dombrecht *et al.*, 2002). Con trabajos previos en *R. etli* se logró identificar una serie de elementos que participan en la transcripción de genes importantes

para la simbiosis y definido como funcionan en la regulación de sus genes blanco; el grupo de la Dra. Girard se han caracterizado los reguladores tipo CRP/FNR, identificando sitios de unión que contienen las regiones promotoras de los genes blancos y caracterizando cascadas de regulación que involucran este tipo de genes (Granados-Baeza *et al.*, 2007; Girard *et al.*, 2000). Adicionalmente, el genoma de *R. etli* CFN42 cuenta con 479 genes (7.93% de los ORF's) involucrados en la transcripción, en *E. coli* K12 esta clase funcional representa solo el 5.4% con 241 genes. El gran número de genes en *R. etli* cuyos productos participan en la regulación genética nos indica el amplio panorama de esta bacteria que aún nos falta por elucidar.

Metodologías más directas también se han utilizado para delimitar regulones, sigmulones y estimulones, entre estas se encuentran los microarreglos de ADN. El primer microarreglo para una bacteria fijadora se construyó en *S. meliloti*, en el cual se analizaron 34 genes de regulación conocida, el ensayo se realizó en condiciones de microaerobiosis (Berges *et al.*, 2003); el mismo año se utilizaron 214 genes seleccionados por su probable participación en la simbiosis, para detectar su expresión en diferentes condiciones biológicas incluyendo la simbiosis (Ampe *et al.*, 2003), posteriormente en el 2004 se presentó el microarreglo completo de esta bacteria, 6205 genes representados por oligonucleótidos de 70 bases (Becker *et al.*, 2004). En el 2006, se presentó el microarreglo de *B. japonicum*, utilizado para identificar los genes en microaerobiosis dependientes de NifA (Hauser *et al.*, 2006). Los últimos microarreglos reportados de rizobios fueron para *R. leguminosarum* (Karunakaran *et al.*, 2009) y *R. etli* CFN42 (Martinez-Salazar *et al.*, 2009), para este último también fue diseñado un microarreglo del tipo "tiling" que cubre todo el genoma de *R. etli* CFN42, y su primer reporte se basó en la identificación de microRNAs (Michiels *et al.*, 2010). En lo que respecta a la delimitación de regulones en este tipo de bacterias, uno de los primeros trabajos fue realizado en *S. meliloti* con la mutante *fixJ*, cuyo producto es un regulador fundamental en la simbiosis, en el mismo trabajo los autores delimitaron el regulón NifA a 19 elementos (Bobik *et al.*, 2006). En *B. japonicum*, fue descrito mediante esta metodología el regulón RegS-RegR, que a su vez contiene el regulón NifA-RpoN, compuesto por 65 blancos, principalmente genes cuyos productos participan en la fijación

de nitrógeno y otros procesos, el regulón fue definido en microaerobiosis y en anaerobiosis, ambas condiciones de vida libre (Bobik *et al.*, 2006; Hauser *et al.*, 2007). En *R. etli* CFN42 en el 2011 definieron los genes afectados en una mutante *relA* que codifica para la alarmona (ppGpp) (Vercruyse *et al.*, 2011).

Se ha realizado trabajos previos en diferentes tiempos de la simbiosis con la finalidad de observar la expresión genética de los genes durante el desarrollo del nódulo. Dos de este tipo de estudios se realizaron en nódulos indeterminados. En *R. leguminosarum* se analizaron los tiempos 7, 15, 21 y 28 ddi, la expresión de los genes de fijación de nitrógeno fueron detectados en todos los tiempos pero con un incremento en los nódulos maduros (15, 21 y 28 ddi), en total determinaron la expresión diferencial de 386 genes repartidos de acuerdo a los niveles y tiempo de expresión en 9 grupos, a su vez estos grupos fueron reagrupados en 2 grandes grupos, uno que incluye los genes que se expresaron en una etapa temprana/intermedia y otro que representa a los genes expresados en nódulos maduros (Karunakaran *et al.*, 2009). En *S. meliloti* 1021, que también presenta nódulos indeterminados en plantas de alfalfa (*Medicago sativa*), se analizaron nódulos de los tiempos 5, 8, y una mezcla de nódulos de 14 a 18 ddi, además los de una mutante en el gen *bacA* a los 11 ddi, la mutación de este gen afecta la fijación de nitrógeno desde tiempos tempranos. Al igual que en *R. leguminosarum*, detectaron 2 grandes grupos de genes, en uno ubican a los genes expresados en nódulos de 5 ddi y los de la mutante *bacA*, en el otro se agrupan los genes que alcanzan su máxima expresión en nódulos de 8 y 14 a 18 ddi, en total reportan 235 genes divididos en 7 grupos, aunque 3 de ellos sólo contienen 1 o 2 elementos por grupo; los genes *fix* y *nif* se localizan en su mayoría en el grupo 4, y se expresan a los 8, 14-18 ddi (Capela *et al.*, 2006). En *B. japonicum*, bacteria que induce nódulos determinados en asociación con soya, se reportaron estudios de los tiempos 10, 13, 21 y 31 ddi, obteniéndose 3 grupos de genes: los sobreexpresados a los 10 y 13 ddi, los sobreexpresados a los 21 y 31 ddi; y los sobreexpresados a los 13 ddi, sin embargo, no determinaron los genes reprimidos en los tiempos 10, 21 y 31 ddi respecto a los 13 ddi, en dicho trabajo presentan 191 genes que modifican su expresión en los tiempos analizados (Pessi *et al.*, 2007).

R. etli CFN42, produce nódulos determinados en asociación con plantas de frijol, planta de gran importancia agrícola, sobre todo para los países en vías de desarrollo. En este trabajo diseñamos y construimos el microarreglo de *R. etli* CFN42 con la intención de: i) definir el regulón RpoN-NifA, los cuales son reguladores primordiales para la activación de los genes de la fijación de nitrógeno, el análisis se realizó en vida libre (microaerobiosis) y en la simbiosis (11 ddi), ii) presentamos el transcriptoma de *R. etli* comparando la simbiosis y la vida libre; iii) finalmente, con el propósito de observar el desarrollo del nódulo y su regulación, se estudió el transcriptoma en diferentes tiempos de la simbiosis con *P. vulgaris*.

8. OBJETIVO GENERAL

El objetivo de este trabajo es analizar de manera global la expresión del genoma de *R. etli* CFN42 en condiciones de vida libre y durante la simbiosis con *P. vulgaris* L. Con lo anterior se obtendrá la información que nos permita proponer los tipos de mecanismos de regulación presentes en esta bacteria, así como adquirir el conocimiento de los eventos metabólicos que se presentan en la bacteria en la vida libre pero principalmente en la simbiosis.

Objetivos particulares

Diseñar y construir el microarreglo de *R. etli* CFN42 que contenga cada uno de los marcos de lectura abiertos anotados para este genoma.

Delimitar el regulón RpoN-NifA de *R. etli* CFN42 en vida libre y durante la simbiosis utilizando herramientas de genómica funcional.

Analizar la expresión del genoma de *R. etli* mediante el uso de microarreglos comparando la vida libre y la simbiosis.

Obtener y analizar los perfiles de expresión en diferentes etapas (tiempos) de la fijación de nitrógeno y en la vida libre. Así mismo, identificar y analizar los sitios de regulación para cada perfil.

9. MATERIAL Y METODOS ADICIONALES A LOS PRESENTADOS EN LOS ARTICULOS

Tinción para determinar viabilidad celular

Para determinar la viabilidad celular de las bacterias *R. etli* CFN42 y la mutante *R. etli* CFNX247 (*nifA*-) en el nódulo, se utilizó el método “Live/Dead BacLight (BL) Viability kit” (Molecular Probes Inc.). Este método utiliza dos fluorocromos, el SYTO 9 que da una tinción verde (bacteria viable) y el yoduro de propidio o PI, que genera una tinción roja (célula muerta o dañada y por lo tanto tiene una membrana permeable). Para observar los nódulos se utilizó un microscopio confocal de la Unidad de microscopía del Instituto de Biotecnología de la UNAM.

Análisis de los microarreglos

Para el análisis, los datos fueron divididos en 3 grandes grupos: Grupo 1, datos analizados utilizando el programa GenArise desarrollado en el Instituto de Fisiología Celular, UNAM (<http://www.ifc.unam.mx/genArise/>); los datos de este grupo incluyen la expresión de *R. etli* CFN42 comparada con *R. etli* CFNX247 (*nifA*-) en ambas condiciones, simbiosis 11 ddi y vida libre en medio mínimo (MM) 10 hrs en bajo oxígeno (BO). El mismo programa utilizamos para procesar los datos de *R. etli* CFN42 comparada con *R. etli* pJMS37 (RpoE4-) y con *R. etli* pFAJ1708 (sobreexpresión RpoE4). El método que el programa utiliza se describe en ambos artículos (Martínez-Salazar *et al.*, 2009 y Salazar *et al.*, 2010). De cada condición se analizaron tres réplicas experimentales, con su respectiva inversión de cianinas o “swap”, cada réplica fue analizada de manera independiente y al final solo presentamos los genes que coincidieron en las 3 réplicas utilizadas. El grupo 2 incluye los datos de la comparación entre *R. etli* CFN42 en vida libre cultivado en MM y colectado a las 6 hrs de cultivo (MM 6h), comparado con *R. etli* CFN42 de nódulos a los 18 días después de inocular (ddi), este grupo fue analizado con el programa R2.8.1 junto con BioConductor (<http://www.bioconductor.org>). El detalle del análisis se encuentra en el artículo No. 2 en el cual se presentan resultados de esta tesis titulado “Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling”, a diferencia del análisis para el grupo 1 en este grupo se

analizaron las réplicas como un conjunto. Finalmente, el grupo 3 corresponde a los datos de los diferentes tiempos de colecta de los nódulos (11, 18, 25, 28 y 35 ddi), además de una condición de vida libre MM 6h, los datos fueron analizados de manera independiente, es decir, de las 3 réplicas de cada condición primero se filtraron por semejanza al tomar para cada punto solo las 2 réplicas experimentales que tuvieran la correlación más cercana a 1 (Gosset *et al.*, 2004). A partir de estas réplicas se obtuvo un promedio, en este punto los datos fueron normalizados entre los 5 tiempos de la simbiosis y la condición de vida libre utilizando la fórmula de la media (Kroll y Wölfl, 2002). Posteriormente los datos normalizados fueron ordenados por expresión y se eligieron para cada tiempo sólo los genes que presentaran una señal de 3 veces de cambio respecto al fondo (Karunakaran *et al.*, 2009; Capela *et al.*, 2006). Estos datos fueron los utilizados para comparar entre los diferentes tiempos, obteniendo un valor de log₂, finalmente el listado de genes con su valor de log₂ se analizó con un agrupamiento tipo “K-means” utilizando el programa “Cluster v3.0” y para visualización utilizamos el programa “java TreeView” ambos de libre acceso y desarrollados por el grupo del Dr. Michael Einsen. Los datos de los 5 tiempos de la simbiosis y el tiempo de la vida libre, así como la comparación entre vida libre y el tiempo de 18 ddi, también se ordenaron por expresión de mayor a menor, para ser analizados posteriormente con el programa Gene Set Enrichment Analysis (GSEA, Broad Institute, Cambridge, MA). Para utilizar este programa en *R. etli* se requiere agregar manualmente el grupo de genes que corresponden a cada vía, en nuestro modelo la base de datos que utilizamos fue la “Kyoto Encyclopedia of Genes and Genomes” (KEGG, www.genome.jp/KEGG), en ésta se indica que *R. etli* CFN42 cuenta con 108 vías que incluyen 3096 genes, aunque debemos aclarar que en ocasiones un gen puede estar en más de una vía. El ambiente de GSEA es gráfico y si los datos son concluyentes da un valor de certidumbre “FDR” y un “p-value” (Subramanian *et al.*, 2005).

Búsqueda sitios de regulación

Para la búsqueda de los sitios de regulación se utilizaron principalmente 2 programas, el MEME-Mast cuya metodología se detalla en el artículo “Characterization of the NifA-RpoN regulon in *R. etli* in free life and in symbiosis with *Phaseolus vulgaris*”. El otro programa

que se utilizó fue el “Regulatory Sequence Analysis Tools (RSA-Tools)” disponible de manera libre en <http://embnet.ccg.unam.mx/rsa-tools/> (van Helden, 2003). Para identificar los probables sitios de unión en las regiones promotoras de los genes que modificaron su expresión se utilizaron diversas matrices de reguladores conocidos para rizobios, por ejemplo NifA, RpoN, Fnr, cajas Nod, RpoE y otros. Pero también se realizó una búsqueda para identificar sitios en genomas recién secuenciados y con poca información de sitios de regulación basados en la referencia de Li y colaboradores del 2002, la técnica se basa en la búsqueda de dos segmentos cortos que representan los sitios más conservados que intervienen en la unión, esto sin la necesidad de obtener grupos de genes co-regulados.

10. RESULTADOS

10.1. Artículo: “Characterization of the NifA-RpoN regulón in *Rhizobium etli* in free life and in symbiosis with *Phaseolus vulgaris*”

En esta sección presentamos un artículo publicado en la revista “Applied and Environmental Microbiology”. En este presentamos la construcción del microarreglo de oligonucleótidos para *R. etli* CFN42. En el trabajo definimos mediante herramientas de genómica funcional el regulón RpoN-NifA de *R. etli* CFN42.

Characterization of the NifA-RpoN Regulon in *Rhizobium etli* in Free Life and in Symbiosis with *Phaseolus vulgaris*^{∇†}

Emmanuel Salazar,¹ J. Javier Díaz-Mejía,^{2,3} Gabriel Moreno-Hagelsieb,³ Gabriel Martínez-Batallar,¹ Yolanda Mora,¹ Jaime Mora,¹ and Sergio Encarnación^{1*}

Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México¹;
Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, 160 College St.,
Toronto, Ontario M5S-3E1, Canada²; and Department of Biology, Wilfrid Laurier University,
Waterloo, Ontario, Canada³

Received 20 August 2009/Accepted 20 April 2010

The NifA-RpoN complex is a master regulator of the nitrogen fixation genes in alphaproteobacteria. Based on the complete *Rhizobium etli* genome sequence, we constructed an *R. etli* CFN42 oligonucleotide (70-mer) microarray and utilized this tool, reverse transcription (RT)-PCR analysis (transcriptomics), proteomics, and bioinformatics to decipher the NifA-RpoN regulon under microaerobic conditions (free life) and in symbiosis with bean plants. The *R. etli* NifA-RpoN regulon was determined to contain 78 genes, including the genes involved in nitrogen fixation, and the analyses revealed 42 new NifA-RpoN-dependent genes. More importantly, this study demonstrated that the NifA-RpoN regulon is composed of genes and proteins that have very diverse functions, that play fundamental and previously less appreciated roles in regulating the normal physiology of the cell, and that have important functions in providing adequate conditions for efficient nitrogen fixation in symbiosis. The *R. etli* NifA-RpoN regulon defined here has some components in common with other NifA-RpoN regulons described previously, but the vast majority of the components have been found only in the *R. etli* regulon, suggesting that they have a specific role in this bacterium and particular requirements during nitrogen fixation compared with other symbiotic bacterial models.

Rhizobium etli is a nitrogen-fixing soil bacterium that is able to form a root nodule symbiosis with leguminous plants, specifically *Phaseolus vulgaris* (common bean), one of the most important crops in Mexico and Latin America (51). The *R. etli*-*P. vulgaris* symbiosis is a complex and biologically important relationship in which the bacterium fixes atmospheric nitrogen and the plant provides carbon compounds and protection against some environmental stresses. The detailed mechanisms that allow this interaction are not completely understood (7). In *R. etli* CFN42 and other nitrogen-fixing symbiotic bacteria, NifA-RpoN is a master regulator of nitrogen fixation genes, and NifA plays a central role in ensuring expression of the nitrogen fixation apparatus during symbiosis. The NifA protein belongs to the enhancer-binding protein family of transcriptional regulators that activate gene expression in concert with RNA polymerase containing the specialized sigma factor σ^{54} (RpoN), which allows the polymerase core to recognize $-24/-12$ promoters (14).

The *R. etli* CFN42 *nifA* gene is located on the symbiotic plasmid (pCFN42d) (23). Similar to *R. etli* CFN42, *R. etli* CNPAF512 has two *rpoN* genes encoding the alternative sigma factor σ^{54} , which are differentially regulated during symbiosis and free-living growth (40). During free-living growth RpoN1, encoded on the chromosome, is required for growth on several

nitrogen and carbon sources (41). When *rpoN2*, which is located on the symbiotic plasmid, is inactivated, there is a sharp decrease in nitrogen fixation (40), indicating that this gene has an essential role in bacteroids (16, 17). The NifA enhancer-binding protein controls transcriptional activation of *rpoN2* under free-living microaerobic conditions and during symbiosis (15). The *rpoN2* gene of *R. etli* CNPAF512 is orthologous to the *R. etli* CFN42 *rpoN2* (*rpoNd*) gene, a gene located on pCFN42d.

In symbiotic diazotrophs, such as *Sinorhizobium meliloti*, transcription of *nifA* and *fix* genes is controlled predominantly by the oxygen-responsive two-component FixL-FixJ system. In *S. meliloti* NifA, along with FixK, controls 19 symbiotic targets (6). In *Bradyrhizobium japonicum* RegS-RegR controls the NifA-RpoN regulon, which regulates 65 targets for nitrogen fixation and other processes in microaerobiosis and anaerobiosis (free-living conditions) (25, 26). In *Rhizobium leguminosarum*, *nifA* is autoregulated (37), while in *Azorhizobium caulinodans* it is under the direct control of FixK (32).

In *R. etli*, *nifA* gene expression has features markedly distinct from the features in other rhizobia, and no genetic elements involved in its regulation have been identified. However, NifA-dependent gene activation occurs only at low oxygen concentrations (43, 55). To date, only a few targets, including *nifHDK*, *iscN*, *nifUS*, *fixABC*, *prxS-rpoN2*, *melA*, and *bacS*, have been reported to be members of the *R. etli* NifA-RpoN2 regulon during symbiosis with *P. vulgaris* (15, 17, 27, 31, 43, 55).

Using the *R. etli* CFN42 genome sequence (24), we designed and constructed a microarray to analyze transcription of the NifA regulon of *R. etli* during symbiosis. The microarray chip contains 70-base (70-mer) oligonucleotides representing all

* Corresponding author. Mailing address: Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. Phone: 52-7773-29-18-99. Fax: 52-7773-17-50-94. E-mail: encarnac@ccg.unam.mx.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

[∇] Published ahead of print on 7 May 2010.

TABLE 1. Oligonucleotide set selection criteria and characteristics of the *R. etli* CFN42 microarray

Oligonucleotide selection criterion	Value	No. (%) of oligonucleotides in genome set satisfying the criterion
Length	70 bases	6,034 (100)
Melting temp	80 ± 5°C	6,034 (100)
Level of cross-hybridization identity to all other CDSs	≤70%	5,886 (97.55)
No. of contiguous base matches with all other CDSs	≤20	5,994 (98.51)
Distance from 5' and 3' ends	>50 bases	6,014 (99.67)
Stem length in potential hairpin	≤7 bases	6,034 (100)
Total		6,034 (100)

6,034 open reading frames (ORFs) that have been detected in the genome sequence of this organism.

The main goal of this work was to identify and more strictly define the *R. etli* NifA-RpoN regulon. This was done by comparing wild-type strain CFN42 with a knockout mutant lacking the NifA regulatory protein grown under microaerobic free-living conditions and during symbiosis. This study is the first study of *R. etli* that used global approaches, such as transcriptomics (DNA microarray and real-time reverse transcription [RT]-PCR), in combination with proteomics and bioinformatics. The regulon described here includes some genes that are shared with other NifA-RpoN regulons. However, more of the components found in this work have been found only in the *R. etli* CFN42 regulon and have specific functions in this species.

MATERIALS AND METHODS

Design and construction of the *R. etli* microarray. The *R. etli* complete array oligonucleotide set contains 6,034 arrayable 70-mers representing the genome of strain CFN42 and was designed using the following NCBI reference sequences: NC 007761 (chromosome), NC 007761 (pCFN42f), NC 007765 (pCFN42e), NC 007764 (pCFN42c), NC 007763 (pCFN42b), NC 007762 (pCFN42a), and NC 004041 (pCFN42d) (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Rhizobium_etli_CFN_42/).

The 70-mer oligonucleotides were designed by using several *ad hoc* programs written in PERL and other software, as follows. The first program read the DNA sequences corresponding to each of the annotated genes of the *R. etli* CFN42 genome. If the size of the gene permitted, the program then cut 50 bases from both the 5' and 3' ends. After this, the program scanned, base by base, each possible 70-mer of the sequence to ensure that there was no single-base repeat longer than 7 bp and the melting temperature (T_m) was 80 ± 5°C (the formula used for calculating the T_m was $T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times [G+C]/\text{length} - 500/\text{length}$, where $[\text{Na}^+]$ was 0.1 M and the length was 70). The resulting candidate oligonucleotides were compared with the complete genome sequence using NCBI BLASTALL (BLASTALL-PBLASTN) (1) with a maximum e value of 1.0 (−e 1.0) and no masking of low-information segments (−F F). The BLAST results were used to ensure that no candidate oligonucleotide could cross-hybridize in the microarray. We used the program PALINDROME from the EMBOSS software suite (44, 49) to find hairpins and filter out any candidate oligonucleotides with potential hairpins more than 9 bases long. Our programs also ensured that no oligonucleotide had more than 20 bases in common with any other coding sequence (Table 1).

When two or more oligonucleotide sequences for the same ORF satisfied all of the selection rules, the oligonucleotide was selected with the following parameters: a T_m close to 80°C and minimum cross-hybridization identity. A control that included 4 positive controls and 12 negative controls was also designed using the *Drosophila melanogaster* genome. All oligonucleotides were commercially synthesized without modification by MWG-Biotech (Ebersberg, Germany), and 50 μM solutions were prepared using Micro Spotting solution (ArrayIt Brand Products) and spotted in duplicate onto SuperAmine-coated slides (25 by

75 mm; TeleChem International, Inc.) using 24 columns and 22 rows per grid (the total number of grids was 24) using a high-speed robot at the microarray facility at Instituto de Fisiología Celular, Universidad Nacional Autónoma de México. Finally, the slides were fixed at 80°C for 4 h. For prehybridization, the slides were rehydrated with water vapor at 60°C and fixed with two cycles of UV light (1,200 J). After 2 min of boiling at 92°C, the slides were washed with 95% ethanol for 1 min and prehybridized in 5× SSC, 0.1% SDS, 1% bovine serum albumin (BSA) for 1 h at 42°C (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). The slides were washed and dried prior to hybridization (35, 50).

Bacterial strains and growth conditions. The bacterial strains used were *R. etli* wild-type strain CFN42 (24) and mutant strain CFN4247 ($\Delta nifA::\Omega\text{Sp}/\text{Sm}$) (22). For growth under microaerobic conditions, cultures were incubated in 150-ml bottles that were closed with an airtight stoppers and flushed with several volumes of an oxygen-argon (1:99, vol/vol) mixture. The cultures were grown with shaking (200 rpm) for 10 h at 30°C as previously described by Girard et al. (21, 22).

Plant experiments. Three-day-old *P. vulgaris* cv. Negro Jamapa seedlings were inoculated with *R. etli* CFN42 or the *nifA* mutant strain as previously described by Peralta et al. (45). Eleven days postinoculation (dpi), nodules were picked from the roots and immediately frozen in liquid nitrogen and stored at −70°C until they were used. Bacteria were isolated from nodules, and their identities were verified by using antibiotic resistance.

RNA isolation, synthesis of labeled cDNA, and microarray hybridization. Microarray experiments were carried out using three independently isolated RNA preparations from independent cultures and sets of plants. Approximately 3 g of nodules was immersed in liquid nitrogen and macerated. Total RNA was isolated by acid hot phenol extraction as described previously by de Vries et al. (13). For microaerobic free-living conditions, 50-ml portions of bacterial cell cultures were collected, and total RNA was isolated using an RNeasy minikit (Qiagen, Hilden, Germany). The RNA concentration was determined by measuring the absorbance at 260 nm. The integrity of RNA was determined by running samples on a 1.3% agarose gel. Ten micrograms of RNA was differentially labeled with Cy3-dCTP and Cy5-dCTP using a CyScribe first-strand cDNA labeling kit (Amersham Biosciences). Pairs of Cy3- and Cy5-labeled cDNA samples were mixed and hybridized with the array as described by Hegde et al. (28). After the arrays were washed, they were scanned using a pixel size of 10 μm with a Scan Array Lite microarray scanner (Perkin-Elmer, Boston, MA).

DNA microarray analysis. Spot detection, determination of mean signals and mean local background intensities, image segmentation, and signal quantification were performed for the microarray images using the Array-Pro Analyzer 4.0 software (Media Cybernetics, L.P.). Microarray data were analyzed with the genArise software, which was developed in the Computing Unit of the Instituto de Fisiología Celular, Universidad Nacional Autónoma de México (<http://www.ifc.unam.mx/genarise/>). The local background value was subtracted from the intensity of each spot, and Lowess normalization was applied to the slide. The normalized ratio of the expression of the experimental sample to the expression of the control was calculated with genArise software for each experimental replicate. This software identifies differentially expressed genes by calculating an intensity-dependent z score. It uses a sliding window algorithm to calculate the mean and standard deviation within a window surrounding each data point and determines a z score which measures the number of standard deviations that a data point is from the mean: $z_i = [R_i \cdot \text{mean}(R)]/\text{sd}(R)$, where z_i is the z score for each element, $\text{mean}(R)$ is the mean log ratio, R_i is the log ratio for each element, and $\text{sd}(R)$ is the standard deviation of the log ratio. Using this criterion, the elements in all experiments with a z score of >2 standard deviations were considered significantly differentially expressed genes. To eliminate dye effects, labeling with different dyes was performed for single replicates. However, no significant differences between Cy3 and Cy5 in incorporation, fluorescence yield, or stability were observed.

The data sets for all genes that were significantly downregulated in the NifA mutant strain, including all expression level ratios, are shown in the supplemental material. The complete data set for the transcriptome analysis has been deposited in the GEO database (<http://www.ncbi.nlm.nih.gov/geo.html>) under accession number GPL10081 for the platform, accession number GSE20440 for the microaerobiosis data, and accession number GSE2063 for the symbiosis data.

In silico analysis. In order to identify putative regulatory binding sites in the upstream regions (UTRs) of the coding sequences (CDSs) of the *R. etli* genome, we compared a set of regulatory binding sites previously identified for NifA and RpoN in *R. etli* and other nitrogen-fixing symbiotic bacteria (4, 23, 40, 42) with the UTRs of *R. etli*. The MEME v 3.5.0 program (2) was used to find motifs (sequence patterns that occurred repeatedly in a group of sequences) in the regulatory binding sites and to construct search matrices, which represented the probability that each possible letter was at each position in the pattern. With this

set of matrices, we searched for putative regulatory binding sites in the UTRs (from bp -500 to bp 200 with respect to the ATG transcriptional start codon) of *R. etli*, using the MAST v 3.5.0 program (2, 3). The search was conducted iteratively by adding the discovered putative regulatory binding sites to the previously described binding sites to construct a new set of matrices and repeating the search until there was convergence (when no more putative regulatory binding sites were found).

Proteomics methodology. Bacteroids were purified from root nodules by centrifugation through self-generated Percoll gradients (48). Bacteroid proteins were obtained by sonication at 24 kHz for 5 cycles consisting of 1 min on and 1 min off at 4°C in a Vibra Cell (Sonics, United States) in the presence of a 106-protease inhibitor (Complete tablets; Roche Diagnostics GmbH, Mannheim, Germany). To further limit proteolysis, protein isolation was performed using phenol extraction (30). To solubilize and obtain completely denatured and reduced proteins, pellets were dried and resuspended as previously reported (19). Prior to electrophoresis samples were mixed with 7 M urea, 2 M thiourea, 4% 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 2 mM tributyl phosphine (TBP), 2% ampholytes, and 60 mM dithiothreitol (DTT). The methods used for sample preparation, analytical and preparative two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE), and image analysis have been described previously (18), and pH gradients were determined by using a 2D SDS-PAGE standard (Sigma, United States). For the first dimension approximately 500 µg of total protein was loaded. The gels were stained with Coomassie blue R-250, and protein spots on all of the gels were detected at a resolution of 127 by 127 µm using a PDI image analysis system and PD-Quest software (Protein Databases, Inc., Huntington Station, NY). We were interested in spots that showed at least a 2-fold change and met the conditions of a statistical Student test (level of significance, 95%). Selected spots from Coomassie blue-stained preparative 2D gels were excised manually and prepared for mass spectrum analysis (19). All experiments were performed three times. Mass spectra were obtained using a Bruker Daltonics Autoflex (Bruker Daltonics, Billerica, MA) operated in the delayed extraction and reflectron mode. Spectra were externally calibrated using a peptide calibration standard (Bruker Daltonics 206095). Peak lists of the tryptic peptide masses were generated and searched against the NCBI nr databases or with Rhizobase (<http://bacteria.kazusa.or.jp/rhizo/>) using the Mascot search program (Matrix Science, Ltd., London United Kingdom).

Real-time RT-PCR. We used real-time quantitative PCR to obtain an independent assessment of the expression of selected genes. The cDNA used for microarrays or freshly prepared cDNA was used as a template for real-time PCR. The following primers were used: *nifHa*-RE1SP00308f (5'-GGG CAG AAG ATC CTG ATC GT-3') and *nifHa*-RE1SP00308r (5'-ATC TCC TGG GCC TTG TTC TC-3') for the *nifH* gene; *nifKb*-RE1SP00200f (5'-CCG GAA TAC AAG CAG ATG CT-3') and *nifKb*-RE1SP00200r (5'-CTC CTT GAA GTG ACG CGA CA-3') for the *nifK* gene; *rpoN*-RE1SP00218f (5'-CCA GCT CAT CGC GTC TAT TC-3') and *rpoN*-RE1SP00218r (5'-CGA AAG GTT GCC TGT GTC AC-3') for the *rpoN2* gene; *yp017*-RE1SP00100 (5'-GAC GAC CGA CGA CTA TTT CA-3') for the *prxS* gene; *virB4*-RE1SP00146f (5'-GGC GTT TCC TTC GAG ACA TC-3') and *virB4*-RE1SP00146r (5'-CAA GTT CGG TAT CGG ATC GG-3') for the *virB4d* gene; *yhd00040*-RE1SP00079f (5'-GAT CGA TCT GGC ACG TAT GT-3') and *yhd00040*-RE1SP00079r (5'-TAG ACT CTC GGC AAG ACG TT-3') for the *yhd00040* gene; *flgKch*-RECH00680f (5'-CAC GCA GAG CAG TGT CGT AT-3') and *flgKch*-RECH00680r (5'-GGC GAC GTT TCA TAG TCG TT-3') for the *flgKch* gene; *ypf00267*-REPF00546f (5'-TCT ACT GGC TCA CGA ACA CG-3') and *ypf00267*-REPF00546r (5'-TTG TCG AGC CTG TTG AAG TG-3') for the *ypf00267* gene; *yha00045*-RHEA00136F (5'-TCG AGC GGT GGC AAT CTG GC-3') and *yha00045*-RHEA00136R (5'-ACA GCC GGT TGT CGC CGT TG-3') for the *yha00045* gene; and *hisC*-RE1SP0000233f (5'-CGA TGG CGA GAC AGC TAA AT-3') and *hisC*-RE1SP0000233r (5'-ATC ATC GCA ACG CTA TCT CC-3') for the *hisC* gene. Each reaction mixture contained 12.5 µl SYBR green PCR master mixture (Applied Biosystems), 3.5 µl H₂O, forward and reverse primers in 5 µl, and the template in 4 µl. PCRs were performed with the ABI Prism 7700 sequence detection system (Applied Biosystems) using the following program: 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. The dissociation protocol was 95°C for 15 s and 60°C for 20 s, followed by ramp from 60°C to 95°C for 20 min. The transcript of the histidinol phosphate aminotransferase protein (*hisCd*) was used as an internal (unregulated) reference for relative quantification. This gene was selected as a reference because its expression is constitutive under all of the conditions tested (free living and symbiosis) (data not shown) and, additionally, it is in the pSym replicon, which contains practically all members of the NifA regulon. We also used *rpoA* as a reference and obtained similar results. The results of real-time RT-PCR

were analyzed using the $\Delta\Delta C_T$ method (52), and the data are expressed below as relative levels of expression. All reactions were done in triplicate.

RESULTS AND DISCUSSION

***R. etli* whole-genome array.** The *R. etli* CFN42 oligonucleotide (70-mer) microarray was designed on the basis of the genome sequence and annotation published by Gonzalez et al. (24). This array comprises a collection of 6,051 probes, 6,034 of which are from the annotated predicted coding sequences (CDSs). A total of 5,776 (95.72%) of the CDSs were represented by a unique probe set and satisfied all of the selection criteria mentioned in Materials and Methods (Table 1). Only 258 (4.28%) oligonucleotides did not satisfy one or more of these criteria (Table 1). A total of 148 genes were not represented by exclusive probe sets, mostly because their sequences were too similar to other sequences. Similar oligonucleotide set results were obtained for the *S. meliloti* oligonucleotide set array (Array Ready Oligo Set version 1.0; Qiagen, Germany), for which 95.3% of the oligonucleotides met all of the criteria and 4.7% of the oligonucleotides did not meet the criteria (Qiagen). The reliable microarray described here is a useful tool for functional genomics studies of one of the most agriculturally important bacteria.

Transcriptome analysis of the NifA regulon in symbiosis. To dissect the NifA-RpoN regulon under symbiotic conditions, we first analyzed *rpoN* gene expression in both strains during symbiosis by using real-time RT-PCR. Previously, two differentially regulated *rpoN* genes were identified in *R. etli*. The first copy of the *rpoN* gene, *rpoN1*, is required for housekeeping functions and allows cells to grow on several nitrogen and carbon sources under free-living conditions. The second copy (*rpoN2*) is essential for symbiotic nitrogen fixation (15). Using real-time RT-PCR, we analyzed *rpoN2* expression in both strains (the wild-type and *nifA* mutant strains) during symbiosis (Fig. 1a) and observed that in *R. etli* CFN42 *rpoN2* expression depends on the NifA protein because it did not occur in the *R. etli nifA* mutant under free-living and symbiotic conditions (Fig. 1a and 1b). Similar data were reported previously by Michiels et al. (40) and Dombrecht et al. (15) for *R. etli* CNPAF512, and, in agreement with these reports, we obtained for *R. etli* wild-type strain CFN42 a messenger PCR product that included the *prxS* and *rpoN2* genes (Fig. 1c).

To examine the NifA-RpoN regulon under symbiotic conditions, the transcriptomes of bacteroids isolated from 11-dpi nodules of *P. vulgaris* infected with either the wild type or the *nifA* mutant strain were compared. We selected this time point for our study because (i) Puppo and colleagues previously found that determinate nodules like bean nodules are initiated by meristem cells in the outer cortex, but cell division stops 10 days after infection, when nitrogen fixation begins (47), (ii) Valderrama et al. (55) found that the nitrogenase genes in *R. etli* are induced only 10 days after inoculation of the bacterium onto bean roots, and (iii) there are not drastic changes in the physiology of the *nifA* mutant at 11 dpi like those observed in Fix⁻ nodules obtained at later times, for which early senescence and in most cases cell death have been documented (9, 22). We demonstrated that there was nodule integrity and bacteroid viability by using electron microscopy and confocal microscopy with fluorescent dyes (propidium iodide) in nod-

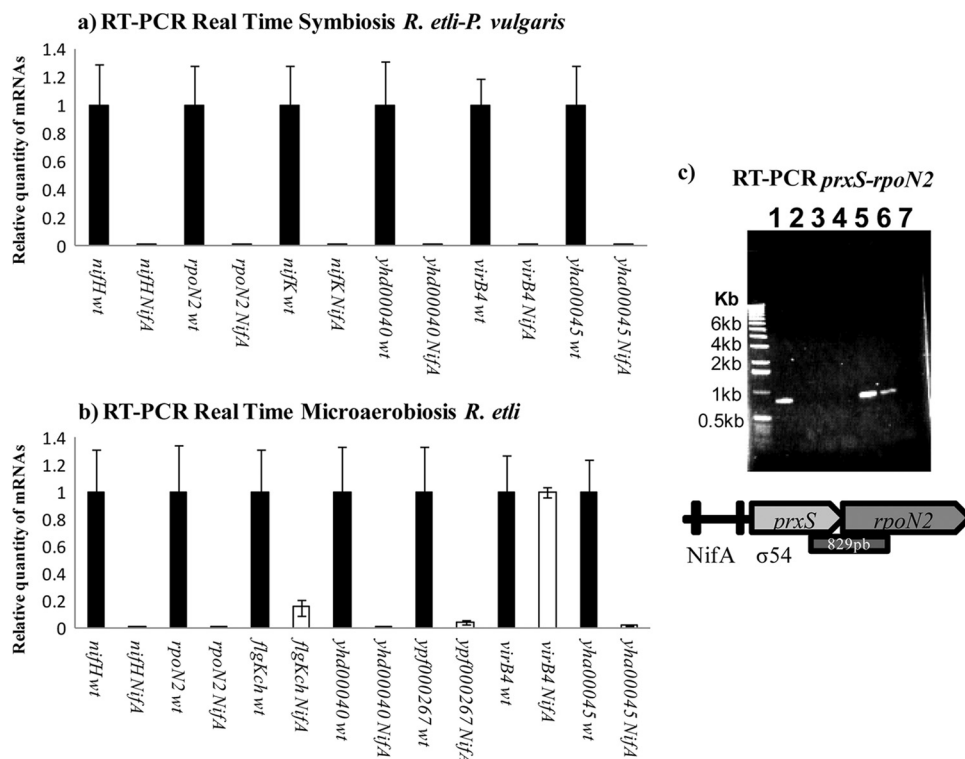


FIG. 1. (a and b) Relative levels of transcripts in *R. etli* wild-type and *nifA* mutant strains in symbiosis (a) and under microaerobic culture conditions (b). The *hisCd* transcript was used as an internal (unregulated) reference for relative quantification. (c) Analysis of the *prxS-rpoN2* operon by RT-PCR. Lane 1, amplification of 829-bp product from bacteroids of the wild-type strain; lane 2, bacteroids from the *nifA* strain; lanes 3 and 4, preparation without retrotranscriptase enzyme; lane 5, total DNA; lanes 6 and 7, wild-type and *nifA* strains, respectively, under microaerobic conditions.

ules inoculated with the *nifA* mutant strain at 11 dpi (data not shown).

Total RNA was extracted from bacteroids, and cDNA was synthesized, differentially labeled with the fluorophore Cy3 (RNA from bacteroids of the *R. etli* wild-type strain) or Cy5 (RNA from bacteroids of the *R. etli* *nifA* mutant strain), and then hybridized to a *Rhizobium etli*_CFN42_6051_v1.0 DNA microarray, as described in Materials and Methods. Three biological replicates with one dye swap were performed. We found that 75 genes showed reduced expression in the *nifA* mutant in symbiosis if we established a z score \log_2 ratio cutoff of ≥ 2.0 using the geneArise software (see Table S1 in the supplemental material). When the physical organization of the induced genes was analyzed in detail, we found that some of these genes are members of operons (*iscN-nifSUW*, *fixABCX*, and *cpxP1P2P3-yhd00049-cpxP4* operons) and some are monocistronic units (*yha00045*, *ypd00055*, and *yhd00073*).

To support the conclusion that NifA is involved in transcriptional regulation of the potential target genes, we measured the differential gene expression by using real-time RT-PCR assays. As expected, the mRNA levels of practically all of the genes tested (*nifH*, *nifK*, *virB4d*, and *yha00045*) were clearly lower in the *nifA* mutant strain than in the wild-type strain (Fig. 1a). On the other hand, only three genes were upregulated in the *nifA* mutant compared to the wild-type strain in symbiosis (see Table S2 in the supplemental material).

In *R. etli*, as in other rhizobia, *nifHDK* and other genes

involved in nitrogen fixation, like *fixABCX*, are controlled by the oxygen-responsive NifA activator protein (22, 55). Our transcriptome analysis results for symbiosis showed that all of these genes were downregulated in bacteroids of the *nifA* mutant strain. However, the most significant finding was the suggestion that 36 genes not reported previously in other bacterial NifA regulons are controlled by NifA (Table 2).

Comparison of microaerobically grown wild-type and *nifA* cells. To obtain a more complete picture of the genes whose expression is regulated by NifA-RpoN, we carried out a new set of microarray experiments in which we compared the global transcriptional profiles of the wild-type strain and the *NifA* mutant strain grown under free-living microaerobic conditions as described in Materials and Methods. By using a z score ratio cutoff of ≥ 2 for genes whose expression changed significantly in each of three microarray replicates, we identified 38 genes whose levels of transcript expression were lower in the *nifA* mutant than in the wild type (see Table S3 in the supplemental material). We also validated some of the results for genes such as *flgKch*, *ypf00267*, *yhd00040*, and *yha00045* using real-time RT-PCR; in contrast, *rpoN2* and *nifH* were found to be differentially expressed only when RT-PCR was used (Fig. 1a and 1b).

When both approaches (microarrays and real-time RT-PCR) were used, the *flgKch* and *ypf00267* genes were found to be downregulated in the *nifA* mutant strain in microaerobiosis but not in symbiosis. It has been suggested that *flgKch* is part of

TABLE 2. Stringent and extended sets of genes in the *R. etli* NifA-RpoN regulon^a

Designation	Gene	Function	z score (symbiosis) ^b	z score (free living) ^b	σ^{54c}	NifA ^c	Probable operon
Stringent set							
RHE_CH00680	<i>flgKch^e</i>	Flagellar hook-associated protein		-2.704	DM	DM	<i>flgKch</i>
RHE_CH00962	<i>thrB^e</i>	Homoserine kinase		-2.598	DM	DM	<i>thrBa</i>
RHE_CH02240	<i>yhch00663^e</i>	Hypothetical conserved protein		-3.88	DM	DM	<i>yhch00663</i>
RHE_CH02986	<i>ypch01047</i>	Probable sugar ABC transporter, ATP-binding protein	-2.001		DM	DM	<i>ypch01047</i>
RHE_PA00136 ^d	<i>yha00045^f</i>	Hypothetical conserved protein	-2.625	-4.831	DA	DM	<i>yha00045</i>
RHE_PD00076 ^d	<i>meIA^f</i>	Tyrosinase, monophenol monooxygenase	-2.107	-2.214	DA	DA	<i>meIA-yhd00038</i>
RHE_PD00077	<i>yhd00038</i>	Hypothetical protein	-2.464				<i>meIA-yhd00038</i>
RHE_PD00078	<i>yhd00039^f</i>	Hypothetical conserved protein	-2.947	-2.876	DA	DM	<i>yhd00039</i>
RHE_PD00079	<i>yhd00040^e</i>	Hypothetical conserved protein		-3.042	DA	DM	<i>yhd00040</i>
RHE_PD00105	<i>ypd00023</i>	Putative two-component sensor protein	-2.832		DM	DM	<i>ypd00023</i>
RHE_PD00146	<i>virB4d</i>	Transport secretion system IV, VirB4 protein	-2.303		DM	DA	<i>virB4d</i>
RHE_PD00186	<i>yhd00073^f</i>	Hypothetical conserved protein	-3.353	-4.547	DA	DA	<i>yhd00073</i>
RHE_PD00197	<i>nifX</i>	Iron-molybdenum cofactor processing protein	-2.946				<i>nifHDK3E2NX</i>
RHE_PD00198	<i>nifN</i>	Nitrogenase reductase iron-molybdenum cofactor synthesis protein	-3.159				<i>nifHDK3E2NX</i>
RHE_PD00199	<i>nifE2</i>	Nitrogenase reductase iron-molybdenum cofactor synthesis protein	-3.304				<i>nifHDK3E2NX</i>
RHE_PD00200 ^d	<i>nifK2</i>	Nitrogenase molybdenum-iron protein beta chain	-3.208				<i>nifHDK3E2NX</i>
RHE_PD00201 ^d	<i>nifD3</i>	Nitrogenase reductase component II, iron-molybdenum alpha subunit protein	-3.994				<i>nifHDK3E2NX</i>
RHE_PD00202 ^d	<i>nifH3</i>	Nitrogenase reductase component II, iron protein	-4.421		DA	DA	<i>nifHDK3E2NX</i>
RHE_PD00214	<i>ypd00045^f</i>	Putative ferredoxin protein	-4.099	-3.368			<i>cpxA-ypd00045</i>
RHE_PD00215	<i>cpxA^f</i>	Cytochrome P450 monooxygenase	-2.646	-4.04	DA	DA	<i>cpxA-ypd00045</i>
RHE_PD00217 ^d	<i>prxS</i>	Putative hydroperoxide reductase	-2.067		DA	DA	<i>prxS-rpoN2</i>
RHE_PD00218 ^d	<i>rpoNd</i>	RNA polymerase sigma factor (sigma-54)	-4.32				<i>prxS-rpoN2</i>
RHE_PD00220	<i>iscN^f</i>	Fe-S cofactor nitrogenase synthesis protein	-3.146	-3.759	DA	DA	<i>iscN-nifUSW</i>
RHE_PD00221	<i>nifU^d</i>	Pyridoxal cysteine desulfurase	-3.572	-3.01			<i>iscN-nifUSW</i>
RHE_PD00222	<i>nifSd</i>	Pyridoxal-phosphate-dependent aminotransferase nitrogen fixation protein	-3.569				<i>iscN-nifUSW</i>
RHE_PD00223	<i>nifW</i>	Nitrogenase stabilizing/protective protein	-3.836				<i>iscN-nifUSW</i>
RHE_PD00224	<i>fixA^f</i>	Electron transfer flavoprotein beta subunit	-3.533	-4.948	DA	DA	<i>fixABCX</i>
RHE_PD00225	<i>fixB^f</i>	Electron transfer flavoprotein alpha subunit	-3.557	-4.419			<i>fixBCX</i>
RHE_PD00226	<i>fixC^f</i>	Electron transfer flavoprotein-quinone oxidoreductase	-3.018	-5.223			<i>fixABCX</i>
RHE_PD00227	<i>fixX^f</i>	Ferredoxin-like protein	-3.848	-4.834			<i>fixABCX</i>
RHE_PD00229	<i>nifB^f</i>	Fe-Mo cofactor biosynthesis protein	-3.465	-2.775	DA	DA	<i>nifB-fdxN-nifZ</i>
RHE_PD00230	<i>fdxN</i>	Ferredoxin protein	-3.379				<i>nifB-fdxN-nifZ</i>
RHE_PD00231	<i>nifZ</i>	Probable chaperone-like protein	-2.768				<i>nifB-fdxN-nifZ</i>
RHE_PD00232	<i>yhd00084</i>	Hypothetical conserved protein	-3.557		DM	DM	<i>yhd00084</i>
RHE_PD00244	<i>ctrE^f</i>	Polyprenyl synthetase	-3.093	-3.305	DA	DM	<i>ctrE</i>
RHE_PD00245	<i>ypd00047</i>	Putative oxidoreductase	-2.726				<i>yhd00088-yhd00089-ypd00047</i>
RHE_PD00246	<i>yhd00088</i>	Hypothetical conserved protein	-2.73				<i>yhd00088-yhd00089-ypd00047</i>
RHE_PD00247	<i>yhd00089</i>	Hypothetical conserved protein	-2.745		DM	DA	<i>yhd00088-yhd00089-ypd00047</i>
RHE_PD00249	<i>cpxP4</i>	Cytochrome P450 protein	-2.498				<i>cpxP1P2P3-yhd00049P4</i>
RHE_PD00250	<i>ypd00049</i>	Probable oxidoreductase	-2.93				<i>cpxP1P2P3-yhd00049P4</i>
RHE_PD00251	<i>cpxP3</i>	Cytochrome P450 monooxygenase	-3.228				<i>cpxP1P2P3-yhd00049P4</i>
RHE_PD00252 ^d	<i>cpxP2</i>	Cytochrome p450 monooxygenase	-3.165				<i>cpxP1P2P3-yhd00049P4</i>
RHE_PD00253	<i>cpxP1</i>	Cytochrome P450 truncated protein	-3.424		DA	DA	<i>cpxP1P2P3-yhd00049P4</i>
RHE_PD00260 ^d	<i>nifH2</i>	Nitrogenase reductase component II, iron protein	-2.871		DA	DA	<i>nifHD2-hemNd2-yhd00090</i>
RHE_PD00261	<i>nifD2</i>	Nitrogenase reductase component I truncated protein	-3.124				<i>nifHD2-hemNd2-yhd00090</i>
RHE_PD00262	<i>hemNd2</i>	Coproporphyrinogen III oxidase	-2.866				<i>nifHD2-hemNd2-yhd00090</i>
RHE_PD00263	<i>yhd00090</i>	Hypothetical protein	-3.138				<i>nifHD2-hemNd2-yhd00090</i>
RHE_PD00301	<i>bacS^f</i>	Bacteroid protein S	-2.915	-3.905	DA	DA	<i>bacS</i>
RHE_PD00302	<i>fdxB</i>	Electron transfer ferredoxin B protein	-3.721				<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00303	<i>yhd00102</i>	Hypothetical conserved protein	-2.47				<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00304	<i>yhd00103</i>	Hypothetical conserved protein	-2.293				<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00305	<i>nifE1</i>	Nitrogenase reductase iron-molybdenum cofactor synthesis truncated protein	-2.108				<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00306 ^d	<i>nifK1</i>	Nitrogenase component I, iron-molybdenum beta subunit protein	-2.984				<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00307 ^d	<i>nifD1</i>	Nitrogenase component I, iron-molybdenum protein, alpha subunit protein	-2.6				<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00308 ^d	<i>nifH1</i>	Nitrogenase reductase component II, iron protein	-3.315		DA	DA	<i>nifHDKE1-yhd00103-yhd00102-fdxB</i>
RHE_PD00312 ^d	<i>yhd00105</i>	Hypothetical protein	-2.917		DA	DA	<i>yhd00105</i>
RHE_PD00313	<i>yi04c-I100043</i>	Probable insertion sequence transposase	-2.124		DM	DM	<i>yi04c-I10043-yi04c-I10044</i>
RHE_PD00314	<i>yi04c-I100044</i>	Probable insertion sequence transposase	-2.099				<i>yi04c-I10043-yi04c-I10044</i>
RHE_PD00332	<i>ypd00053^f</i>	Probable secretion protein, HlyD family	-2.041	-2.593	DA	DA	<i>ypd00053-ypd00054</i>

Continued on following page

TABLE 2—Continued

Designation	Gene	Function	z score (symbiosis) ^b	z score (free living) ^b	σ^{54c}	NifA ^c	Probable operon
RHE_PD00333	<i>ypd00054</i>	Probable bacteriocin/lantibiotic exporter ABC transporter, ATP binding protein	-3.117				<i>ypd00053-ypd00054</i>
RHE_PD00335	<i>yhd00111</i>	Hypothetical conserved protein	-3.812		DA	DM	<i>yhd00111</i>
RHE_PD00339	<i>ypd00055^f</i>	Probable water channel protein	-2.792	-4.418	DA	DM	<i>ypd00055</i>
RHE_PD00340	<i>yhd00114^f</i>	Hypothetical conserved protein	-2.43	-3.795	DA	DA	<i>yhd00114</i>
RHE_PD00343	<i>yhd00115^f</i>	Hypothetical protein	-2.512	-3.444			<i>yhd00116-yhd00115</i>
RHE_PD00344	<i>yhd00116</i>	Hypothetical protein	-2.127		DA	DM	<i>yhd00116-yhd00115</i>
RHE_PE00331	<i>ype00167^e</i>	Putative peptide ABC transporter, permease protein		-1.97	DM	DM	<i>ype00167</i>
RHE_PF00546	<i>ypf00267^e</i>	Putative epoxide hydrolase protein		-2.578	DM	DM	<i>ypf00267</i>
Extended set							
RHE_CH02239	<i>yhch00662^e</i>	Hypothetical conserved protein		-3.685	ND	DM	<i>yhch00662</i>
RHE_PD00106	<i>ypd00024</i>	Putative metallo-hydrolase	-2.915		ND	DM	<i>ypd00024</i>
RHE_PD00110	<i>ypd00026</i>	Putative hydrolase	-2.42		ND	DM	<i>ypd00026</i>
RHE_PD00123	<i>yi14a-II00019</i>	Probable insertion sequence transposase	-2.972		DA	ND	<i>yi14a-II00019</i>
RHE_PD00139	<i>ardCd^d</i>	Conjugal transfer antirestriction protein	-2.672	-3.02	DM	ND	<i>ardCd</i>
RHE_PD00205	<i>yhd00078</i>	Hypothetical conserved protein	-3.077		ND	DM	<i>yhd00078</i>
RHE_PD00213	<i>yhd00082</i>	Hypothetical conserved protein	-2.231		ND	DA	<i>yhd00082</i>
RHE_PD00315	<i>ypd00051</i>	Probable methionyl-tRNA synthetase	-2.272		ND	DM	<i>ypd00051</i>
RHE_PD00341	<i>aptD</i>	Adenine phosphoribosyltransferase	-2.478		ND	DM	<i>aptD-ypd00056</i>
RHE_PD00342	<i>ypd00056^f</i>	Putative biosynthetic heme protein	-2.611	-3.302			<i>aptD-ypd00056</i>
RHE_PF00533	<i>ypf00260^e</i>	Probable two-component response regulator protein		-2.023	ND	DM	<i>ypf00260</i>

^a All genes were differentially expressed as determined by a z score of ≤ 2 for the *nifA* mutant compared with the wild-type strain. The genes in the stringent set contained both NifA and RpoN binding sites. The genes in the extended set contained either NifA or RpoN motifs, and expression was corroborated by proteomics where indicated.

^b Detected by *in silico* analysis (MEME-MAST program).

^c DM, detected manually; DA, detected by *in silico* analysis (MEME-MAST program); ND, not detected.

^d Corroborated by proteomics.

^e Gene expressed under microaerobic conditions.

^f Gene expressed in symbiosis and under microaerobic conditions.

the *flgEKL-flaF-flbT-flgDch* operon (24), but using microarrays we did not detect expression of the other members of this operon. Regarding incomplete operon expression in microaerobiosis, Bobik et al. (6) reported that in some cases it is not possible to detect expression of all members of an operon using microarrays due to low levels of expression. However, a Northern blot analysis showed that there was only one band corresponding to the size of *flgKch* mRNA, 1,487 bp (data not show). Similar results were obtained for the *cpxA* gene, which is a member of the probable *cpxA-ypd0045* operon, and for *iscN* of the *iscN-nifUSW* operon. Similar to the results of the experiments under symbiotic conditions, in which we detected 3 genes that were induced in the *nifA* mutant strain, under microaerobic conditions we detected induced expression of 37 genes in the mutant, which were distributed throughout the partitioned genome of *R. etli* CFN42 (see Table S4 in the supplemental material); this might have been a result of indirect regulatory effects originating from the mutant genotype.

As mentioned above, in our microarray analysis we considered only genes whose expression in three replicates was constant. However, when the data were analyzed separately, we observed that almost all of the genes detected as part of the NifA regulon in symbiosis were induced differentially under microaerobic conditions. These results and those obtained for *B. japonicum* (25) and *S. meliloti* (5) indicate that microaerobic conditions for free-living growth are probably not compatible with maximal expression of rhizobial NifA proteins. As we mentioned above and in agreement with previous reports (40), using RT-PCR we observed that in *R. etli* *rpoN2* expression in symbiosis and under microaerobic conditions depends on the

NifA protein (Fig. 1a and 1b). The level of amino acid identity of RpoN1 and RpoN2 is 56%, and this could suggest that RpoN1 has less capacity for binding to RpoN2 promoters and *vice versa*. Moreover, in *Rhodobacter capsulatus* (46) RpoN1 and RpoN2 have different targets, determined in part by the capacity of each protein to bind to its corresponding site. Specificity is determined mainly by region 1, and in *R. etli* there is a major difference between RpoN1 and RpoN2 in this region. On the other hand, it has been reported that *rpoN1* expression is reduced during microaerobiosis and symbiosis, suggesting that in addition to RpoN1, RpoN2 negatively regulates *rpoN1* expression (40). Another feature that affects gene expression compared with expression during microaerobiosis is the very low oxygen concentration in the nodule (20, 34), which has been proposed to produce differential regulation in the FixLJ-FixK2-FixK1 cascade of *B. japonicum* (39), of which NifA is a part.

We propose that NifA has greater stability and activity in symbiosis, which leads to increased expression of RpoN2, and as a result of the targets that were identified in each experimental replicate, we suggest that the putative targets of RpoN1 are not expressed in the absence of the *rpoN1* product. We assumed that we did not use the best conditions for NifA-RpoN2 function in microaerobiosis, and while we detected some targets under our conditions, the expression of NifA-RpoN2 was not as stable as it was in symbiosis, which caused the replicates to vary substantially.

***In silico* search for RpoN and NifA binding sites.** The array results reported above suggest that the effect of NifA on gene expression could be due to direct or indirect regulation of the

TABLE 3. Regulatory RpoN-NifA binding sites detected in the genome of *R. etli* CFN42 using the MEME-MAST program and manually

Gene	RpoN binding site	NifA binding site
<i>cpxP1</i>	CGGCACGCGATTTGCT	TGTCGGCATTGAGACT
<i>yhd00084</i>	CGGCGCCACGGGTGCT ^a	TGTCGGCTTACGTACC ^a
<i>thrB</i>	CGGGCAGAGAGATGCT ^a	TGTTATGCCACGACCA ^a
<i>nodA</i>	TGGCAGGCCGATTGAT	TGTCGACGCCTGGACA
<i>ctrE</i>	TGGCACACGGTTTCAT	TGTCGCCATTGAAACA ^a
<i>ypd00053</i>	TGGCACGAATCTTGAA	TGTCGTGCTTGGCACA
<i>fixA</i>	TGGCACGAATGATGCA	TGTCGGGAAGCTTACA
<i>iscN</i>	TGGCAGCGGTTTTGAA	TGTCGCGCAAGCGACA
<i>yha00045</i>	TGGCACGCTGATATCC	TGTCGCCATCAAATCA ^a
<i>nifH1</i>	TGGCACGGTTTTGAA	TGTCTGATCGGCGACA
<i>nifH2</i>	TGGCAGCGGTTTTGAA	TGGCGGTTTCGAGACA
<i>nifH3</i>	TGGCACGGTTTTGAA	TGTCTGATCGGCGACA
<i>yhd00105</i>	TGGCACGGTTGTTGCG	TGTTCCGGTTTGGCACA
<i>melA</i>	TGGCAGCTCGCTTGCT	TGTCGAGATCAGACA
<i>yhd00039</i>	TGGCACTGGATTTGCA	TGGTCAACAGTTTACA ^a
<i>yhd00040</i>	TGGCATAGACATTGCA	TGTA AACCTGACCAACA ^a
<i>cpxA</i>	TGGCATGAACTTTGAA	TGTTAGATACGCGACA
<i>prxS</i>	TGGCATGACGCTTGCT	TGTGCAGTTTCAAACA
<i>yhd00111</i>	TGGCATGAGTCTTGAA	TGTCGGTTTACGACA ^a
<i>nifB</i>	TGGCATGGGTTTTGCG	TGTCGGAAGCTTGACA
<i>bacS</i>	TGGCATGGTGTGTTGCT	TGTCGGGCATCCGCCA
<i>yhd00073</i>	TGGCATGGTGTGTTGCT	TGTCGGGCATCCGCCA
<i>yhd00114</i>	TGGCATGGTGTGTTGCT	TGTCGGGCATCCGCCA
<i>ypd00023</i>	TGGCCCTCGTAAAGGA ^a	TCTCGCCAAACGGGACA ^a
<i>yi04c-100043</i>	TGGCCGAACAGATGCT ^a	TCTCACCATTGCCACA ^a
<i>ypch01047</i>	TGGCCGGCCTCGAGCT	TGCCGTTCTCTCACA
<i>ypf00267</i>	TGGCGAAATACTGGG ^a	TGGTATCGGGAATACA ^a
<i>yhd00089</i>	TGGCGCAGGCCGTGCA ^a	TGTGTTCCGATTCCACG
<i>flgKch</i>	TGGCGCCGCTTTGCA ^a	TGTCGATCCGGTGACG ^a
<i>ype00167</i>	TGGCGGGCGCCAGCA ^a	TGGTACCAGTACAACA ^a
<i>ypd00010</i>	TGGGAGGAACTATGAG	TGTCGGGCCTACTACA
<i>virB4d</i>	TGGGGCGGCGCCACGT ^a	TGGATGTTTTGGCTACA
<i>yhch00663</i>	TGGCCCTGCTTTTGCT ^a	TGTCATAACCGGCAAAA ^a
<i>ypd00055</i>	TGGTACGACGTTTGCT	TGGCGGATGCCCGACA ^a
<i>yhd00116</i>	TTGCTGCAGATACGCA	TGTCACGCAGATCGCA ^a

^a Detected manually.

corresponding promoters by NifA. In order to obtain insight into which promoters are regulated directly and which promoters are regulated indirectly, a genome-wide DNA motif search was performed.

The bacterial alternative sigma factor RpoN recognizes and binds to a $-24/-12$ -type promoter with the following consensus sequence: 5'-TGGCACG-N₄-TTGCW-3', where the -24 GG and the -12 GC dinucleotides are indicated by bold type (4). RpoN binds to these promoter regions and interacts with the NifA protein that recognizes and binds to a specific upstream sequence, TGT-N₁₀-ACA; all of the sequences used to locate NifA and RpoN sites and sequence logos (11) created with these sequences are shown in Fig. S1 in the supplemental material and in Table 3.

Using the MEME and MAST programs, the search was restricted to the regions from bp -500 to bp 200 from the start codons of annotated genes. A total of 154 RpoN binding sites and 97 NifA binding sites were predicted (see Table S5 in the supplemental material); 17 of these sites had both motifs, while 18 additional binding sites were detected manually (7 sites with NifA binding motifs, 2 sites with RpoN binding motifs, and 9 sites with both binding motifs), resulting in a total of 35 targets (Table 3). The genome locations of the RpoN- and NifA-dependent sequence elements identified were cross-examined using the genes identified by the transcriptome analysis that had a z score cutoff of 2.0. As shown in Table 3, we concluded

that of the 35 NifA-RpoN sites identified, 33 (94.2%) are apparently functional.

In symbiosis a total of 61 genes associated with both motifs showed significant changes in microarray experiments, and these genes are members of 14 operons and 12 monocistronic units (Table 2). In contrast, there were 5 genes that were differentially expressed in the NifA mutant strain but for which no NifA- or RpoN-dependent regulatory elements could be found (see Table S6 in the supplemental material). In microaerobiosis we identified 38 genes whose levels of transcript expression were lower in the *nifA* mutant than in the wild-type strain (see Table S3 in the supplemental material). However, for 9 genes neither a NifA binding site nor an RpoN binding site was detected (see Table S6 in the supplemental material), and just 19 targets (25 genes) were identified as targets that had both motifs, which corresponds to 65.7% of the total number of genes whose expression was reduced in the *nifA* mutant in microaerobiosis; additionally, in 4 more cases only one of the motifs was present. It is relevant that the expression of 19 of the 25 genes with both putative transcription factor binding sites was reduced in the *nifA* mutant in symbiosis. On the other hand, the expression of *virB4d*, which also possessed putative NifA and RpoN motifs, was reduced in the *nifA* mutant background in symbiosis as determined by microarrays and RT-PCR assays, in contrast to the findings for microaerobiosis (Fig. 1a and 1b). Also, NifA and RpoN regulatory binding sites were predicted to be in the promoter region of *yhd00040*, but by using microarray analysis a low level of expression in the *nifA* mutant strain in microaerobiosis but not in symbiosis was found, which is not consistent with the results of a real-time RT-PCR analysis, in which we detected differential expression under both conditions (Fig. 1a and 1b). Since real-time RT-PCR is a more sensitive technique than microarray assays, it is possible that we detected expression of this gene only by the former method.

A number of NifA-dependent genes were detected only under microaerobic conditions (*yhd00040*, *flgKch*, *thrB*, *ype00167*, *yhch00663*, and *ypf00267*), suggesting that some NifA regulation could be specific for the free-living state of the bacterium. Nevertheless, some of the genes whose expression was not detected in symbiosis were expressed at low levels under microaerobic conditions. Consequently, we cannot exclude the possibility that the lack of detection of these genes under symbiotic conditions might actually have been due to a technical limitation, given that the overall sensitivity of the arrays is better with free-living RNAs than with symbiotic RNAs. For this reason, whether NifA contributes to biochemical functions specific for free-living *R. etli* remains to be determined.

In our opinion, these data define the NifA regulon in *R. etli* more reliably than other NifA regulon studies have, because our analysis included results obtained both under free-living (microaerobic) conditions and in symbiosis. On the other hand, we found that the *nodA* and *ypd0010* genes, containing both NifA and RpoN binding sites, were not differentially expressed in symbiosis or microaerobiosis (data not shown), which could suggest that there is an alternative NifA-dependent regulation process. *nodA* is also negatively regulated by NoliR (10, 33), and a search for NoliR consensus sites in the *R. etli* genome sequence using MEME and MAST identified three motifs (data not shown), one of which is located upstream of *nodA*. In

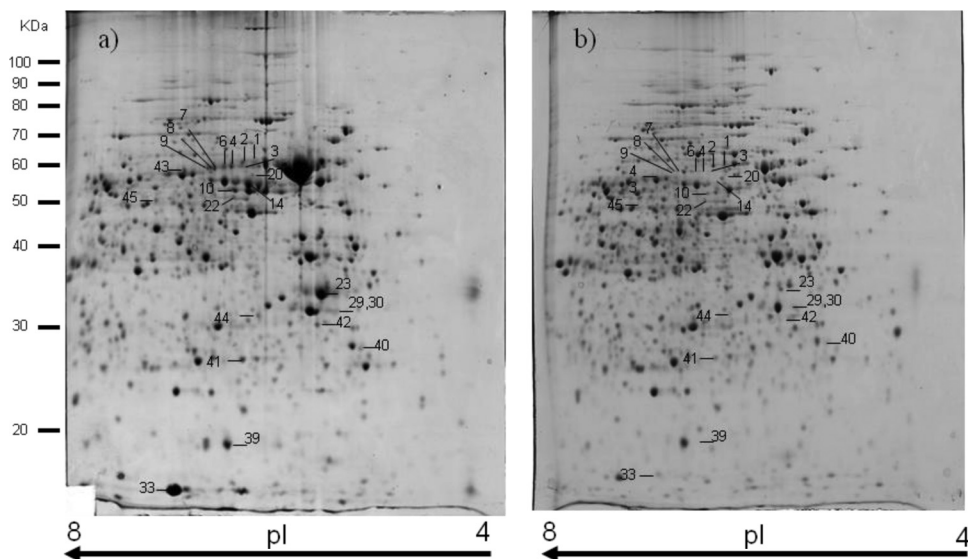


FIG. 2. Proteome profiles of 11-dpi bacteroids from the *R. etli* wild-type (a) and *nifA* mutant (b) strains. The numbered protein spots were excised from the gel of the wild-type strain and were identified using MALDI-TOF mass spectrometry. Three independent experiments were performed, and representative 2D gel electrophoresis gels are shown.

addition, *nolR* is transcribed significantly in symbiosis and microaerobiosis (data not shown). Also, the findings for the *ypd00010-yhd00018* operon, which contained both promoter sites but did not exhibit differential expression either in symbiosis or under free-living conditions, suggest that there is an alternative NifA-dependent regulation process.

Identification of NifA-regulated proteins in symbiosis by proteome analysis. Bacteroid extracts of wild-type strain CFN42 and the *nifA* mutant strain were obtained and analyzed by 2D PAGE. Three independent biological samples from each strain were used for protein profiling. For each sample, 2D gel electrophoresis was carried out using a pH range of 4 to 8, and the Coomassie blue-stained gels were analyzed using PD-Quest software. On average, more than 800 protein spots were observed in each 2D electrophoresis gel prepared for both strains. In this study we selected only spots that showed at least a 2-fold change and met the conditions of a statistical Student test (significance level, 95%). Seventy-eight polypeptide spots that were present in the wild-type strain gel and not in the *nifA* mutant gel were identified as probable NifA-regulated protein spots (Fig. 2a and 2b). In contrast, 18 polypeptide spots were found only in the *nifA* mutant strain gel (data not shown). A total of 41 proteins that were differentially expressed in the wild-type strain and the *nifA* mutant were identified on the basis of their tryptic peptide masses using matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) analysis (Table 4). The genes encoding 24 (58%) of these proteins were associated with NifA and RpoN motifs, and there were significant severalfold changes in the expression of these proteins when transcriptomics experiments were performed. These proteins included 7 different electrophoretic variants of NifD, 4 electrophoretic variants of MelA, 4 electrophoretic variants of Yha00045, 3 electrophoretic variants of NifH, 2 electrophoretic variants of NifK, and one electrophoretic variant each of PrxS, RpoN2, CpxP2, and Yhd00105 (Fig. 2 and Table 4). However, using 2D PAGE it

was not possible to identify proteins like FdxN, IscN, NifW, and FdxB, which were members of a group of 22 proteins identified with microarrays as members of the NifA-RpoN regulon, because they had isoelectric points or molecular weights outside the electrophoretic resolution range utilized in this work (see Table S7 in the supplemental material). Our proteomics experiments confirmed that elements such as CpxP2, Yha00045, and Yhd00105, which are proposed for the first time in this work to be regulated by NifA-RpoN, are members of the NifA-RpoN regulon (which was suggested by transcriptomics findings and the presence of RpoN and NifA binding motifs). Also, in several cases we identified isoforms of the same protein, suggesting that there are regulatory post-transductional modifications. Detection of these modifications is an additional advantage of the proteomics approach used in this work, and for some of these proteins posttransductional modifications have not been reported previously. Further analysis is required to identify the types of modifications in these proteins. Conversely, the results for 17 proteins found only in the wild-type strain (see Table S8 in the supplemental material) did not correlate with either the predicted motif results or transcriptomics analysis results. We know that NifA-dependent σ^{54} motifs can differ from consensus sites, as recently shown using gel retardation analysis and detailed mutagenesis of the functional NifA promoter of the *hupSL* genes of *R. leguminosarum* bv. *viciae* (36); thus, experiments to confirm that these elements are members of NifA-RpoN regulon are necessary.

Global characterization of the *R. etli* NifA-RpoN regulon. Previous studies documented that there is an imperfect correlation between the motif scores and fold changes, so promoter sequence data alone are not sufficient to predict transcription efficiency (25). Because of this, to delineate the *R. etli* NifA-RpoN regulon (stringent set), we selected as members of this regulon (stringent set) only genes that had significant fold changes in transcriptomics experiments (symbiosis and/or free

TABLE 4. NifA-RpoN regulon proteins identified using MALDI-TOF mass spectrometry^a

Spot	Designation	Gene	Mascot score	% Coverage	Peptide no. ^b	<i>M_r</i> (calculated)	pI (calculated)
33	RHE_PA00136	<i>yha00045</i>	56	24	4/11	17.04	6.49
34	RHE_PA00136	<i>yha00045</i>	71	28	4/12	17.03	6.32
35	RHE_PA00136	<i>yha00045</i>	64	36	6/20	17.15	5.93
36	RHE_PA00136	<i>yha00045</i>	63	28	4/18	17.18	5.73
10	RHE_PD00076	<i>melA</i>	79	10	5/9	50.80	5.79
12	RHE_PD00076	<i>melA</i>	58	9	4/10	50.69	5.69
16	RHE_PD00076	<i>melA</i>	63	9	4/8	51.23	5.64
17	RHE_PD00076	<i>melA</i>	69	10	5/13	51.08	5.49
13	RHE_PD00200	<i>nifK2</i>	54	9	4/13	51.98	5.69
39	RHE_PD00217	<i>prxS</i>	143	29	9/17	20.10	5.85
43	RHE_PD00218	<i>rpoN2</i>	91	17	5/10	55.50	6.30
45	RHE_PD00252	<i>cpxP2</i>	80	17	8/20	48.08	6.80
14	RHE_PD00306	<i>nifK1</i>	119	24	10/38	53.06	5.68
4	RHE_PD00307	<i>nifD1</i>	195	30	11/16	57.23	5.81
5	RHE_PD00307	<i>nifD1</i>	89	23	8/45	55.55	5.83
6	RHE_PD00307	<i>nifD1</i>	165	24	9/12	56.94	5.88
7	RHE_PD00307	<i>nifD1</i>	153	30	11/30	56.49	5.99
8	RHE_PD00307	<i>nifD1</i>	138	28	10/31	56.45	6.04
9	RHE_PD00307	<i>nifD1</i>	151	28	10/24	56.34	6.07
2	RHE_PD00307	<i>nifD1</i>	139	24	9/19	58.90	5.72
25	RHE_PD00308	<i>nifH1</i>	76	22	6/27	33.05	4.77
23	RHE_PD00308	<i>nifH1</i>	82	27	4/8	32.58	4.97
24	RHE_PD00308	<i>nifH1</i>	78	27	4/10	32.85	4.86
44	RHE_PD00312	<i>yhd00105</i>	108	20	8/15	30.10	5.60

^a The spots were excised from gels obtained using wild-type bacteroids of *R. etli*.

^b Number of peptides detected by MALDI-TOF/theoretical peptide number.

life) and/or whose product was detected by proteomics and that contained RpoN- and NifA-dependent binding sites, which resulted in identification of a total of 67 genes; 42 of these genes were symbiosis specific, 19 were found under both symbiosis and microaerobic (free life) conditions, and 6 were microaerobiosis specific (Table 2). The extended set included genes that exhibited differential microarray expression (z score ratio cutoff, ≥ 2) in the *nifA* mutant strain, in contrast to the wild type in symbiosis or microaerobiosis, but contained only NifA or RpoN motifs. In this category there were 11 genes, 2 of which were detected by microarrays specifically in microaerobiosis. Additionally, similar to the results of the proteomics analysis, in which 17 proteins were found only in the wild-type strain, 5 genes were differentially expressed under symbiosis conditions and 9 genes were differentially expressed under microaerobic conditions in this strain; however, no NifA- and/or RpoN-dependent regulatory elements could be found for these genes (see Table S6 in the supplemental material), suggesting that (i) there may be indirect regulation, (ii) NifA and RpoN promoters are too divergent from the promoters used in our analysis, (iii) there were artifacts of the microarray method (transcriptomics), and/or (iv) there was basal induction or no induction of some genes in ineffective nodules (symbiosis). We also did not consider genes in theoretical operons which did not meet the criteria for the ≥ 2.0 -change cutoff in gene expression in the microarray analysis members of this regulon.

In agreement with the findings of Hu et al. (29) and Curatti et al. (12), the NifA-RpoN stringent regulon of *R. etli* described in this work contains all of the necessary elements required to fix nitrogen, including *nifHDK*, *nifB*, *nifE*, *nifN*, *nifX*, *nifW*, *nifU*, and *nifS*. Moreover, we confirmed the presence of the genes previ-

ously reported to be NifA-RpoN dependent in *R. etli*, including *bacS*, *prxS*, *rpoN2*, and *iscN* (16, 31, 43), and the presence of operons predicted *in silico*, including the *iscN-nifUSW* (16), *nifHDK1-yhd00103-yhd00102-fdxB*, *nifHDK3E2NX*, and *nifHD2-hemNd2-yhd00090* operons. For the last three operons, RNA transcripts that were the correct length were detected in Northern blots (data not shown). In summary, the NifA-RpoN regulon (extended set) of *R. etli* was determined to contain 78 genes, including 50 genes in 15 operons and 28 monocistronic genes.

Comparison of NifA-RpoN regulons in nitrogen-fixing bacteria. The NifA-RpoN regulons in different rhizobia have been characterized previously (6, 25, 26). In *S. meliloti*, 19 of the genes regulated by FixJ appeared to be NifA-regulated genes, and analysis of the intergenic sequence upstream of these genes revealed the presence of NifA boxes (TGT-N₁₀-ACA) in five targets (6). In *B. japonicum*, the NifA-RpoN regulon is composed of 65 elements, including 32 regulated genes with significant NifA and RpoN binding sites (stringent set), 2 manually identified RpoN and NifA sites, and 31 regulated genes with significant RpoN binding sites, but it has no NifA binding site (extended set) (26). Remarkably, the total NifA regulon in *R. etli* described in this work consisted of 67 genes in the stringent set and 78 genes in the extended set.

The majority (90%) of the members of the *R. etli* NifA-RpoN regulon (stringent set) occur only on the symbiotic plasmid (60 genes are on pCFN42d, 4 genes are localized on the chromosome, and one gene each is on pCFN42a, pCFN42e, and pCFN42f). *R. etli* has a small symbiotic region compared with other species. In *S. meliloti*, the 19 NifA-dependent genes described previously are on pSymA (6), a replicon five times larger than pCFN42d, and in *B. japonicum* the symbiotic island

contains 34 genes regulated by NifA (stringent set) and is twice as large as pCFN42d. To date, the *R. etli* NifA-RpoN regulon (stringent set) is the largest NifA regulon that has been described, and its members are located in a smaller proportion of the genome. The *yha00045* gene, which contains functional NifA and RpoN binding sites, which was determined by proteomic and microarrays studies to be expressed only in the wild-type strain, and whose expression under symbiosis and microaerobic conditions was validated by RT-PCR, is the only *R. etli* regulon gene detected (stringent set) on pCFN42a. This replicon has been reported to be indispensable for the transfer of the symbiotic plasmid (8, 54), and both plasmids have lower G+C contents (58%) than the rest of the *R. etli* genome (61.5%) (24). These data suggest that both replicons were acquired during the same event, and our data support the hypothesis that there is a coordinated relationship between pCFN42a and pCFN42d.

The *R. etli* NifA-RpoN regulon (stringent set) contains at least one gene in each of 16 functional classes (53); in comparison, the members of the *B. japonicum* regulon belong to only 12 classes (extended set), and the members of the *S. meliloti* regulon (stringent set) belong to only 6 functional classes. Classical genes necessary for nitrogen fixation, like the *nifHDKENX* and *nifB-fdxN-nifZ* operons and *fixA*, are present in all regulons. Additionally, we observed 21 other genes that are present in both the *B. japonicum* and *R. etli* NifA-RpoN regulons; in comparison, there are only 13 genes that are present in both the *R. etli* and *S. meliloti* regulons. Hence, 43 (54.4%), 49 (76.6%), and 6 (31.6%) of the genes were present only in the *R. etli*, *B. japonicum*, and *S. meliloti* regulons, respectively, which indicates that there are important differences and specificities in these regulons.

Interestingly, in the *R. etli* stringent set 47% of the genes were categorized as hypothetical genes, conserved hypothetical genes, or genes with unknown functions (see Fig. S2 in the supplemental material). For the 35 remaining genes, functional categories, such as energy production and conversion (14 elements), were significantly overrepresented. Also present were genes encoding products involved in functions like transport (divided into six clusters of orthologous groups of genes [COGS], as well as inorganic ion transport), metabolism (see Fig. S2 in the supplemental material), and secondary metabolite biosynthesis, transport, and metabolism (see Fig. S2 in the supplemental material), and there were three or four elements in each COG (see Fig. S2 in the supplemental material).

For several coding regions with unknown functions, some of the proteins contained domains that indicated that they are involved in various important physiological functions; these coding regions included *ypd00053* (multidrug resistance efflux pump domain), *ypd00054* (putative bacteriocin/antibiotic exporters), and *ypd00055* (similar to *Mesorhizobium loti* mll6943, which contains a domain conserved in aquaporin proteins, some of which are involved in urea, methylammonium, lactic acid, or ammonia transport) (38). Experimental demonstration of these roles would provide further insights into the NifA-RpoN regulon in *R. etli*.

The abundance of hypothetical genes that are components of the *R. etli* NifA-RpoN regulon is not surprising since the NifA-RpoN regulon of *B. japonicum* is composed of 25 hypothetical genes (26), but only two of these genes (bir1954 and

bir2725) had sequence identity (51% and 57%, respectively) with ORFs in *R. etli*. However, neither of these genes was determined by us to be part of the *R. etli* NifA-RpoN regulon. This result confirmed that the expression of genes for symbiosis is different in *B. japonicum* and *R. etli*.

Conclusion. One of the purposes of this study was to develop and evaluate a relatively low-cost whole-genome DNA microarray system for *R. etli*. Here, we describe the most complete characterization to date of the NifA-RpoN regulon, the first regulon in rhizobia described using transcriptomic profiling, proteome analysis, and bioinformatics. We show that there are 67 members of the stringent set in the *R. etli* NifA-RpoN regulon and 78 members in the extended set. Our studies, done in symbiosis and in free life, revealed 42 new NifA-RpoN-dependent genes. The *R. etli* NifA-RpoN regulon contains components that are also present in other bacterial regulons described previously, although the majority of the components described here are present only in the *R. etli* regulon, which suggests that there are specific requirements for nitrogen fixation in different bacterial models.

In the past, NifA has been characterized largely as a regulator of nitrogen fixation, a finding which was confirmed in the present work. More importantly, however, this study demonstrates that the NifA-RpoN regulon is composed of genes with very diverse functions. Our results reveal that in addition to providing adequate conditions for efficient nitrogen fixation, NifA also plays a fundamental and heretofore less appreciated role in regulation of the normal physiology of the cell.

ACKNOWLEDGMENTS

We thank Michael F. Dunn and David Krogmann for useful scientific comments; Sandra Contreras, Raúl Noguez, and Miguel Elizalde for MALDI-TOF protein identification; María de Lourdes Girard and Nicolas Gómez-Hernández for technical advice concerning microaerobic experiments; and Oliver Castillo, J. L. Zitlalpopoca, and Hadau Sánchez for plant experiments and greenhouse support.

This research was supported by DGAPA-UNAM grant IN222707 and by CONACYT grants 60641 and ING206512. E.S.B. was a recipient of a Ph.D. studentship from the Consejo Nacional de Ciencia y Tecnología (México).

REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Bailey, T. L., and C. Elkan. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**:28–36.
- Bailey, T. L., and M. Gribskov. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**:48–54.
- Barrios, H., B. Valderrama, and E. Morett. 1999. Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res.* **27**:4305–4313.
- Becker, A., H. Berges, E. Krol, C. Bruand, S. Ruberg, D. Capela, E. Lauber, E. Meilhoc, F. Ampe, F. J. de Bruijn, J. Fourment, A. Francez-Charlot, D. Kahn, H. Kuster, C. Liebe, A. Puhler, S. Weidner, and J. Batut. 2004. Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol. Plant-Microbe Interact.* **17**:292–303.
- Bobik, C., E. Meilhoc, and J. Batut. 2006. FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *J. Bacteriol.* **188**:4890–4902.
- Braeken, K., M. Fauvart, M. Vercautse, S. Beullens, I. Lambrichts, and J. Michiels. 2008. Pleiotropic effects of a *rel* mutation on stress survival of *Rhizobium etli* CNPAF512. *BMC Microbiol.* **8**:219.
- Brom, S., A. Garcia-de los Santos, L. Cervantes, R. Palacios, and D. Romero. 2000. In *Rhizobium etli* symbiotic plasmid transfer, nodulation competitiveness and cellular growth require interaction among different replicons. *Plasmid* **44**:34–43.
- Cermola, M., E. Fedorova, R. Tate, A. Riccio, R. Favre, and E. J. Patriarca.

2000. Nodule invasion and symbiosome differentiation during *Rhizobium etli*-*Phaseolus vulgaris* symbiosis. *Mol. Plant-Microbe Interact.* **13**:733–741.
10. Chen, H., K. Gao, E. Kondorosi, A. Kondorosi, and B. G. Rolfe. 2005. Functional genomic analysis of global regulator NolR in *Sinorhizobium meliloti*. *Mol. Plant-Microbe Interact.* **18**:1340–1352.
 11. Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res.* **14**:1188–1190.
 12. Curatti, L., J. A. Hernandez, R. Y. Igarashi, B. Soboh, D. Zhao, and L. M. Rubio. 2007. In vitro synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proc. Natl. Acad. Sci. U. S. A.* **104**:17626–17631.
 13. de Vries, S., H. Hoge, and T. Bisseling. 1993. Isolation of total and polysomal RNA from plant tissues, p. 1–13. *In* S. B. Gelvin, R. A. Schilperoort, and D. P. S. Verma (ed.), *Plant molecular biology manual*. Kluwer Academic Publishers, Dordrecht, Netherlands.
 14. Dixon, R., and D. Kahn. 2004. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2**:621–631.
 15. Dombrecht, B., C. Heusdens, S. Beullens, C. Verreth, E. Mulkers, P. Proost, J. Vanderleyden, and J. Michiels. 2005. Defence of *Rhizobium etli* bacteroids against oxidative stress involves a complexly regulated atypical 2-Cys peroxidoredoxin. *Mol. Microbiol.* **55**:1207–1221.
 16. Dombrecht, B., K. Marchal, J. Vanderleyden, and J. Michiels. 2002. Prediction and overview of the RpoN-regulon in closely related species of the Rhizobiales. *Genome Biol.* **3**:RESEARCH0076.
 17. Dombrecht, B., M. Z. Tesfay, C. Verreth, C. Heusdens, M. C. Napoles, J. Vanderleyden, and J. Michiels. 2002. The *Rhizobium etli* gene *iscN* is highly expressed in bacteroids and required for nitrogen fixation. *Mol. Genet. Genomics* **267**:820–828.
 18. Encarnacion, S., Y. Guzman, M. F. Dunn, M. Hernandez, M. del Carmen Vargas, and J. Mora. 2003. Proteome analysis of aerobic and fermentative metabolism in *Rhizobium etli* CE3. *Proteomics* **3**:1077–1085.
 19. Encarnacion, S., M. Hernandez, G. Martinez-Batallar, S. Contreras, C. Vargas Mdel, and J. Mora. 2005. Comparative proteomics using 2-D gel electrophoresis and mass spectrometry as tools to dissect stimulons and regulons in bacteria with sequenced or partially sequenced genomes. *Biol. Proced. Online* **7**:117–135.
 20. Gabel, C., M. A. Bittinger, and R. J. Maier. 1994. Cytochrome aa3 gene regulation in members of the family Rhizobiaceae: comparison of copper and oxygen effects in *Bradyrhizobium japonicum* and *Rhizobium tropici*. *Appl. Environ. Microbiol.* **60**:141–148.
 21. Girard, L., S. Brom, A. Davalos, O. Lopez, M. Soberon, and D. Romero. 2000. Differential regulation of *fixN*-reiterated genes in *Rhizobium etli* by a novel *fixL*-*fixK* cascade. *Mol. Plant-Microbe Interact.* **13**:1283–1292.
 22. Girard, L., B. Valderrama, R. Palacios, D. Romero, and G. Dávila. 1996. Transcriptional activity of the symbiotic plasmid of *Rhizobium etli* is affected by different environmental conditions. *Microbiology* **142**:2847–2856.
 23. Gonzalez, V., P. Bustos, M. A. Ramirez-Romero, A. Medrano-Soto, H. Salgado, I. Hernandez-Gonzalez, J. C. Hernandez-Celis, V. Quintero, G. Moreno-Hagelsieb, L. Girard, O. Rodriguez, M. Flores, M. A. Cevallos, J. Collado-Vides, D. Romero, and G. Davila. 2003. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol.* **4**:R36.
 24. Gonzalez, V., R. I. Santamaria, P. Bustos, I. Hernandez-Gonzalez, A. Medrano-Soto, G. Moreno-Hagelsieb, S. C. Janga, M. A. Ramirez, V. Jimenez-Jacinto, J. Collado-Vides, and G. Davila. 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci. U. S. A.* **103**:3834–3839.
 25. Hauser, F., A. Lindemann, S. Vuilleumier, A. Patrignani, R. Schlapbach, H. M. Fischer, and H. Hennecke. 2006. Design and validation of a partial-genome microarray for transcriptional profiling of the *Bradyrhizobium japonicum* symbiotic gene region. *Mol. Genet. Genomics* **275**:55–67.
 26. Hauser, F., G. Pessi, M. Friberg, C. Weber, N. Rusca, A. Lindemann, H. M. Fischer, and H. Hennecke. 2007. Dissection of the *Bradyrhizobium japonicum* NifA+ σ 54 regulon, and identification of a ferredoxin gene (*fixtN*) for symbiotic nitrogen fixation. *Mol. Genet. Genomics* **278**:255–271.
 27. Hawkins, F. K., and A. W. Johnston. 1988. Transcription of a *Rhizobium leguminosarum* biovar *phaseoli* gene needed for melanin synthesis is activated by *nifA* of *Rhizobium* and *Klebsiella pneumoniae*. *Mol. Microbiol.* **2**:331–337.
 28. Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush. 2000. A concise guide to cDNA microarray analysis. *Biotechnology* **29**:548–550, 552–554, 556.
 29. Hu, Y., A. W. Fay, C. C. Lee, J. Yoshizawa, and M. W. Ribbe. 2008. Assembly of nitrogenase MoFe protein. *Biochemistry* **47**:3973–3981.
 30. Hurkman, W. J., and C. K. Tanaka. 1986. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiol.* **81**:802–806.
 31. Jahn, O. J., G. Davila, D. Romero, and K. D. Noel. 2003. BacS: an abundant bacteroid protein in *Rhizobium etli* whose expression ex planta requires *nifA*. *Mol. Plant-Microbe Interact.* **16**:65–73.
 32. Kaminski, P. A., and C. Elmerich. 1998. The control of *Azorhizobium caulinodans nifA* expression by oxygen, ammonia and by the HF-1-like protein, NrfA. *Mol. Microbiol.* **28**:603–613.
 33. Li, F., B. Hou, and G. Hong. 2008. Symbiotic plasmid is required for NolR to fully repress nodulation genes in *Rhizobium leguminosarum* A34. *Acta Biochim. Biophys. Sin. (Shanghai)* **40**:901–907.
 34. L'Taief, B., B. Sifi, M. Zaman-Allah, J. J. Drevon, and M. Lachaal. 2007. Effect of salinity on root-nodule conductance to the oxygen diffusion in the *Cicer arietinum*-*Mesorhizobium ciceri* symbiosis. *J. Plant Physiol.* **164**:1028–1036.
 35. Luna-Moreno, D., O. Vazquez-Martinez, A. Baez-Ruiz, J. Ramirez, and M. Diaz-Munoz. 2007. Food restricted schedules promote differential lipoperoxidative activity in rat hepatic subcellular fractions. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **146**:632–643.
 36. Martinez, M., M. V. Colombo, J. M. Palacios, J. Imperial, and T. Ruiz-Argueso. 2008. Novel arrangement of enhancer sequences for NifA-dependent activation of the hydrogenase gene promoter in *Rhizobium leguminosarum* bv. viciae. *J. Bacteriol.* **190**:3185–3191.
 37. Martinez, M., J. M. Palacios, J. Imperial, and T. Ruiz-Argueso. 2004. Symbiotic autoregulation of *nifA* expression in *Rhizobium leguminosarum* bv. viciae. *J. Bacteriol.* **186**:6586–6594.
 38. Maurel, C., L. Verdoucq, D. T. Luu, and V. Santoni. 2008. Plant aquaporins: membrane channels with multiple integrated functions. *Annu. Rev. Plant Biol.* **59**:595–624.
 39. Mesa, S., F. Hauser, M. Friberg, E. Malaguti, H. M. Fischer, and H. Hennecke. 2008. Comprehensive assessment of the regulons controlled by the FixLJ-FixK2-FixK1 cascade in *Bradyrhizobium japonicum*. *J. Bacteriol.* **190**:6568–6579.
 40. Michiels, J., M. Moris, B. Dombrecht, C. Verreth, and J. Vanderleyden. 1998. Differential regulation of *Rhizobium etli rpoN2* gene expression during symbiosis and free-living growth. *J. Bacteriol.* **180**:3620–3628.
 41. Michiels, J., T. Van Soom, I. D'Hooghe, B. Dombrecht, T. Benhassine, P. de Wilde, and J. Vanderleyden. 1998. The *Rhizobium etli rpoN* locus: DNA sequence analysis and phenotypic characterization of *rpoN*, *ptsN*, and *ptsA* mutants. *J. Bacteriol.* **180**:1729–1740.
 42. Morett, E., and M. Buck. 1988. NifA-dependent in vivo protection demonstrates that the upstream activator sequence of *nif* promoters is a protein binding site. *Proc. Natl. Acad. Sci. U. S. A.* **85**:9401–9405.
 43. Moris, M., B. Dombrecht, C. Xi, J. Vanderleyden, and J. Michiels. 2004. Regulatory role of *Rhizobium etli* CNPAF512 *fnrN* during symbiosis. *Appl. Environ. Microbiol.* **70**:1287–1296.
 44. Olson, S. A. 2002. EMBOSS opens up sequence analysis. *European Molecular Biology Open Software Suite, Brief. Bioinform.* **3**:87–91.
 45. Peralta, H., Y. Mora, E. Salazar, S. Encarnacion, R. Palacios, and J. Mora. 2004. Engineering the *nifH* promoter region and abolishing poly-beta-hydroxybutyrate accumulation in *Rhizobium etli* enhance nitrogen fixation in symbiosis with *Phaseolus vulgaris*. *Appl. Environ. Microbiol.* **70**:3272–3281.
 46. Poggio, S., A. Osorio, G. Dreyfus, and L. Camarena. 2006. Transcriptional specificity of RpoN1 and RpoN2 involves differential recognition of the promoter sequences and specific interaction with the cognate activator proteins. *J. Biol. Chem.* **281**:27205–27215.
 47. Puppo, A., K. Groten, F. Bastian, R. Carzaniga, M. Soussi, M. M. Lucas, M. R. de Felipe, J. Harrison, H. Vanacker, and C. H. Foyer. 2005. Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. *New Phytol.* **165**:683–701.
 48. Reibach, P. H., P. L. Mask, and J. G. Streeter. 1981. A rapid one-step method for the isolation of bacteroids from root nodules of soybean plants, utilizing self-generating Percoll gradients. *Can. J. Microbiol.* **27**:491–495.
 49. Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**:276–277.
 50. Rojas-Mayorquin, A. E., N. M. Torres-Ruiz, D. Ortuno-Sahagun, and G. Gudino-Cabrera. 2008. Microarray analysis of striatal embryonic stem cells induced to differentiate by ensheathing cell conditioned media. *Dev. Dyn.* **237**:979–994.
 51. Segovia, L., J. P. Young, and E. Martinez-Romero. 1993. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *Int. J. Syst. Bacteriol.* **43**:374–377.
 52. Soldevila, G., I. Licona, A. Salgado, M. Ramirez, R. Chavez, and E. Garcia-Zepeda. 2004. Impaired chemokine-induced migration during T-cell development in the absence of Jak 3. *Immunology* **112**:191–200.
 53. Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**:33–36.
 54. Tun-Garrido, C., P. Bustos, V. Gonzalez, and S. Brom. 2003. Conjugative transfer of p42a from *Rhizobium etli* CFN42, which is required for mobilization of the symbiotic plasmid, is regulated by quorum sensing. *J. Bacteriol.* **185**:1681–1692.
 55. Valderrama, B., A. Davalos, L. Girard, E. Morett, and J. Mora. 1996. Regulatory proteins and cis-acting elements involved in the transcriptional control of *Rhizobium etli* reiterated *nifH* genes. *J. Bacteriol.* **178**:3119–3126.

11. RESULTADOS ADICIONALES

11.1 Artículo: “Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling”

En esta sección presentamos el artículo publicado en la revista “BMC Systems Biology”. En este se integran los resultados de transcriptómica y proteómica con el análisis *in silico* para la reconstrucción y predicción (caracterización) metabólica de *R. etli* CFN42 en la simbiosis. Mi principal aportación fue en la parte de transcriptómica.

RESEARCH ARTICLE

Open Access

Systems biology of bacterial nitrogen fixation: High-throughput technology and its integrative description with constraint-based modeling

Osbaldo Resendis-Antonio^{*}, Magdalena Hernández, Emmanuel Salazar, Sandra Contreras, Gabriel Martínez Batallar, Yolanda Mora and Sergio Encarnación^{*}

Abstract

Background: Bacterial nitrogen fixation is the biological process by which atmospheric nitrogen is uptaken by bacteroids located in plant root nodules and converted into ammonium through the enzymatic activity of nitrogenase. In practice, this biological process serves as a natural form of fertilization and its optimization has significant implications in sustainable agricultural programs. Currently, the advent of high-throughput technology supplies with valuable data that contribute to understanding the metabolic activity during bacterial nitrogen fixation. This undertaking is not trivial, and the development of computational methods useful in accomplishing an integrative, descriptive and predictive framework is a crucial issue to decoding the principles that regulated the metabolic activity of this biological process.

Results: In this work we present a systems biology description of the metabolic activity in bacterial nitrogen fixation. This was accomplished by an integrative analysis involving high-throughput data and constraint-based modeling to characterize the metabolic activity in *Rhizobium etli* bacteroids located at the root nodules of *Phaseolus vulgaris* (bean plant). Proteome and transcriptome technologies led us to identify 415 proteins and 689 up-regulated genes that orchestrate this biological process. Taking into account these data, we: 1) extended the metabolic reconstruction reported for *R. etli*; 2) simulated the metabolic activity during symbiotic nitrogen fixation; and 3) evaluated the *in silico* results in terms of bacteria phenotype. Notably, constraint-based modeling simulated nitrogen fixation activity in such a way that 76.83% of the enzymes and 69.48% of the genes were experimentally justified. Finally, to further assess the predictive scope of the computational model, gene deletion analysis was carried out on nine metabolic enzymes. Our model concluded that an altered metabolic activity on these enzymes induced different effects in nitrogen fixation, all of these in qualitative agreement with observations made in *R. etli* and other *Rhizobiaceas*.

Conclusions: In this work we present a genome scale study of the metabolic activity in bacterial nitrogen fixation. This approach leads us to construct a computational model that serves as a guide for 1) integrating high-throughput data, 2) describing and predicting metabolic activity, and 3) designing experiments to explore the genotype-phenotype relationship in bacterial nitrogen fixation.

^{*} Correspondence: resendis@ccg.unam.mx; encarnac@ccg.unam.mx
Programa de Genómica Funcional de Procaríotes. Centro de Ciencias
Genómicas-UNAM. Av. Universidad s/n, Col. Chamilpa, Cuernavaca Morelos,
C.P. 62210. Mexico

Background

Biological nitrogen fixation carried out by *Rhizobiaceas* represents nearly 70 percent of the entire nitrogen transformation required for maintaining life in our biosphere. Simultaneously, nitrogen fixation driven by these bacteria constitutes an appealing and natural strategy for developing sustainable agricultural programs due to its cost-effectiveness in crop improvement and its more environmentally friendly effects in comparison to those produced by chemical fertilizers [1]. Based on these fundamental and practical issues, the study of bacterial nitrogen fixation is one active line of research that in the post-genomic era demands new paradigms capable of surveying in a systematic fashion the metabolic organization by which this process occurs in nature.

At a molecular level, symbiotic nitrogen fixation arises as a consequence of the coordinated action of a variety of genes, proteins and metabolites that in turn activate signal transduction cascades and transcriptional factors inside bacteroids. At the end of the day, the consequences are the activation and repression of certain metabolic pathways whose end products are required for counteracting the microenvironmental conditions prevailing inside nodules [2-4]. The advent of high-throughput technologies has fostered the genome scale analysis for bacterial nitrogen fixation, and the output data constitute valuable material in deciphering their metabolic organization at different biological layers [5,6]. Although some significant results have been achieved in interpreting the high-throughput data, their overwhelming numbers and heterogeneous composition represent a challenge for inferring biological knowledge in a coherent and systematic fashion. This challenge is, indeed, a central issue in systems biology, and its solution demands integrative efforts among genome scale data, physiological knowledge and computational modeling [7-11].

With the purpose of contributing to this integrative challenge, in this paper we present a systems biology description in bacterial nitrogen fixation. In particular, it integrates high-throughput technology and flux balance analysis in order to explore the metabolic activity of *Rhizobium etli* bacteroids while they fix nitrogen in symbiotic association with *Phaseolus vulgaris* (common bean plant) [11]. To survey the bacterial phenotype and sketch the genetic and metabolic profile during nitrogen fixation, transcriptome and proteome technologies were carried out for *R. etli* bacteroids selected at 18 days after inoculation with root plants of *P. vulgaris* (see details in experimental procedure and methods). We selected this interval of time based on experimental knowledge that has indicated it as an average for maximum enzymatic activity of nitrogenase in *R. etli* bacteroids. To identify

those genes with a significant role in nitrogen fixation, we accomplished a comparative analysis between the gene expression profile at the nitrogen fixation stage and under free-living conditions in *R. etli*, this last condition mainly defined by succinate and ammonia as carbon and nitrogen sources, respectively (see methods). Simultaneously, the protein profile inside bacteroids was obtained, also at 18 days after plant inoculation. A set of genes with significant participation in bacterial nitrogen fixation was defined by combining those genes differentially expressed in the two physiological conditions—free life and nitrogen fixation— and those codifying for the proteins detected inside bacteroids. This same set of genes served as our benchmark for extending the metabolic reconstruction for *R. etli* metabolism (*iOR363*) and evaluating the consistency of the metabolic capacities inferred by the *in silico* analysis [8]. To assess the predictive scope of the model, we qualitatively compared the metabolic activity predicted by constraint-based modeling against that which was deduced from the high-throughput data obtained for *R. etli*. Overall, our study represents a significant effort toward the reconstruction of a systems biology platform for studying metabolic activity in bacterial nitrogen fixation. It is characterized by its capacity to integrate and describe high-throughput data and predict the metabolic mechanism underlying bacterial nitrogen fixation.

Results

High-throughput technology to guide the Metabolic Reconstruction

To characterize the gene expression during nitrogen fixation in *R. etli*, we compared each gene's activity in the free-living condition and in bacteroids driving nitrogen fixation selected at 18 days after inoculation with *P. vulgaris*. Data from microarray experiments were stored at the data depository *GEO* (<http://www.ncbi.nlm.nih.gov/geo/>) with access numbers GPL10081 for *R. etli* platform and GSE21638 for free life and symbiosis data. Even though a variety of sophisticated regulatory mechanisms may occur at diverse levels of biological organization [12], we have assumed that those genes with a significant over-expression indicate functional mechanisms for accomplishing nitrogen fixation. Under this criteria, we identified 689 genes (approximately 11% of the *R. etli* genome) whose transcriptional activity significantly increases during the biological process. To survey the role that these genes have in supporting nitrogen fixation, we classified them in accordance to the functional categories defined for *Rhizobiaceas* [13,14], see panel (A) in Figure 1 and Additional File 1. As expected, the majority of the *nif* and *fix* genes in bacteroids and other genes required for translation

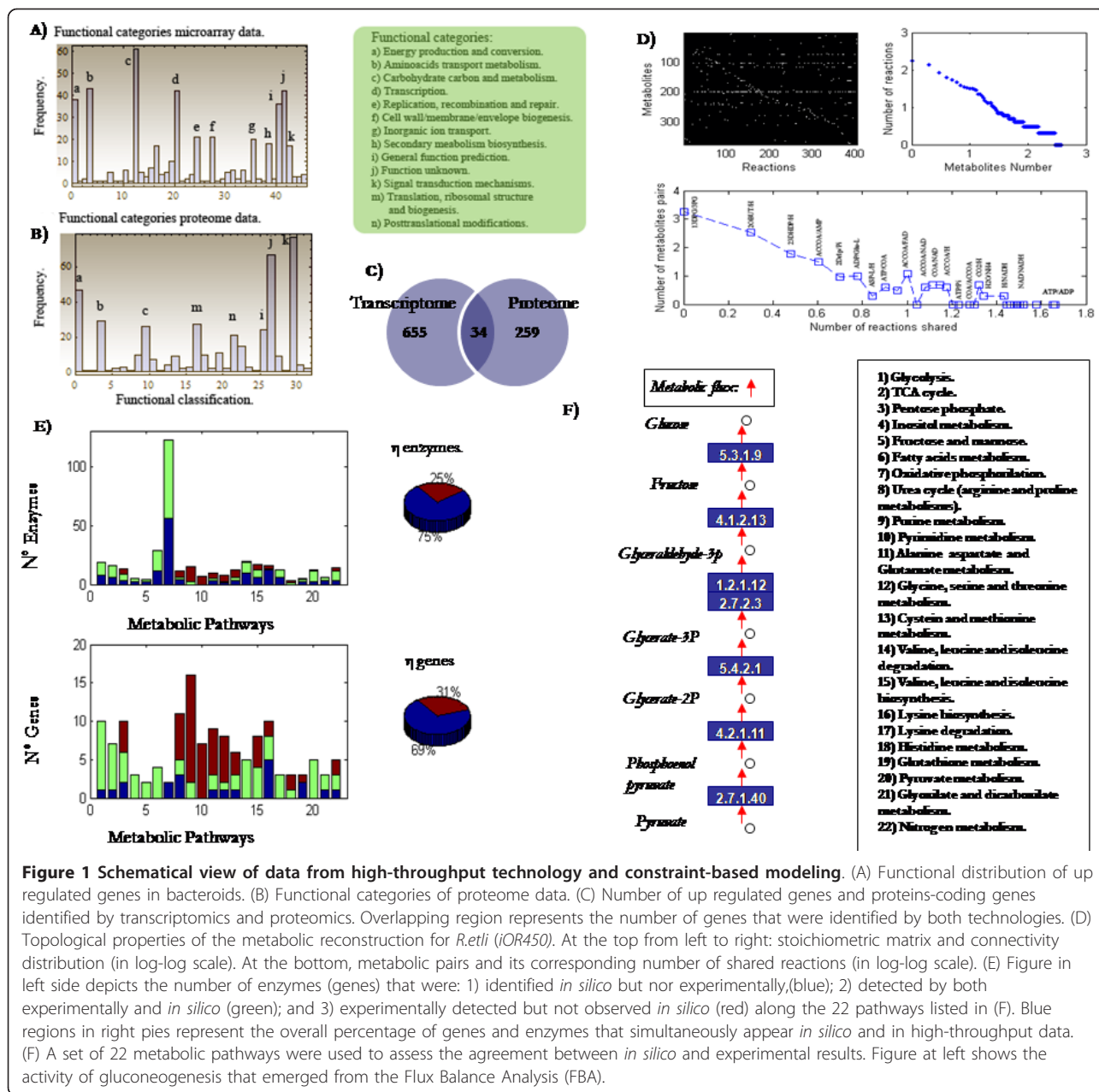


Figure 1 Schematic view of data from high-throughput technology and constraint-based modeling. (A) Functional distribution of up regulated genes in bacteroids. (B) Functional categories of proteome data. (C) Number of up regulated genes and proteins-coding genes identified by transcriptomics and proteomics. Overlapping region represents the number of genes that were identified by both technologies. (D) Topological properties of the metabolic reconstruction for *R. etli* (iOR450). At the top from left to right: stoichiometric matrix and connectivity distribution (in log-log scale). At the bottom, metabolic pairs and its corresponding number of shared reactions (in log-log scale). (E) Figure in left side depicts the number of enzymes (genes) that were: 1) identified *in silico* but nor experimentally, (blue); 2) detected by both experimentally and *in silico* (green); and 3) experimentally detected but not observed *in silico* (red) along the 22 pathways listed in (F). Blue regions in right pies represent the overall percentage of genes and enzymes that simultaneously appear *in silico* and in high-throughput data. (F) A set of 22 metabolic pathways were used to assess the agreement between *in silico* and experimental results. Figure at left shows the activity of gluconeogenesis that emerged from the Flux Balance Analysis (FBA).

initiation, elongation, and termination were up-regulated inside nodules. Furthermore, our data suggest that the expression of genes forming part of translation initiation, elongation and termination machinery was not absent although it was significantly reduced in the nodule bacteria, a common observation reported in *Bradyrhizobium japonicum*, *Sinorhizobium meliloti* and *Mesorhizobium loti* bacteroids [15-17]. In accordance with the induction of cell-division inhibitor protein *minD*, a significant number of housekeeping genes down-regulate their expression at nitrogen fixation

stages, and from microarray data we concluded that a slower rate of general metabolism, see Additional File 1.

To give a broader view of the biological activity inside the bacteroid, proteome analysis was conducted for *R. etli* bacteroids similarly recollected from nodules selected at 18 days after inoculation in root plants of *P. vulgaris* [18], see experimental procedure and methods. In total, proteome studies led us to identify and characterize 415 spot proteins in the bacteroids that suggested the expression of 293 genes during nitrogen fixation, see Figure 1 (B) and Additional File 2.

Both technologies—transcriptome and proteome—contributed to supply a broader biological landscape regarding bacterial nitrogen fixation. However, it is necessary to be aware of some differences in the experimental design underlying both technologies in order to integrate and interpret this data in a coherent fashion. While microarray technology resulted from a comparative analysis of two physiological conditions (free life and nitrogen fixation stages), proteome data identified the most abundant proteins present exclusively during nitrogen fixation stages. As Figure 1 (C) and Additional File 3 show, a scarce overlapping between the genes identified by both data sets is observed due to the experimental distinctness inherent in each technology. Thus, in order to identify those genes and enzymes with a relevant role in bacteroid metabolism and, in turn, form a set of genes that serve as a benchmark for computational assessment, we followed an integrative, rather than, selective strategy. Taking into account both sources of data, we hypothesized that up-regulated genes identified by microarray data and those genes that codify for the identified proteins potentially reveal those genes with a major role in nitrogen fixation. Under this assumption, both technologies led us to integrate a total of 948 genes that have a role in supporting bacterial nitrogen fixation, see Figure 1 and Additional File 1 and 2.

Functional classification of this set of genes ranged from enzymes participating in central metabolism and amino acid production to those maintaining specific pathways of nitrogen fixation such as glycogen and poly- β -hydroxybutyrate (PHB) biosynthesis. In addition, we identified enzymes participating in catabolism and anabolism of amino acids, chemotaxis, ribosome composition, RNA polymerase, DNA replication, nucleotide repairs, secretion systems and fatty acids metabolism. Moreover, a significant number of proteins participating as transporters reflects the intense metabolic crosstalk between plant and bacteroid; for instance, proteins participating in transport of small molecules, such as carbon, hydrogen, phosphate and sugar, fall under this classification, see panels (A) and (B) in Figure 1. We also identified proteins participating in the regulatory mechanism in nitrogen fixation, two components systems, transport and cell surface structure, energy transfer, cellular protection, and the transport and synthesis of polysaccharides. An extended discussion of the functional analysis that emerged from both technologies and its implication at a metabolic level can be reviewed in the Additional File 4.

Expanding *Rhizobium etli* metabolic reconstruction and selecting pathways for its experimental assessment

The data generated by high-throughput technology constitutes a cornerstone in moving toward a descriptive

analysis of nitrogen fixation. Despite the fact that this *top-down* scheme represents a valuable contribution to monitoring cell activity at a genome scale, complementary descriptions are required to integrate these data and survey how genetic perturbations affect nitrogen fixation in a systematic and quantitative fashion (*bottom-up* scheme). Among these quantitative schemes, constraint-based modeling is an appropriate formalism for exploring the cellular metabolic activity and guiding experiments to improve cellular behavior in a rational, coherent and optimal fashion [7,8,19,20]. In order to construct a *bottom-up* scheme for bacterial nitrogen fixation, our strategy consisted of three steps: 1) metabolic reconstruction for *R. etli*; 2) *in silico* modeling of nitrogen fixation, and 3) a cyclic assessment of computational predictions and experimental results.

In terms of metabolic reconstruction, proteome and transcriptome data were used to elaborate on the previous report for *R. etli* [8], thereby making some metabolic improvements and including new metabolic pathways absent in the previous version. To visually identify these metabolic reactions, we proceeded to represent the set of genes identified by high-throughput data and those from *iOR363* reconstruction into each metabolic pathway defined in KEGG database. A comparative analysis among each pathway led us to visualize and highlight their differences. Consistent with the previous metabolic reconstruction, certain reactions were identified in the experimental set of data, while others led us to postulate the activity of new metabolic pathways that were absent in the previous reconstruction [8]. Specifically, high-throughput data strongly indicated the biological activity of fatty acid metabolism, and we therefore included this pathway in the metabolic reconstruction, see supplementary material. Overall, a set of 405 reactions and 450 genes made up the new metabolic reconstruction for *R. etli* (*iOR450*) with which *in silico* simulations and analysis were carried out. Topological properties that emerged from the updated metabolic reconstruction are shown in Figure 1 (D).

To evaluate the concordance between the metabolic activity predicted *in silico* and that interpreted from high-throughput technology, we selected 22 KEGG metabolic pathways [21] that had the highest number of genes experimentally detected by high-throughput data see Figure 1 (F). According to the KEGG database, these 22 metabolic pathways contain 311 genes for *R. etli* of which 76.7% were included in the metabolic reconstruction *iOR450*. This set of genes and their corresponding enzymes constituted the central core for evaluating the coherence between *in silico* predictions and high-throughput data interpretations. Even though *in silico* assessment relies on the activity of 22 metabolic pathways, *in silico* analysis of nitrogen fixation took into

account all the reactions included in the metabolic reconstruction. This latter procedure will be valuable especially for exploring and predicting the metabolic role that additional pathways have on nitrogen fixation.

Constraint-based modeling: evaluating the descriptive and predictive capacities of the metabolic reconstruction

Constraint-based modeling is useful for predicting the metabolic phenotype in microorganisms surviving in specific environmental conditions and/or subject to genetic perturbations [7,22]. With the purpose of evaluating the phenotype capacities of the metabolic reconstruction, flux balance analysis (FBA) was carried out for *R. etli* by imposing physical and chemical constraints to each metabolic reaction and using an objective function that mimics symbiotic nitrogen fixation [8], see method section. As a result of this analysis, a set of enzymes and genes with a significant role in nitrogen fixation was identified *in silico* as those that underlie the metabolic fluxes obtained from FBA. To quantify the agreement of experimental and computational interpretations, we defined a consistency coefficient representing the fraction of genes (η^{Genes}) or enzymes ($\eta^{Enzymes}$) predicted active by FBA and detected by high-throughput technology, see methods section. This parameter ranges from 0 to 1, with 1 representing the highest and 0 the lowest consistency between the genes (or enzymes) detected from high-throughput technology and predicted *in silico*. To evaluate the numerical value of these parameters and estimate the coherence between modeling outputs and high-throughput data during nitrogen fixation, an early metabolic simulation on *iOR450* was carried out using the objective function originally suggested in a previous work, Z^{Fix} [8], i.e.

$$Z^{Fix} = \text{glycogen}[c] + \text{lys}[c] + \text{phb}[c] + \text{ala}[e] + \text{asp}[c] + \text{nh4}[c]$$

where glycogen, lysine, poly-hydroxybutyrate, alanine, aspartate and ammonium are denoted as *glycogen*[c], *lys* [c], *phb*[c], *ala*[e], *asp*[e] and *nh4*[e], respectively. All these metabolites are required to support an effective symbiotic nitrogen fixation [8], and their spatial location is indicated by [c] and [e] for cytoplasm and external compound. As a result of this simulation, we obtained a consistency coefficient of $\eta^{Genes} = 0.6835$ for genes and $\eta^{Enzymes} = 0.702$ for enzymes. Notably, this numerical value implied that 68.35% of the genes and 70.2% of the enzymes predicted *in silico* were consistently identified by high-throughput technology. To evaluate the statistical significance of this correlation, a hypergeometric test was applied in each case. In terms of enzymes, the coefficient reflected that of 74 enzymes predicted *in silico*, 52 were identified by high-throughput data. Meanwhile,

the gene consistency coefficient indicated that of 237 expressed genes, 162 were identified experimentally. In both cases we concluded that these correlations were statistically significant: $p\text{-value} = 8.59 \times 10^{-35}$ and $p\text{-value} = 4.9 \times 10^{-64}$ for genes and enzymes, respectively.

Improving predictability capacity of constraint-based modeling

These results encouraged us to proceed with an analysis of the *in silico* metabolic phenotype during nitrogen fixation, yet some improvements are desirable for ensuring a model with coherent interpretations and accurate predictions. To raise the qualitative agreement between *top-down* (high-throughput data) and *bottom-up* (*in silico* modeling) schemes, we therefore explored the possibility of finding an expanded objective function whose *in silico* phenotype improves the protein consistency coefficient η . To avoid this procedure from becoming a simple computational artifact without a biological foundation, we limited the search to those metabolites whose significant role in the bacterial nitrogen fixation were subject to strong experimental evidence. Thus, guided by a review in the literature, two metabolites were included in the objective function: *L-valine* and *L-histidine* both with a biologically meaningful role in nitrogen fixation. Supporting this assumption, mutagenesis made on the biosynthesis of branched chain amino acids, such as *L-valine*, has been shown to be defective in the initiation of nodule formation on host legumes [23]. In addition, we found evidence that *L-histidine* is a central compound participating in the mechanisms for regulating nitrogen fixation [12], and we noted that its inclusion in the objective function increased the agreement with high-throughput data. We therefore constructed a new objective function to mimic metabolic activity during nitrogen fixation in bacteria, it now integrated by

$$Z^{Fix} = \text{glycogen}[c] + \mathbf{his}[c] + \text{lys}[c] + \text{phb}[c] + \mathbf{val}[c] + \text{ala}[e] + \text{asp}[c] + \text{nh4}[c]$$

where boldface letters indicate those metabolites that were added to the previous objective function. Taking into account this implementation and simulating the flux distribution through FBA as described above, we obtained the following results during nitrogen fixation: $\eta^{Genes} = 0.6948$ and $\eta^{Enzymes} = 0.7683$, see Figure 1(E). In terms of enzyme activity this numeric value indicates that of the 82 metabolic reactions predicted *in silico*, 63 of them were consistently justified by high-throughput data ($p\text{-val} = 3.05 \times 10^{-64}$). Meanwhile the gene consistency coefficient indicated that of 249 expressed genes, 173 were identified by high-throughput data ($p\text{-value} = 4.9 \times 10^{-64}$).

Given this improvement, a detailed comparison between computational predictions and high-throughput data of the 22 metabolic pathways defined in Figure 1 (E) led us to distinguish three possible cases: the presence of 1) genes (enzymes) that were predicted *in silico* but not detected experimentally, 2) genes (enzymes) that were consistently observed in both schemes, and 3) genes (enzymes) that were experimentally detected but not predicted *in silico*, see Figure 1 (E). As explained in the methods section, η is related to the fraction of genes (enzymes) that were consistently observed in both schemes and constitutes the backbone of our modeling assessment. However, the biological explanation for the discrepancies described above (in cases 1 and 3) requires feedback assessment between modeling and experiments. For instance, these discrepancies could be reflecting the presence of post-transcriptional and post-translational regulation during nitrogen fixation and the design of proper experiments will be fundamental to discarding or accepting this hypothesis.

Discussion

A coherent description between *in silico* modeling and high-throughput data is a primary goal for exploring the fundamental principles governing metabolism in *Rhizobiaceas* and predicting their phenotype behavior during nitrogen fixation. In this work we present a systems biology framework capable of exploring the metabolic activity of *R. etli* during nitrogen fixation in symbiosis with *P. vulgaris*. In particular, we present a genome scale model that integrates high-throughput data for describing, simulating and guiding experiments dealing with metabolic activity in bacterial nitrogen fixation. An important issue in constraint-based optimization analysis is the presence of alternate optimal fluxes, in other words the presence of a set of reactions—or flux distributions—that produce the same quantitative objective function. As a consequence of these alternate fluxes, the metabolic output of one pathway can be substituted by others such that macroscopic phenotype remains constant. Therefore, the distinction of the reactions with and without a range of variability is essential to guess the metabolic activity supporting biological phenotype. Hence, in order to characterize the core metabolic activity and compare our *in silico* metabolic interpretations with those emerged from high-throughput data, we carried out flux variability analysis (FVA) [24]. With the purpose to identify those reactions that represent the central core of metabolic activity along the set of alternate solutions, we limited our analysis to those reactions with a range of variability equivalent to zero. This set was such that the minimum and maximum flux variability for each reaction were equivalent and constituted our cornerstone for guiding the metabolic activity

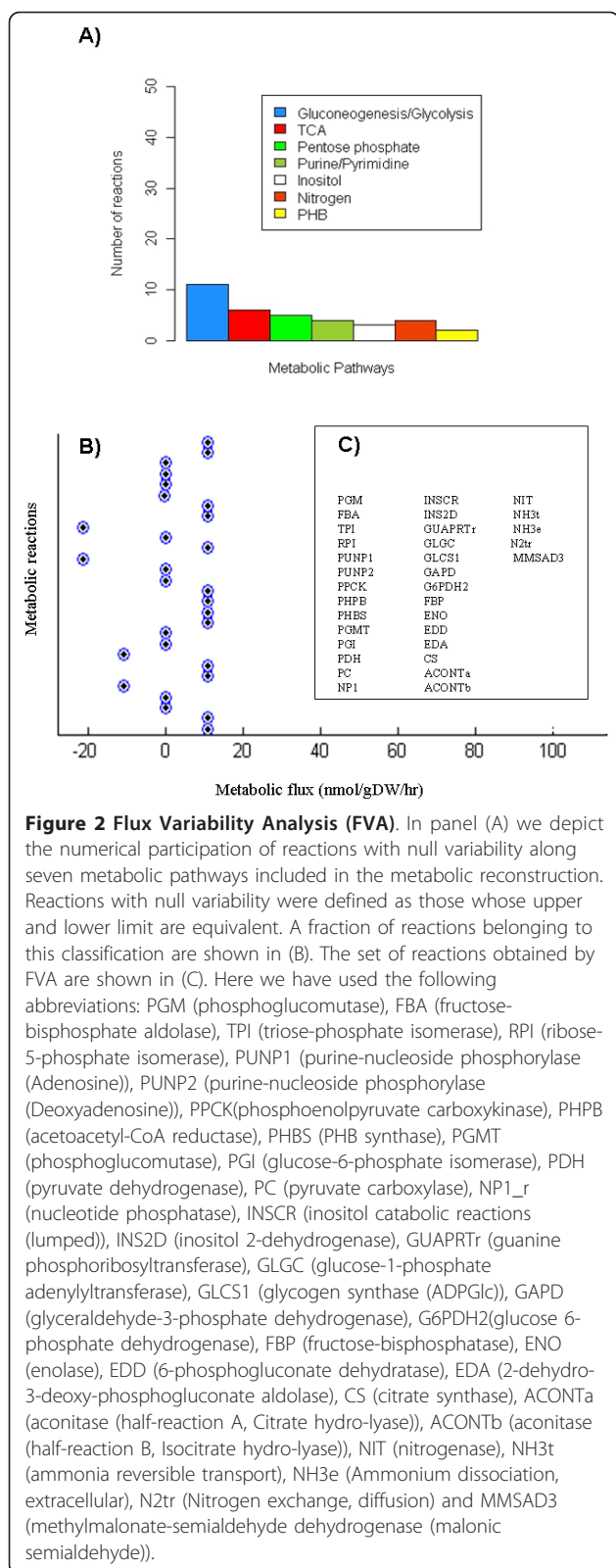
during the biological process. As depicted in Figure 2, the output of this analysis led us to identify some key reactions participating in some metabolic pathways required for sustaining bacterial nitrogen fixation. FVA was carried out with COBRA Toolbox [25]. As a consequence of this study, some concluding remarks immediately follow.

Citric acid cycle

Constraint-based modeling suggested that the TCA cycle is activated during nitrogen fixation by dicarboxylates which constitute the main carbon source in bacteroids [26], see Additional File 5 panel (B) in supplementary material. Consistent with this finding, eight proteins participating in the TCA cycle were detected in the *R. etli* bacteroid by proteome technology (FumC, FumB, LpdAch, SucB, SucA, SucC, Mdh and AcnA). To further assess this agreement, we applied gene deletion analysis to explore to what extent the deletion of some enzymes can qualitatively influence the activity of bacterial nitrogen fixation and if the predicted behavior is biologically coherent with knowledge reported in *Rhizobiaceas*, see method section. Thus, *in silico* gene deletion analysis accomplished on the metabolic reconstruction leads us to conclude that the aconitase hydratase (AcnA) mutant in *R. etli* is not lethal in nitrogen fixation. Despite the fact that this result has not been experimentally proven in *R. etli*, it has been validated in other *Rhizobiaceas* [27]. Furthermore, although isocitrate dehydrogenase (Icd) was not detected by high-throughput technology, *in silico* *icd* mutants in *R. etli* suggest a reduced phenotype on nitrogen fixation. This result is qualitatively in agreement with the fact that *icd* mutants on *S. meliloti* are symbiotically ineffective [28]. Similarly, constraint-based modeling concludes that a reduction of enzymatic activity in pyruvate dehydrogenase (PDH) induces a significant reduction in symbiotic nitrogen fixation but does not impair it as occurs in the case of *S. meliloti* bacteroids [29], see Figure 3 (B). This finding suggests that the role of PDH in the production of acetyl-coenzyme A can be replaced by alternative pathways in *R. etli* bacteroids [5]. The experimental assessment of this hypothesis for *R. etli* metabolisms is a central issue to explore in the future.

Glycolysis, gluconeogenesis and pentose phosphate pathways

A common metabolic trait for some *Rhizobiaceas* is the intense activity of gluconeogenesis pathway [3]. In agreement with this finding, a significant number of gluconeogenic and glycolytic enzymes were identified by high-throughput technology, and constraint-based modeling consistently concluded that gluconeogenesis pathway was actively participating in nitrogen fixation.



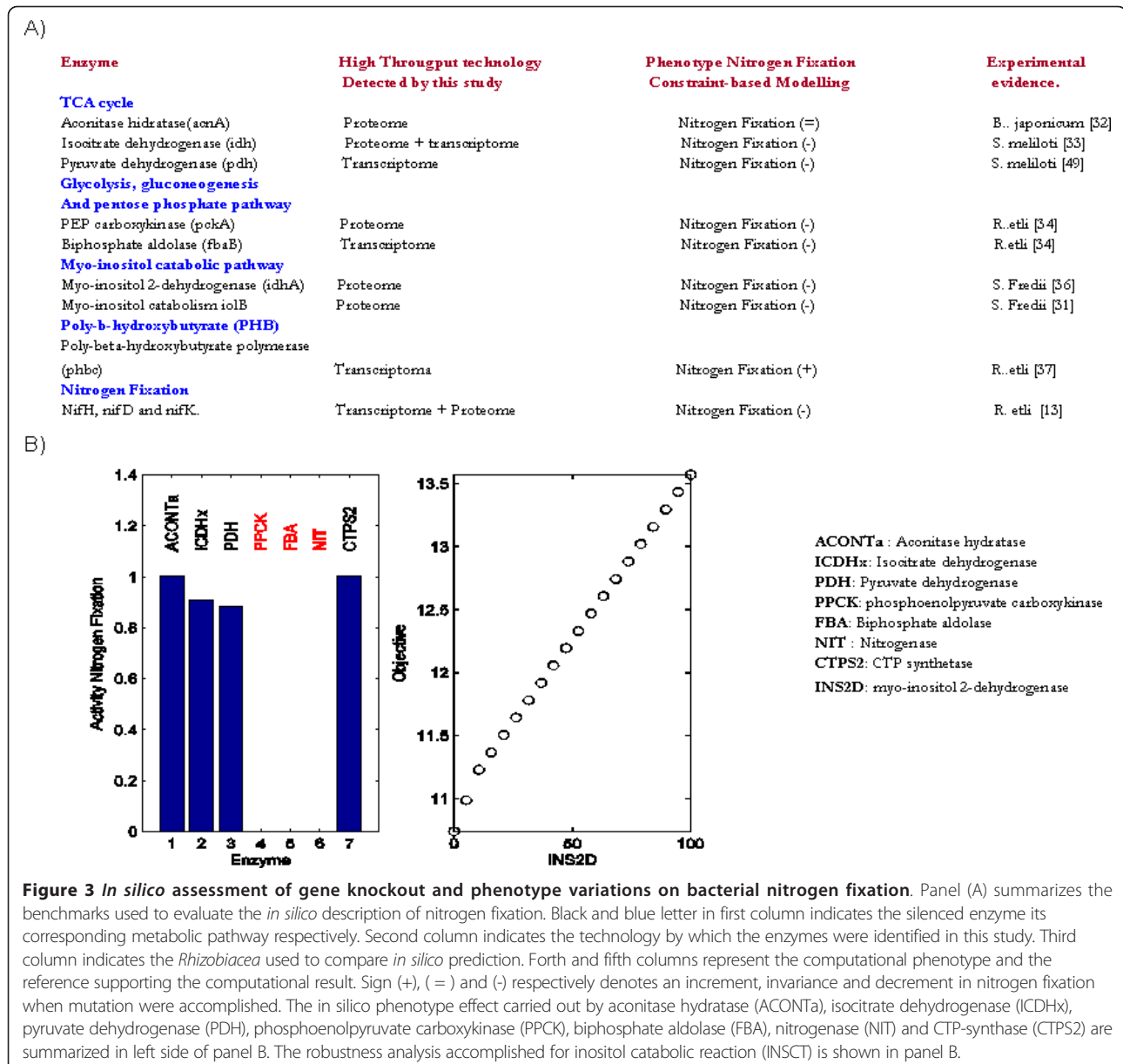
Multiple isoforms of PEP carboxykinase (*pckA*) were detected by proteome technology, see Additional File 2, mirroring their pivotal role in nitrogen fixation and bacteroid differentiation. Thus, *R. etli* CE3 *pckA* mutant produces few nodules into which the infection threads do not appear to penetrate [30]. In qualitative agreement with this report, *in silico* mutation suggests that *pckA* is an essential gene for accomplishing nitrogen fixation in *R. etli*, see Figure 3.

In addition, 6-phosphogluconolactonase (*pgl*), glucose 6-phosphate dehydrogenase (*Zwf1*), its chromosomal homolog (designated by *zwf2*) and one transaldolase (*Tal*) were detected by proteome, supplying evidence that pentose phosphate pathways can be actively participating in nitrogen fixation. Consistent with this finding, in fast-growing *Rhizobiaceas*, there is evidence that pentose phosphate and Entner-Doudoroff pathways work in coordinate action as the probable major routes for the metabolism of sugars [31].

As mentioned before, some glycolytic genes were identified by high-throughput data: two triosephosphate isomerases (*TpiAch* and *TpiAf*), one glyceraldehyde 3-phosphate dehydrogenase (*Gap*), one pyruvate kinase II (*PykA*), one 2-phosphoglycerate dehydratase (enolase), phosphoglycerate mutase (*pgm*), and the bisphosphate aldolase (*fbaB*), see Additional File 1 and 2. Furthermore, there is experimental evidence that the genetic silence of *fbaB* in *R. etli* causes the development of sparse, empty nodules on root beans [30]. Consistent with this fact, computational gene deletion analysis carried out with this gene confirms that *fbaB* has a crucial role in supporting the metabolism of bacterial nitrogen fixation [30], see Figure 3(B). Even though these findings were not enough to postulate an active glycolytic cycle, they may suggest the metabolism of sugar intermediates via other pathways. For example, the presence of a specific transporter for glycerol-3-phosphate (*ugpAch1*, induced 3.88-fold by microarray analysis) indicates that this may be an important source for generating glycolytic intermediates. Similarly, the expression of 6-phosphogluconate dehydrogenase (*Gnd*) suggests the presence of an active pentose pathway, which is another potential channel for the metabolism of glycolytic intermediates.

Myo-inositol catabolic pathway

Myo-inositol is one of the most abundant compounds in the soybean nodule, and accordingly, high-throughput technology successfully detected the presence of myo-inositol 2-dehydrogenase proteins (*IdhA* and *IoIB*) encoding a *myo-inositol* protein in catabolism [32]. In



agreement with this fact, computational analysis of the metabolism in *R. etli* suggests that a decrease of myo-inositol inside the nodule can reduce its capacity to fix nitrogen, see Figure 3(B). This result supports the hypothesis that the presence of myo-inositol in the nodule is essential for growth and maturation of the bacteroid and its metabolic inhibition can lead to both a nonfunctional bacteroid and the reduction of nitrogen-fixation activity [32].

Poly-β-hydroxybutyrate and glycogen accumulation

While most of the bacteroid carbon supplied by the plant is channeled into energy production to fuel nitrogen reduction, in certain types of nodules some carbon

is diverted by the bacteroids into the production of intracellular storage polymers composed of either glycogen or poly-β-hydroxybutyrate (PHB). Our simulations produced PHB, and consistent with our predictions, high-throughput analysis led us to identify the presence of three components related to its metabolic pathway: the polymerase PhbC (poly-beta hydroxybutyrate polymerase protein), a putative polyhydroxybutyrate depolymerase protein (detected by transcriptoma, see Additional File 1) and the acetyl-CoA acetyltransferase (beta-ketothiolase, *phbAch*) detected by proteome. Other reports confirm that metabolic fluxes in PHB and glycogen pathways are such that inhibition of one results in accumulation of the other, a property that was

consistently observed by *in silico* modeling [8,33,34]. The precise role of PHB and glycogen during infection, nodulation, and nitrogen fixation and the factors that induce their accumulation are not yet determined. Future experiments dealing with these pathways are necessary to elucidate their role in bacterial nitrogen fixation.

Nitrogen Fixation

To ensure the proper production of the ammonium required to establish an optimal bacterial-plant symbiosis, constraint-based modeling concludes that central genes involved in nitrogen fixation (*nif* and *fix* genes) are required for an optimal activity. Consistent with this fact, NifH, NifD and NifK were identified in proteome data and detected up-regulated in transcriptome analysis. In addition, an up-regulated gene expression was observed for *nifE* (nitrogenase reductase iron-molibdenum cofactor synthesis truncated protein), *nifN* (nitrogenase reductase iron-molibdenum cofactor synthesis protein), *nifX* (iron-molibdenum cofactor processing protein) and *nifB* (*FeMo* cofactor biosynthesis).

In *R. etli*, the *iscN* gene (Fe-S cofactor nitrogenase synthesis protein) is co-transcribed with *nifU* and *nifS*, and in conjunction, these genes were significantly up-regulated in bacteroids in comparison to bacteria under free-life condition (10.82, 3.92 and 1.99-fold, respectively). Furthermore, the *iscN* mutant in *R. etli* showed a significant reduction in nitrogen fixation [35]. Consistent with this report, *in silico* gene deletion analysis of those genes codifying for nitrogenase mostly reduces nitrogen fixation, see Figure 3.

Amino acid metabolism and transport

A previous report suggests that *Rhizobiaceas* require the availability of 20 amino acids to establish an effective symbiosis with legumes [36]. Some amino acids are synthesized by *Rhizobiaceas* whereas the remaining are supplied by the host plant, a condition that appears to be plant-type specific. High-throughput analysis led us to identify certain proteins required for the synthesis of arginine, tyrosine, tryptophan, phenylalanine and lysine, the latter participating in the objective function defined in constraint-based modeling. On the other side, from the ABC-transporter proteins founded in nodule bacteria, thirteen were involved in amino acid transport, it strongly suggests that the uptake of amino acid is of particular importance in nitrogen fixation. The general amino acid ABC-transporter protein for AapJ (substrate binding protein) was detected by proteome analysis: the *aapJ* gene is part of the *aapJQMP* operon that exists in many *Rhizobiaceas* and has been described in detail in *R. leguminosarum* [37]. BraC1 and braC2, of the branched-chain amino acid ABC transporter, were

detected in bacteroid by proteome and transcriptome technologies (2.85 fold). In *R. leguminosarum* *braDEFG* is required for alanine, histidine, leucine and arginine uptake [38] (two of which form part of the objective function associated with the metabolism of nitrogen fixation in our *in silico* model). Alternately, in *R. leguminosarum*, *braC* mutants are effective in alanine uptake (but are lacking in the uptake of the other three amino acids) [38]. Phenotype behavior for *braC* mutants has not been studied in *R. etli*, but there is evidence that *braD* and *braH* mutants were found to be deficient in glutamine uptake and respiration but proficient in nodulation and nitrogen fixation [30].

Nucleotides metabolism

Purine and pyrimidine pathways are important during the nodulation processes given that most purine or pyrimidine auxotrophs in *Rhizobiaceas* are ineffective in symbiotic nitrogen fixation because they elicit pseudo-nodules devoid of infection threads [39]. Thus, for instance, the *purB* and *purH* genes in *Mesorhizobium loti* are involved in infection thread formation and nodule development in *Lotus japonicus* [40]. In addition, *purB* and *purH* mutants exhibited purine auxotrophy and nodulation deficiency in *L. japonicus* [40]. As Figure 2(A) and Additional File 5 panel (C) shows in the supplementary material, constraint-based modeling concludes that some enzymes in purine and pyrimidine pathways are actively participating in reaching an optimal symbiotic nitrogen fixation. Supporting this finding, several key enzymes were identified in bacteroids by proteome technology. Among them, we identified: phosphoribosylamine-glycine ligase protein (PurD), adenylosuccinate lyase protein (PurB), phosphoribosylformylglycinamide synthetase protein (PurL), adenylosuccinate synthetase protein (PurA), IMP cyclohydrolase/phospho-ribosylaminoimidazole-carboxamide-formyltransferase protein (PurH), adenylylate kinase (ATP-AMP transphosphorylase, Adk) and nucleoside-diphosphate-kinase protein (Ndk).

In the presence of adenine, only the *purH* mutant induced nodule formation, and the *purB* mutant produced few infection threads, suggesting that 5-aminoimidazole-4-carboxamide ribonucleotide biosynthesis catalyzed by PurB is required for the establishment of symbiosis. In addition, *purL* mutants in *S. fredii* HH103 strain does not grow in minimal medium unless the culture is supplemented with thiamin and adenine or an intermediate of purine biosynthesis [41]. Furthermore, gene expression of *purC1*, phosphoribosylaminoimidazole-succinocarboxamide (SAICAR) synthetase protein, *purUch* (formyltetrahydrofolate deformylase protein), *gmk2* (*guanylate kinase* (GMP kinase protein) and *pyrE* (orotate phosphoribosyltransferase protein) were up-

regulated inside bacteroids between 2.3 to 6.35 fold. In *S. meliloti*, nodule development in the case of *pyrE/pyrF* mutants did not reach the extent observed in the parental strain. These results suggest that some of the intermediates and/or enzymes of the *pyrimidine* biosynthetic pathway play a key role in bacteroid transformation and nodule development [42], information that should be taken into account for constructing an improved objective function and ensuring a proper computational description in future analysis.

Fatty acids metabolisms

According to high-throughput data, metabolism of fatty acid can play a significant role in bacterial nitrogen fixation, this being in contrast to the drastic reduction of lipid biosynthesis observed in *B. japonicum* [43]. Thus, a variety of *fab* genes and proteins participating in fatty acid biosynthesis were detected by both methodologies (proteome and transcriptome). For instance, we detected by proteome the MccB subunit of methylcrotonyl-CoA carboxylase protein, acyl-CoA thiolase protein (FadA), enoyl-CoA hydratase protein (FadB1), enoyl-[acyl-carrier-protein] reductase (NADH) protein (FabI2) and S-malonyltransferase protein (FadD); and by transcriptome *fadB2* was induced 3.09-fold. As these findings suggest, fatty acid metabolism could play an important role in bacteroid metabolism given that it can supply a variety of precursors such as components of the *rhizobial* membrane, lipopolysaccharides and coenzymes required in signal transduction. As opposed to the process in other *Rhizobiaceas* where fatty acids can be supplied by the host plant [43], we supply experimental evidence that bacteroids of *R. etli* synthesize and metabolize their fatty acids. The assessment of this hypothesis and the biological implications on bacterial nitrogen fixation constitute an avenue to experimentally verify in the future.

Conclusions

In this study we present a systemic metabolic description of bacterial nitrogen fixation carried out by *R. etli* in symbiosis with *P. vulgaris*, at present the most complete study made in *Rhizobiaceas*. Collectively, high-throughput data suggest the following significant clues: 1) *R. etli* bacteroids are capable of synthesizing several amino acids through integrated carbon and nitrogen metabolisms. In addition, we observe the participation of some minor metabolic pathways such as myo-inositol catabolic pathway, degradation and synthesis of poly- β -hydroxybutyrate and glycogen. 2) Gene expression in bacteroids suggests the presence of a specialized transport system for sugars, proteins and ions. 3) An antioxidant defense mechanism based on peroxiredoxine, regulated by *nifA*, prevails during nitrogen

fixation, as opposed to in free-living condition, where the mechanism is rooted in catalases [44]. 4) *R. etli* over-expresses genes and enzymes required in fatty acid and nucleic acid metabolism, contrary to other studies in bacteroids. Finally, 5) this study contributes a computational model that serves as a useful framework for integrating data, designing experiments and predicting the phenotype during bacterial nitrogen fixation, see Figure 3.

This systemic and integrative approach constitutes a valuable effort toward a systems biology description of the metabolism in bacterial nitrogen fixation; however, to increase our understanding and predictive accuracy some issues should be addressed in the future. Thus, particular attention should be directed toward those enzymes that were predicted metabolically active *in silico* but were not detected experimentally, and conversely, those enzymes that were detected experimentally but not *in silico*, see Figure 1(E). We expect that the study of these differences will be fundamental in postulating, verifying and uncovering mechanisms of regulation, while simultaneously confirming or improving hypotheses derived through *in silico* predictions.

Notably, even though the simulations have been carried out without a detailed numerical description of the coefficients c_i in the objective function—see methods section—we have shown that the *in silico* model is capable of qualitatively predicting the activity of classic metabolic pathways and successfully describing some phenotype behavior in bacterial nitrogen fixation. Even though this represents a significant advance toward a systems biology description of bacterial nitrogen fixation, some improvements should be addressed in future. For instance, additional metabolites with a biological role in nitrogen fixation should be considered in order to obtain a more proper objective function that contributes to uncovering the role that less known metabolic pathways, such as nucleotides and fatty acid metabolisms, have on this biological process. As described here, these improvements will be guided by high-throughput data and the cyclic crosstalk between model and theory, a needed step in integrating, interpreting and generating biological hypotheses in a more accurate fashion.

Overall our study contributes to establishing the bases toward a systems biology platform capable of integrating high-throughput technology and computational simulation of bacterial nitrogen fixation. In particular, we envision that this metabolic reconstruction for *R. etli* (*iOR450*) will contribute to the rational design of optimal experiments that help us understand biological principles and identify those molecular mechanisms in order to improve this biological process, all this from a systems biology perspective.

Methods

Bacterial strains, growth conditions

The bacterial strain used was *R. etli* CFN42 wild type [11]. Culture media and growth conditions for *R. etli*, and plant experiments were accomplished as previously described in reference [45].

Plant experiments

Three-day-old *Phaseolus vulgaris* cv. Negro Jamapa seedlings were inoculated with *R. etli* CFN42 strains as previously described by Peralta et al. [46]. After 18 days post-inoculation (dpi), nodules were picked out from the roots, immediately frozen in liquid nitrogen and stored at -70°C until further use. Bacteria were isolated from nodules and their identities verified by their antibiotic resistances.

RNA isolation and microarray hybridization

Microarray experiments were carried out using three independently isolated RNA preparations from independent cultures and set of plants. Approximately 3 g of nodules were immersed in liquid nitrogen and macerated. Total RNA was isolated by acid hot-phenol extraction as described previously by de Vries et al [47]. For microaerobic free-living conditions, 50 ml of bacterial cell cultures were collected and total RNA isolated using a RNeasy Mini Kit (QIAGEN, Hilden, Germany). RNA concentration was determined by measuring the absorbance at 260 nm. The integrity of RNA was determined by running samples on a 1.3% agarose gel. 10 µg of RNA was differentially labeled with Cy3-dCTP and Cy5-dCTP using a CyScribe First-Strand cDNA labeling kit (Amersham Biosciences). Pairs of Cy3- and Cy5-labeled cDNA samples were mixed and hybridized to a *Rhizobium etli* CFN42_6051_v1.0 DNA microarray as described by Hegde et al. [48,49]. After washing, the arrays were scanned using a pixel size of 10 µm with a Scan Array Lite microarray scanner (Perkin-Elmer, Boston, MA). Three biological replicates with one dye swap were performed. We used real-time quantitative PCR to provide an independent analysis of gene expression for selected genes. Primer sequences and additional experimental protocols are reported in the supplementary material section.

DNA microarray analysis

Spot detection, mean signals, mean local background intensities, image segmentation, and signal quantification were determined for the microarray images using the Array-Pro Analyzer 4.0 software (Media Cybernetics, L.P). Statistical treatment of microarray data was accomplished with bioconductors software (<http://www.bioconductor.org/>). Specifically, microarray

normalization was carried out by applying the *maNormMain* function in the *marray* library. MA-plots before and after normalization are depicted in Additional File 5. Having normalized the gene expressions in the three experimental replicates, differentially expressed genes were identified by the following procedure. First, we calculate the average log-ratio for each gene obtained from the three experimental replicates. Then, we obtained the standardized *z-score* of the log-ratio associated to each gene. The set of genes differentially expressed during nitrogen fixation was selected as those genes with a *z-score* higher than 1.65, see Additional File 5. The complete dataset used in the transcriptome analysis can be downloaded from GEO (<http://www.ncbi.nlm.nih.gov/geo>) with accession numbers: GPL10081 for *Rhizobium etli* platform and GSE21638 for free-life and symbiosis data.

Verification by RT-PCR

We used real-time quantitative PCR to provide an independent assessment of gene expression for selected genes. The cDNA used for microarrays or freshly prepared cDNA was used as a template for Real-time PCR. Primer sequences used were as follows: *fabI2*-RECH000938f (5'-GTA TTG CCA AGG CCA TTC AT-3'), *fabI2*-RECH000938r (5'-CCC ACA GTT TTT CGA CGT TT-3') for the *fabI2* gene. *idhA*-RECH003170f (5'-TTT CTT CAT GAC CCG CTA CA-3'), *idhA*-RECH003170r (5'-TTG ATC AGC TTG CCT TCC TT-3') for the *idhA* gene. *ppk*-RECH001491f (5'-TCC TGG CAC TGA ACA CTC TG-3'), *ppk*-RECH001491r (5'-GAG AAG GAA CTG GAC CAC CA-3') for the *ppk* gene. *hisD*-RECH000581f (5'-GAT CTG AAG CAA GCC ATT CC 3'), *hisD*-RECH000581r (5'-ACA TAA TCG CCG ATG ACC TC-3') for the *hisD* gene. *nifH*-REPD00202f (5'-CCT CGG GCA GAA GAT CCT GA-3'), *nifH*-REPD00202r (5'-CAT CGC CGA GCA CGT CAT AG-3') for the *nifH* gene. *fixA*-REPD00224f (5'-ACA TCA ATG GGC GCG AGA TT-3'), *fixA*-REPD00224r (5'-TGT CGA TCT GCT CCG CCT TT-3') for the *fixA* gene. *cpxP2*-REPD00252f (5'-TCC GTG CCA TTT CAA AGA CC-3'), *cpxP2*-REPD00252r (5'-CCG CCA AAT GAG AAG ATT GC-3') for the *cpxP2* gene. *hisC*-RE1SP0000233f (5'-CGA TGG CGA GAC AGC TAA AT-3'), *hisC*-RE1SP0000233r (5'-ATC ATC GCA ACG CTA TCT CC-3') for the *hisC* gene. Each reaction contained 12.5 µl SYBR green PCR mastermix (Applied Biosystems), 3.5 µl H₂O, forward and reverse primers in a volume of 5 µl, and template in a volume of 4 µl. PCR reactions were run with the ABI Prism 7700 sequence detection system (Applied Biosystems) using the following steps: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles

of 95°C for 15 s and 60°C for 1 min. The dissociation protocol was 95°C for 15 s, 60°C for 20 s, followed by ramp from 60°C to 95°C for 20 min. The transcript of the histidinol phosphate aminotransferase protein (*hisCd*) was used as an internal (unregulated) reference for relative quantification. This gene was selected as a reference because its expression is constitutive in all tested conditions (free live and symbiosis). Results of RT-PCR in real time were analyzed using the $\Delta\Delta CT$ method [49] and the data was presented like relative expression. All reactions were done by triplicate.

Proteomics experiments

Bacteroids purification, protein extraction and two dimensional gel electrophoresis were done as previously described in [44]. Briefly, bacteroids were purified from root nodules by centrifugation through self-generated Percoll gradients. Bacteroid proteins were obtained by sonication at 24 kHz 1 min ON/1 min OFF for 5 cycles at 4°C in a Vibra Cell (Sonics, USA) in the presence of a protease inhibitor (Complete tablets, Roche Diagnostics GmbH, Mannheim, Germany). To further limit proteolysis, protein isolation was performed using phenol extraction. Two dimensional gel electrophoresis (2D-PAGE), was performed like previously described. Gels were stained with Coomassie Blue G-250, scanned with PDI image analysis system, and analyzed with PD-Quest software (Bio-Rad Laboratories, Inc, Hercules, CA.). Selected spots from preparative 2-D gels were excised, digested and the proteins were identified by PMF MALDI-TOF using a Bruker Daltonics Autoflex, following the same methodology mentioned in [44]. The experiments were performed three times. Selected spots from Coomassie stained preparative 2-D gels were excised and processing automatically using the Proteomeer SP spot picker and DP digestion robots (Bruker Daltonics, Billerica MA). Mass spectra were obtained using a Bruker Daltonics Autoflex (Bruker Daltonics Bellerica, Mass. USA) operated in the delayed extraction and reflectron mode. Spectra was externally calibrated using a peptide calibration standard (Bruker Daltonics 206095). Peak lists of the tryptic peptide masses were generated using FlexAnalysis1.2vSD1Patch2 (Bruker Daltonics). The search engine MASCOT server 2.0 was used to compare fingerprints against *Rhizobium etli* CFN42, NC_007761.1, pA, NC_007762.1, pB, NC_007763.1, pC, NC_007764.1, pD, NC_004041.2, pE, NC_007765.1, pF, NC_007766.1 with the following parameters: one missed cleavage allowed, carbamidomethyl cysteine as the fixed modification and oxidation of methionine as the variable modification. We

accepted those proteins with scores greater than 50 and a $p < 0.05$. Proteome data associated with this manuscript can be downloaded from <http://ProteomeCommons.org> Tranche using the following hash:

```
BY/eCcVjwTWN1+m+2ArvJ0QVnesGx5Ekgd4wUOA-SACfm/ueNI7YI3iLf4xz0lnGsepV5LkpMWOQOrZt-jYExlNpQkIBcAAAAAAAABJA = =
```

High-throughput technology and its use for extending metabolic reconstruction and simulating nitrogen fixation

With the purpose of establishing an integrative description between modeling and experimental data, we extended the metabolic reconstruction for *R. etli* by including those reactions whose enzyme activity were supported by high-throughput data. Thus, the fatty acids metabolism was included in the metabolic reconstruction, and some metabolic improvements were made along the network. Additional File 6 enlist the main abbreviations used along this paper. Additional File 7 in supplementary material contains a detailed description of the reactions included in this new metabolic version (*iOR450*). Overall, the updated metabolic reconstruction for *R. etli* consists of 377 metabolites and 450 genes codifying for enzymes participating in 405 metabolic reactions. The gene-protein reaction association for the entire metabolic reconstruction, lower and upper bounds and reversibility information associated to each reaction are shown in Additional File 7.

Constraint-based modeling

Metabolic flux distribution supporting nitrogen fixation in *Rhizobium etli* was predicted *in silico* by constraint-based modeling [8]. Briefly, simulations were carried out assuming a steady-state condition for metabolic fluxes and by constructing a mathematical function that mimics nitrogen fixation. This objective function, Z^{Fix} , consists of certain key compounds required for sustaining nitrogen fixation and others required for mimicking the physiological conditions prevailing in the boundaries of the nodules. Thus, objective function was mathematically written as a linear combination of these metabolites (X_i) and their contribution to nitrogen fixation was weighted by coefficients (c_i), which for simplicity's sake were all selected as a unit. With the purpose of obtaining a computational profile of metabolic fluxes, we assumed that the metabolic state of the bacteroid during nitrogen fixation is one that optimizes the objective function, Z^{Fix} . This latter issue was solved by taking into account linear programming and considering that the fluxes are constrained by their enzymatic and thermodynamic capacities,

$$\max \left[Z^{\text{Fix}} = \sum_{i=1} C_i \cdot X_i \right]$$

such that

$$\begin{aligned} \sum S_{ij} \cdot v_j &= 0 & i = 1, 2 \dots m \\ -\alpha_j \leq v_j \leq \beta_j & & j = 1, 2 \dots n \end{aligned}$$

where S_{ij} represents the entries of the stoichiometric matrix, v_j is the metabolic flux of the j -th reaction and α_j and β_j account for thermodynamic and enzymatic constraints, see Additional File 7. Linear programming was carried out using the *Tomlab* optimization package called from COBRA toolbox in *Matlab* [25].

External metabolites considered for flux balance analysis

In order to explore the phenotype capacities of the bacteria metabolism, we included in the reconstruction certain exchange and sink reactions for limiting our metabolic modeling and representing the microenvironmental conditions in the plant nodules. In general, these can be classified as one of two categories. Class I includes those metabolites that can be exchanged between the bacteroid membrane and the plant environment. Among them, we included carbon dioxide (CO₂), water (H₂O), oxygen (O₂), malate (mal-L) and glutamate (glu-L). In addition, exchange reactions for nitrogen (n2), alanine (ala-L), aspartate (asp-L), succinate and ammonium (NH₄) were included in the reconstruction for representing their possible bidirectional exchange from plant to bacteroids. On the other hand, metabolites in class II include those that contribute to the defining of internal frontiers in the bacteroids. Importantly, these sink reactions were included as a representation of metabolites originating from metabolic processes currently absent in the metabolic reconstruction. Thus, phosphate (pi), myo-inositol (inost), L-histidinol phosphate (hisp), palmitoyl-CoA (pmtcoa), dodecanoyl-CoA (dodecoa), decanoyl-CoA (decoa), octanoyl-CoA (otcoa) and hydrogen (h) fall in this classification.

Definition of consistency coefficient

To assess the agreement between *in silico* predictions and interpretations suggested by high-throughput data, we defined a consistency coefficient, η^{Genes} , that quantifies the fraction of genes that were predicted upregulated *in silico* and simultaneously detected or induced by proteome or transcriptome technologies. Simultaneously, we defined a consistency coefficient that quantifies the fraction of proteins enzymatically active that were predicted by constraint-based modeling and confirmed by high-throughput technology, η^{Enzymes} . To proceed with this evaluation, we denoted $E^j_{\text{kegg}}(G^j_{\text{kegg}})$ as the set of enzymes (genes) that form the j -th metabolic pathways in

KEGG database, with j -th ranging from 1 to 22. Similarly, the set of enzymes (genes) that integrates the i -th metabolic pathway in the reconstruction and the set of enzymes detected by high-throughput data are denoted by $E^j_{\text{Rec}}(G^j_{\text{Rec}})$ and $E^j_{\text{HT}}(G^j_{\text{HT}})$, respectively. Finally, the sets of enzymes and genes obtained from constraint-based modeling were denoted by $E^j_{i\text{Model}}$ and $G^j_{i\text{Model}}$. More specifically, $E^j_{i\text{Model}}$ and $G^j_{i\text{Model}}$ sets were defined as those enzymes and genes participating in the active fluxes obtained from flux balance analysis. In order to evaluate and create a proper framework for comparison between *in silico* predictions and high-throughput data, we defined the consistency coefficient as the fraction of enzymes (genes) that were actively predicted *in silico* and were identified by high-throughput technology. This can be summed up in the following equations:

$$\begin{aligned} \eta_{\text{Gene}} &= \frac{G^j_{i\text{Model}} \cap G^j_{\text{HT}}}{G^j_{i\text{Model}}} \\ \eta_{\text{Enzyme}} &= \frac{E^j_{i\text{Model}} \cap E^j_{\text{HT}}}{E^j_{i\text{Model}}} \end{aligned}$$

Both ratios range from zero to one and constitute our central parameter to assess and quantify the degree of coherence between constraint-based modeling and experimental data.

In silico gene deletion analysis

Computational gene deletion analysis was used to quantify the effects that gene silencing has in supporting bacterial nitrogen fixation. Thus, once the gene to be switched off was selected, we identified its gene-protein reaction association and selected as zero its upper and lower bound in flux activity. Having made this adjustment, we applied flux balance analysis and obtained the new resulting objective function. In order to quantify the participation of this metabolic reaction in bacterial nitrogen fixation, we calculated the percentage of reduced activity of the mutated strain in comparison to the wild type, see Figure 3.

Additional material

Additional file 1: Microarray Data Analysis. This table shows those genes that were over expressed during bacteroid activity in nitrogen fixation.

Additional file 2: Proteome Data. By using mass spectrometry, we identified a set of proteins during bacterial nitrogen fixation for *R. etli*. In each row, we named the protein and presented some of the parameter utilized for concluding the protein identify.

Additional file 3: Intersect between proteome and transcriptome. Genes that were simultaneously identified by proteome and transcriptome technologies.

Additional file 4: This file contains an extended descriptive analysis deduced from the genes identified by proteome and transcriptome data.

Additional file 5: (A) MA plot and representation of Metabolic activity. In this figure we show the MA-plot obtained from microarray data and a selected representation of the metabolic activity predicted by FBA in some metabolic pathways: (B) TCA cycle, (C) purine and pyrimidine metabolism.

Additional file 6: Abbreviations. This file enlists the main abbreviations used along the paper.

Additional file 7: Metabolic Reconstruction for *Rhizobium etli*. This table depicts the gene-protein-reaction association for *Rhizobium etli* metabolic reconstruction, *iOR450*. Overall the reconstruction contains 450 genes codifying for a set of enzymes participating in 405 metabolic reactions and 377 metabolites.

Acknowledgements

The authors thank Prof Jaime Mora Head of the Program of Functional Genomics of Prokaryotes at the Center of Genomic Sciences-UNAM for his support and comments. OR-A also thanks to Prof. B.Ø. Palsson for his guiding and encouraging support during the progress of this study. We thank Oliver Castillo, J. L. Zitalpopoca, and Hadau Sánchez for plant experiments and greenhouse support, and María del Carmen Vargas for technical assistance. Finally, the authors appreciate the valuable comments and suggestions from two anonymous referees during the review process. This work was supported by combined grants from National Council of Science and Technology CONACyT-Mexico, grants 83461 (OR-A) and 60641 (SE), and from PAPIIT-DGAPA-UNAM through grants IN222707 (SE) and IN203809-3 (OR-A).

Authors' contributions

OR-A conceived the metabolic reconstruction for *R. etli* and designed the computational analysis to simulate nitrogen fixation. ES realized microarray experiments, and OR-A contributed to their statistical analysis. ME and GM-B prepared the samples and accomplished the identification and processing in spectrometry analysis of all proteome data. YM guided the experimental cultivation for nodule preparation in *R. etli*. SE conceived the study, design and analyzed the experimental data obtained from proteome and microarray technologies. All authors read and approved the final manuscript.

Received: 10 March 2011 Accepted: 29 July 2011

Published: 29 July 2011

References

1. Diaz RJ, Rosenberg R: **Spreading dead zones and consequences for marine ecosystems.** *Science* 2008, **321**:926-929.
2. Deakin WJ, Broughton WJ: **Symbiotic use of pathogenic strategies: rhizobial protein secretion systems.** *Nat Rev Microbiol* 2009, **7**:312-320.
3. Lodwig E, Poole P: **Metabolism of Rhizobium bacteroids.** *Critical Reviews in Plant Sciences* 2003, **22**:37-78.
4. Prell J, Poole P: **Metabolic changes of rhizobia in legume nodules.** *Trends Microbiol* 2006, **14**:161-168.
5. Sarma AD, Emerich DW: **Global protein expression pattern of Bradyrhizobium japonicum bacteroids: a prelude to functional proteomics.** *Proteomics* 2005, **5**:4170-4184.
6. Oehrlé NW, Sarma AD, Waters JK, Emerich DW: **Proteomic analysis of soybean nodule cytosol.** *Phytochemistry* 2008, **69**:2426-2438.
7. Feist AM, Palsson BO: **The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli.** *Nat Biotechnol* 2008, **26**:659-667.
8. Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BO: **Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli.** *PLoS Comput Biol* 2007, **3**:1887-1895.
9. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA, Palsson B, Osterman A, Godzik A: **Three-dimensional structural view of the central metabolic network of Thermotoga maritima.** *Science* 2009, **325**:1544-1549.
10. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO: **Integrating high-throughput and computational data elucidates bacterial networks.** *Nature* 2004, **429**:92-96.
11. Gonzalez V, Santamaria RI, Bustos P, Hernandez-Gonzalez I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramirez MA, Jimenez-Jacinto V, Collado-Vides J, Davila G: **The partitioned Rhizobium etli genome: genetic and metabolic redundancy in seven interacting replicons.** *Proc Natl Acad Sci USA* 2006, **103**:3834-3839.
12. Dixon R, Kahn D: **Genetic regulation of biological nitrogen fixation.** *Nat Rev Microbiol* 2004, **2**:621-631.
13. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631-637.
14. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
15. Barnett MJ, Toman CJ, Fisher RF, Long SR: **A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction.** *Proc Natl Acad Sci USA* 2004, **101**:16636-16641.
16. Becker A, Berges H, Krol E, Bruand C, Ruberg S, Capela D, Lauber E, Meilhoc E, Ampe F, de Bruijn FJ, Fourment J, Francez-Charlot A, Kahn D, Küster H, Liebe C, Pühler A, Weidner S, Batut J: **Global changes in gene expression in Sinorhizobium meliloti 1021 under microoxic and symbiotic conditions.** *Mol Plant Microbe Interact* 2004, **17**:292-303.
17. Uchiumi T, Ohwada T, Itakura M, Mitsui H, Nukui N, Dawadi P, Kaneko T, Tabata S, Yokoyama T, Tejima K, Saeki K, Omori H, Hayashi M, Maekawa T, Sriprang R, Murooka Y, Tajima S, Simomura K, Nomura M, Suzuki A, Shimoda Y, Sioya K, Abe M, Minamisawa K: **Expression islands clustered on the symbiosis island of the Mesorhizobium loti genome.** *J Bacteriol* 2004, **186**:2439-2448.
18. Encarnacion S, Hernandez M, Martinez-Batallar G, Contreras S, Vargas Mdel C, Mora J: **Comparative proteomics using 2-D gel electrophoresis and mass spectrometry as tools to dissect stimulons and regulons in bacteria with sequenced or partially sequenced genomes.** *Biol Proced Online* 2005, **7**:117-135.
19. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO: **Reconstruction of biochemical networks in microorganisms.** *Nat Rev Microbiol* 2009, **7**:129-143.
20. Hyduke DR, Palsson BO: **Towards genome-scale signalling-network reconstructions.** *Nat Rev Genet* 11:297-307.
21. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
22. Segre D, Vitkup D, Church GM: **Analysis of optimality in natural and perturbed metabolic networks.** *Proc Natl Acad Sci USA* 2002, **99**:15112-15117.
23. de las Nieves, Peltzer M, Roques N, Poinso V, Aguilar OM, Batut J, Capela D: **Auxotrophy accounts for nodulation defect of most Sinorhizobium meliloti mutants in the branched-chain amino acid biosynthesis pathway.** *Mol Plant Microbe Interact* 2008, **21**:1232-1241.
24. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.** *Metab Eng* 2003, **5**:264-276.
25. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.** *Nat Protoc* 2007, **2**:727-738.
26. Dunn MF: **Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia.** *FEMS Microbiol Rev* 1998, **22**:105-123.
27. Thony-Meyer L, Kunzler P: **The Bradyrhizobium japonicum aconitase gene (acnA) is important for free-living growth but not for an effective root nodule symbiosis.** *J Bacteriol* 1996, **178**:6166-6172.
28. McDermott TR, Kahn ML: **Cloning and mutagenesis of the Rhizobium meliloti isocitrate dehydrogenase gene.** *J Bacteriol* 1992, **174**:4790-4797.
29. Soto MJ, Sanjuan J, Olivares J: **The disruption of a gene encoding a putative arylesterase impairs pyruvate dehydrogenase complex activity and nitrogen fixation in Sinorhizobium meliloti.** *Mol Plant Microbe Interact* 2001, **14**:811-815.
30. Tate R, Ferraioli S, Filosa S, Cermola M, Riccio A, Iaccarino M, Patriarca EJ: **Glutamine utilization by Rhizobium etli.** *Mol Plant Microbe Interact* 2004, **17**:720-728.

31. Romanov VI, Hernandez-Lucas I, Martinez-Romero E: **Carbon Metabolism Enzymes of Rhizobium tropici Cultures and Bacteroids.** *Appl Environ Microbiol* 1994, **60**:2339-2342.
32. Jiang G, Krishnan AH, Kim YW, Wacek TJ, Krishnan HB: **A functional myo-inositol dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of Sinorhizobium fredii USDA191 to nodulate soybean (Glycine max [L.] Merr.).** *J Bacteriol* 2001, **183**:2595-2604.
33. Cevallos MA, Encarnacion S, Leija A, Mora Y, Mora J: **Genetic and physiological characterization of a Rhizobium etli mutant strain unable to synthesize poly-beta-hydroxybutyrate.** *J Bacteriol* 1996, **178**:1646-1654.
34. Encarnacion S, del Carmen Vargas M, Dunn MF, Davalos A, Mendoza G, Mora Y, Mora J: **AniA regulates reserve polymer accumulation and global protein expression in Rhizobium etli.** *J Bacteriol* 2002, **184**:2287-2295.
35. Dombrecht B, Tesfay MZ, Verreth C, Heusdens C, Napoles MC, Vanderleyden J, Michiels J: **The Rhizobium etli gene iscN is highly expressed in bacteroids and required for nitrogen fixation.** *Mol Genet Genomics* 2002, **267**:820-828.
36. Randhawa GS, Hassani R: **Role of rhizobial biosynthetic pathways of amino acids, nucleotide bases and vitamins in symbiosis.** *Indian J Exp Biol* 2002, **40**:755-764.
37. Walshaw DL, Poole PS: **The general L-amino acid permease of Rhizobium leguminosarum is an ABC uptake system that also influences efflux of solutes.** *Mol Microbiol* 1996, **21**:1239-1252.
38. Hosie AH, Allaway D, Galloway CS, Dunsby HA, Poole PS: **Rhizobium leguminosarum has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family.** *J Bacteriol* 2002, **184**:4071-4080.
39. Newman JD, Diebold RJ, Schultz BW, Noel KD: **Infection of soybean and pea nodules by Rhizobium spp. purine auxotrophs in the presence of 5-aminoimidazole-4-carboxamide riboside.** *J Bacteriol* 1994, **176**:3286-3294.
40. Okazaki S, Hattori Y, Saeki K: **The Mesorhizobium loti purB gene is involved in infection thread formation and nodule development in Lotus japonicus.** *J Bacteriol* 2007, **189**:8347-8352.
41. Buendia-Claveria AM, Moussaid A, Ollero FJ, Vinardell JM, Torres A, Moreno J, Gil-Serrano AM, Rodriguez-Carvajal MA, Tejero-Mateo P, Peart JL, Brewin NJ, Ruiz-Sainz JE: **A purL mutant of Sinorhizobium fredii HH103 is symbiotically defective and altered in its lipopolysaccharide.** *Microbiology* 2003, **149**:1807-1818.
42. Vineetha KE, Vij N, Prasad CK, Hassani R, Randhawa GS: **Ultrastructural studies on nodules induced by pyrimidine auxotrophs of Sinorhizobium meliloti.** *Indian J Exp Biol* 2001, **39**:371-377.
43. Sarma AD, Emerich DW: **A comparative proteomic evaluation of culture grown vs nodule isolated Bradyrhizobium japonicum.** *Proteomics* 2006, **6**:3008-3028.
44. Salazar E, Diaz-Mejia JJ, Moreno-Hagelsieb G, Martinez-Batallar G, Mora Y, Mora J, Encarnacion S: **Characterization of the NifA-RpoN regulon in Rhizobium etli in free life and in symbiosis with Phaseolus vulgaris.** *Appl Environ Microbiol* 2010, **76**:4510-4520.
45. Encarnacion S, Dunn M, Willms K, Mora J: **Fermentative and aerobic metabolism in Rhizobium etli.** *J Bacteriol* 1995, **177**:3058-3066.
46. Peralta H, Mora Y, Salazar E, Encarnacion S, Palacios R, Mora J: **Engineering the nifH promoter region and abolishing poly-beta-hydroxybutyrate accumulation in Rhizobium etli enhance nitrogen fixation in symbiosis with Phaseolus vulgaris.** *Appl Environ Microbiol* 2004, **70**:3272-3281.
47. Vries S, Hoge H, Bisseling : **Isolation of total and polysomal RNA from plant tissues.** *Plant Molecular Biology Manual* 1988, **6**:1-13.
48. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J: **A concise guide to cDNA microarray analysis.** *Biotechniques* 2000, **29**:548-550, 552-544, 556 passim.
49. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**:402-408.

doi:10.1186/1752-0509-5-120

Cite this article as: Resendis-Antonio et al.: Systems biology of bacterial nitrogen fixation: High-throughput technology and its integrative description with constraint-based modeling. *BMC Systems Biology* 2011 5:120.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



11.2. Análisis de la comparación de la vida libre (VL) contra Simbiosis 18 ddi

La capacidad de medir la expresión de miles de genes hizo de los microarreglos una herramienta eficaz para evaluar la transcripción durante la simbiosis y otras condiciones de los organismos involucrados en la fijación biológica de nitrógeno. Como previamente mencionamos *R. etli* tiene la capacidad de vivir en vida libre y en simbiosis con *P. vulgaris*, por lo que utilizando el microarreglo diseñado y construido por nuestro grupo consideramos relevante evaluar la expresión a nivel transcripcional de *R. etli* comparando estas dos condiciones. Esta parte de la tesis se realizó con la finalidad de ampliar el análisis presentado en el artículo No. 2 (Resendis-Antonio *et al.*, 2011), en el cual se presentan los datos de los genes mayormente expresados en simbiosis (18 ddi).

El análisis de los datos de la comparación entre vida libre (6hrs MM) y simbiosis (18 ddi) nos aportó la siguiente información: el total de genes expresados en la simbiosis es de 689 y de vida libre son 385. En *B. japonicum* un análisis similar muestra que 692 genes fueron inducidos en los bacteroides respecto a la aerobiosis (VL), de estos el 54% son controlados por RegR. El experimento se hizo en nódulos de 21 ddi (Lindemann *et al.*, 2007). En *R. etli*, los 689 genes expresados en simbiosis se distribuyen de la siguiente manera: 363 (52.69%) están localizados en cromosoma, 21(3.05%) en el plásmido A, 23 (3.34%) en el plásmido B, 27 (3.92%) en el plásmido C, 91 (13.21%) en el plásmido D, 85 (12.34%) en el plásmido E y 79 (11.47%) en el plásmido F. En tanto en vida libre: 298 (77.4%) están localizados en el cromosoma, 9 (2.34%) en el plásmido A, 8 (2.08%) en el plásmido B, 10 (2.60%) en el plásmido C, 10 (2.60%) en el plásmido D, 15 (3.90%) en el plásmido E y 35 (9.09%) en el plásmido F. Lo que estos datos nos indican es que en vida libre la diferencia del porcentaje de los genes ubicados en el cromosoma que son diferencialmente expresados con respecto a simbiosis es de más del 20% (52% contra el 77%); por el contrario, los genes que pertenecen al plásmido D y que se expresan en vida libre sólo son 10, lo que representa el 2.6% del total, en tanto en simbiosis se observaron 91 genes expresados que pertenecen a este replicón (13.21%), lo que nos sugiere que nuestra estrategia de análisis es la indicada (Fig. 4).

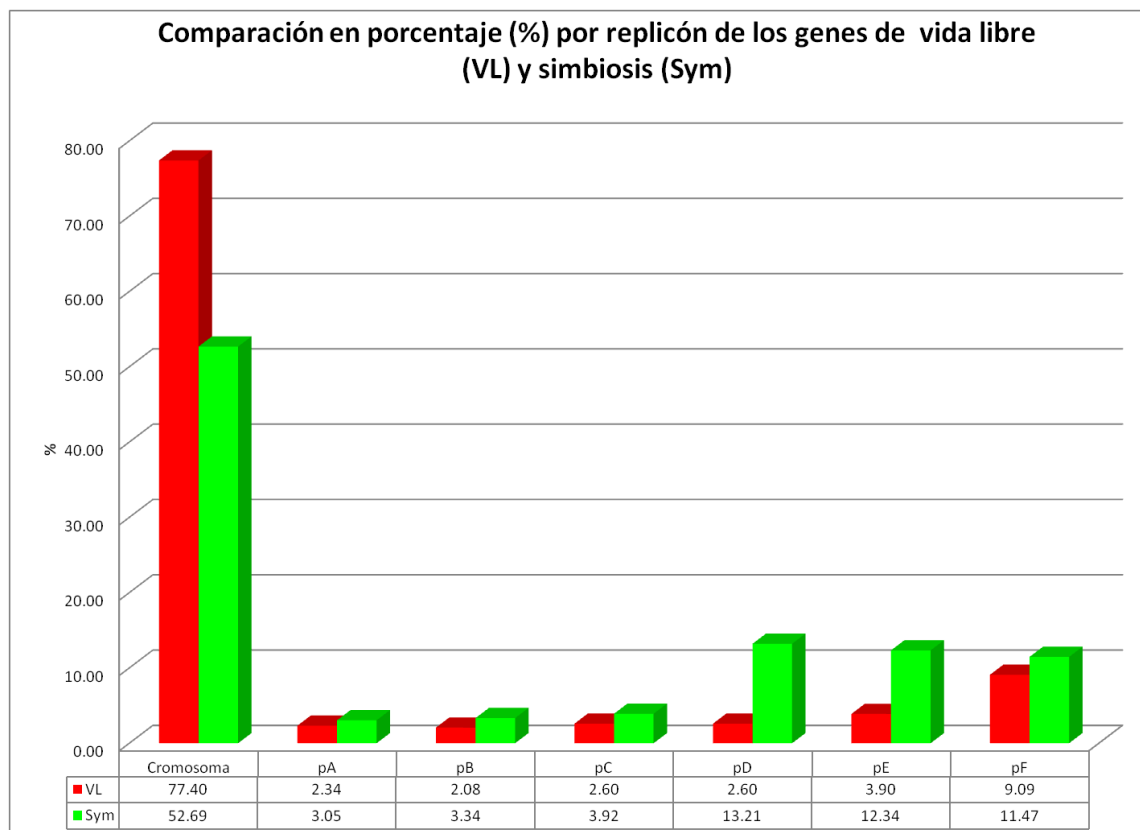


Figura 4. Comparación en porcentaje de los genes expresados en cada replicón del genoma de *R. etli*, en vida libre (VL, barras rojas) y los genes expresados en simbiosis (Sym, barras verdes). Los porcentajes se obtuvieron tomando como el 100% el total de cada condición (385 genes VL y 689 genes Sym).

En cuanto a los grupos funcionales (COGS), los que más contrastan en el número de elementos que los constituyen en la vida libre (6hrs MM) y simbiosis (18ddi) son: Modificación postraduccional, recambio de proteínas y chaperonas, en vida libre hay 22 elementos de este grupo (5.71%) en tanto en simbiosis se observan sobre-expresados solo 8 elementos de este COG, lo que corresponde al 1.16%. Traducción es el COG con el porcentaje de elementos con más diferencia entre estas dos condiciones, en donde 40 elementos son de vida libre (10.39%) y 10 que corresponden a la simbiosis (1.45%). El COG de control del ciclo celular, mitosis y meiosis contiene 8 elementos en vida libre (2.08%) por sólo 2 (0.29%) en simbiosis. Por último, el COG que corresponde a tráfico y secreción intracelular está constituido por 7 elementos en vida libre y sólo por 3 de simbiosis, el 1.82% y el 0.44% respectivamente. De los genes expresados en simbiosis el grupo que destaca son los genes que no se encuentran clasificados en COGS (No en

COGS), con 16 elementos expresados en la vida libre por 134 en simbiosis, que representan el 4.16% y el 19.45% respectivamente. En vida libre el COG-Q (transporte, biosíntesis y catabolismo de metabolitos secundarios) está constituido por 4 elementos y por 21 en la simbiosis, el 1.04% y el 3.05% respectivamente. Adicionalmente, el COG de motilidad celular tiene un mayor número de elementos en simbiosis con 13 (el 1.89%) mientras que sólo 2 (0.52%) están presentes en vida libre. Con este análisis no se observan diferencia en el número de elementos que constituyen el COG-C (producción y conversión de energía), porque el grupo incluye genes a los genes que participan en la fijación de nitrógeno (genes *nif*) expresados en simbiosis, y en la vida libre se expresan otros como los genes *atp*. En la tabla 2 y en la figuras 5 y 6 se presentan los COG's arriba descritos, así como los números y porcentajes de todos los COG's.

Tabla 2. Números totales de elementos en los COGS en vida libre (VL) y simbiosis (Sym), así como en porcentaje y relación entre ambas condiciones utilizando logaritmo base 2 (columna 6).

COG	VL	Sym	VL(%)	Sym(%)	Rel VL-Sym (Log2)
Amino acid transport and metabolism	35	57	9.09	8.27	0.14
Carbohydrate transport and metabolism	32	76	8.31	11.03	-0.41
Cell cycle control, mitosis and meiosis	8	2	2.08	0.29	2.84
Cell motility	2	13	0.52	1.89	-1.86
Cell wall/membrane biogenesis	23	26	5.97	3.77	0.66
Chromatin structure and dynamics	1	0	0.26	0.00	4.16
Coenzyme transport and metabolism	12	9	3.12	1.31	1.25
Defense mechanisms	6	5	1.56	0.73	1.10
Energy production and conversion	29	49	7.53	7.11	0.08
Function unknown	37	43	9.61	6.24	0.62
General function prediction only	29	47	7.53	6.82	0.14
Inorganic ion transport and metabolism	20	26	5.19	3.77	0.46
Intracellular trafficking and secretion	7	3	1.82	0.44	2.06
Lipid transport and metabolism	13	32	3.38	4.64	-0.46
Not in COGs	16	134	4.16	19.45	-2.23
Nucleotide transport and metabolism	3	9	0.78	1.31	-0.75
Posttranslational modification, protein turnover, chaperones	22	8	5.71	1.16	2.30
Replication, recombination and repair	11	30	2.86	4.35	-0.61
Secondary metabolites biosynthesis, transport and catabolism	4	21	1.04	3.05	-1.55
Signal transduction mechanisms	13	31	3.38	4.50	-0.41
Transcription	22	58	5.71	8.42	-0.56
Translation	40	10	10.39	1.45	2.84
Totales	385	689	100	100	---

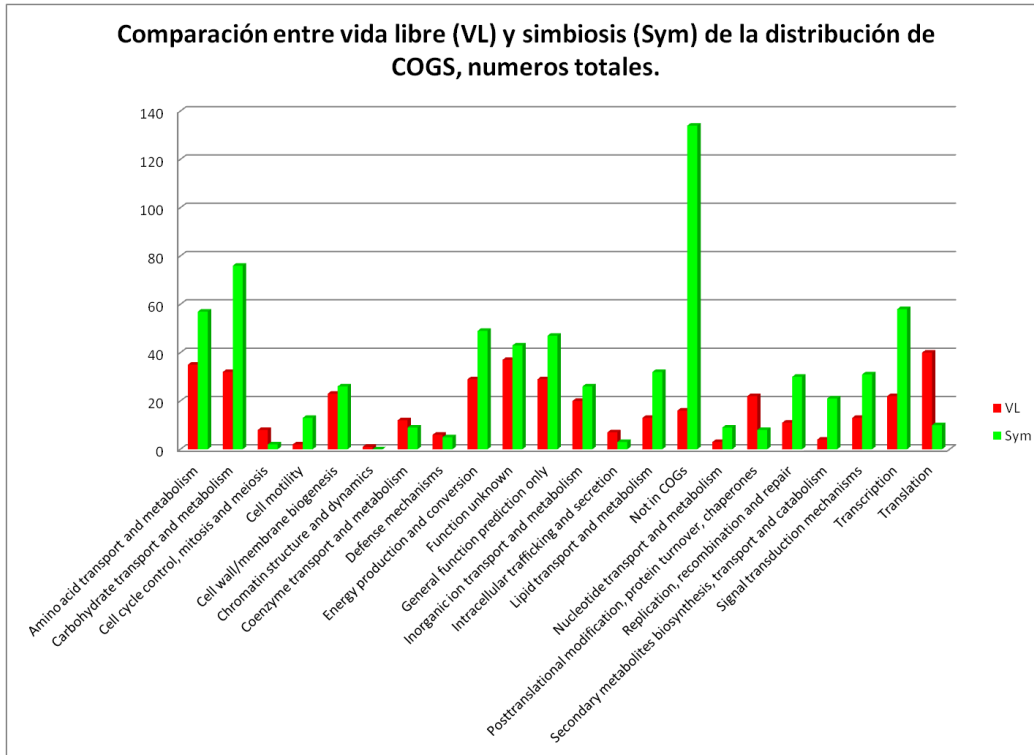


Figura 5. Números totales de la distribución de elementos en COGS, vida libre (VL), barras rojas y en simbiosis (Sym), barras verdes.

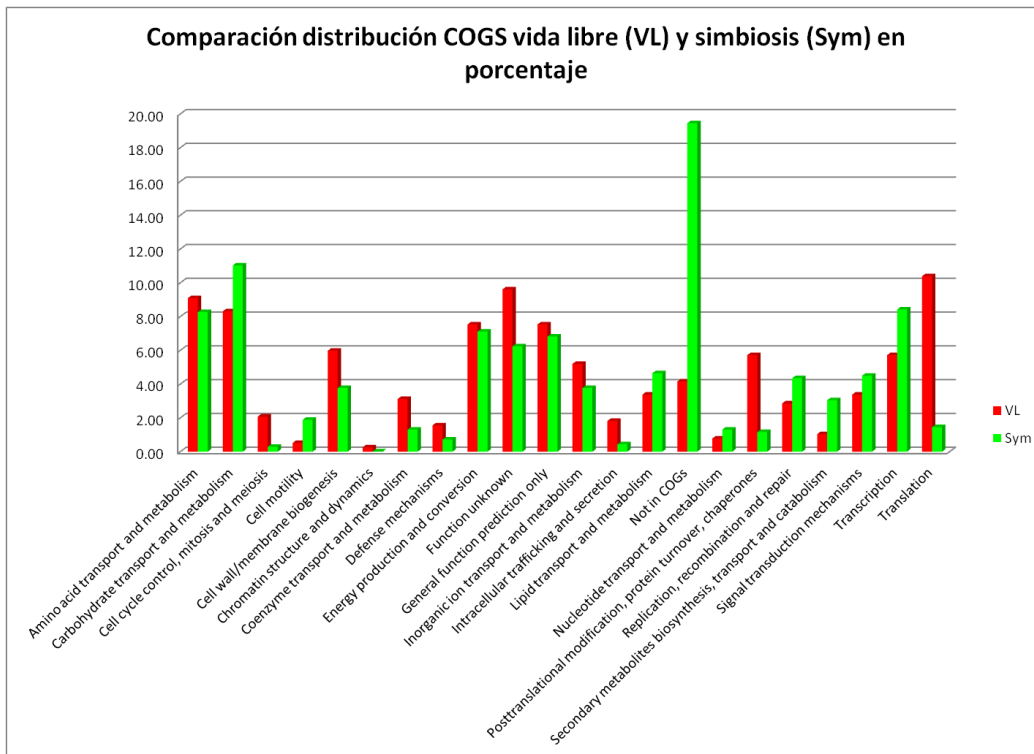


Figura 6. Porcentajes de elementos distribuidos en COGS, en vida libre (VL), barras rojas y en simbiosis (Sym), barras verdes.

A excepción del COG D (Motilidad celular), el resto de los COG's con mayor contraste en el número de elementos y porcentaje de los mismos entre la vida libre y la simbiosis, coincide con los transcriptomas previos reportados en *B. japonicum* (Lindemann *et al.*, 2007), *R. leguminosarum* (Karunakaran *et al.*, 2009), y *S. meliloti* (Cabanés *et al.*, 2006). Con la finalidad de realizar una búsqueda más exhaustiva y que a su vez nos permitiera obtener información global y adicional a la de un análisis de transcriptoma convencional, utilizamos el programa GSEA (Subramanian *et al.*, 2005). Este programa nos permite, a partir de resultados de expresión y con los genes ordenados de mayor a menor o viceversa, extraer información de enriquecimiento de alguna vía en particular, es decir, con base a una plantilla en donde se agrupan a los genes por vía, obteniéndose aquellos que presentan una expresión similar y pertenecen a una misma vía. Para utilizar este programa, no se requiere que los datos tengan un corte estadístico, ya que los datos se adicionan ordenados por expresión, y el programa le da un mayor peso al resultado biológico, el resultado final contiene valores estadísticos que ayudan a la interpretación y selección de los mismos (Subramanian *et al.*, 2005).

Los transcriptomas de un organismo en diferentes condiciones por lo general mantienen un número mayoritario de genes que se expresan en común; a diferencia de lo anterior, en nuestro modelo de estudio, y en particular cuando comparamos la expresión de células de *R. etli* en vida libre (6h MM) contra la expresión en simbiosis (18 ddi), la expresión diferencial observada es de las más contrastantes. Esto dificulta establecer un corte estadístico excluya genes relevantes, en el artículo No. 2 el z-score fue ± 1.65 (Resendis-Antonio, 2011), por lo que se requirió hacer un análisis que mostrara en su totalidad las diferencias de expresión entre ambas condiciones, por ello utilizamos los 6034 genes anotados para *R. etli* y los ordenamos por su valor de z-score de mayor a menor utilizando en conjunto el programa GSEA.

La base de datos para *R. etli* en KEGG, contiene 115 vías compuestas desde 5 hasta 323 elementos. Estas 115 vías se utilizaron como plantilla para obtener los grupos enriquecidos de cada condición, con los siguientes resultados: el grupo con un mayor enriquecimiento lo observamos en vida libre, a este pertenecen los genes que codifican

para proteínas ribosomales, de 55 genes de este grupo presentes en el genoma de *R. etli* CFN42 48 se enriquecieron en la condición de medio mínimo a las 6 hrs (VL). En la simbiosis, uno de los grupos enriquecidos es el metabolismo de nitrógeno que involucra a varios de los genes de la fijación de nitrógeno. El resto de los grupos que se enriquecieron para simbiosis y vida libre se presentan ordenados por sus valores de enriquecimiento en las tablas 3 y 4.

Tabla 3. Vías enriquecidas en la simbiosis utilizando el programa GSEA.

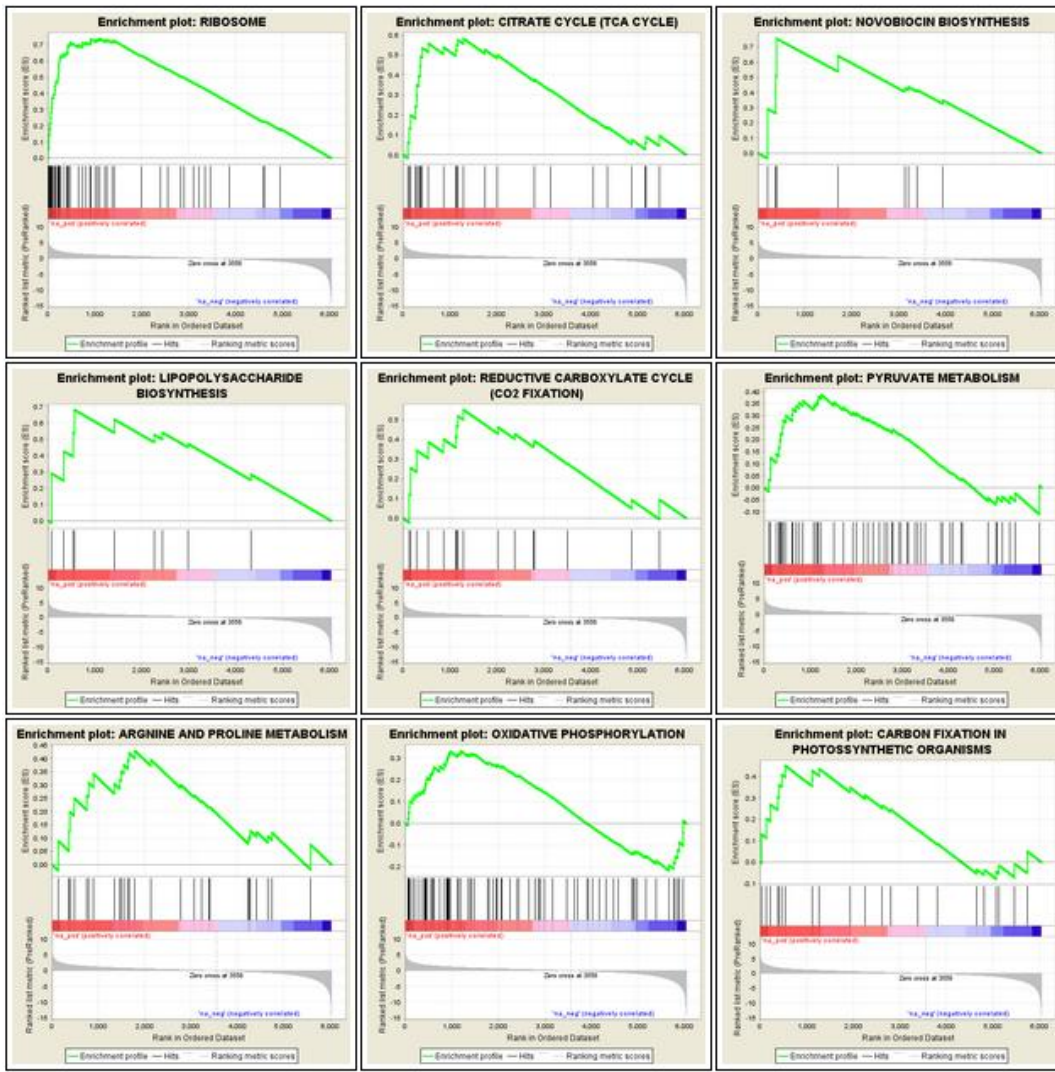
NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
TETRACHLOROETHANE DEGRADATION	20	-0.8031195	-2.080933	0	0	0	461
1,4-DICHLOROBENZENE DEGRADATION	10	-0.83729213	-1.854909	0	0.018115647	0.039	794
NITROGEN METABOLISM	45	-0.5334849	-1.5969121	0.011940299	0.42215335	0.737	470
TOLUENE AND XYLENE DEGRADATION	9	-0.7277483	-1.5769241	0.035250463	0.3775375	0.8	794
BISPHENOL A DEGRADATION	21	-0.59844905	-1.5593297	0.03642384	0.3500766	0.855	1079
STARCH AND SUCROSE METABOLISM	35	-0.5333244	-1.5524485	0.02366864	0.31120047	0.874	563
UBIQUINONE AND MENAQUINONE BIOSYNTHESIS	8	-0.73340666	-1.55051	0.047186933	0.2712068	0.878	15
GAMMA-HEXACHLOROCYCLOHEXANE DEGRADATION	18	-0.60600567	-1.5108784	0.04393673	0.33224687	0.947	794
NAPHTHALENE AND ANTHRACENE DEGRADATION	18	-0.5779138	-1.4993365	0.044982698	0.32073358	0.955	794
1- AND 2-METHYLNAPHTHALENE DEGRADATION	32	-0.53657466	-1.4965844	0.03937008	0.29355392	0.958	825
BILE ACID BIOSYNTHESIS	41	-0.4989194	-1.489862	0.030927835	0.2827193	0.965	1153
FATTY ACID BIOSYNTHESIS	27	-0.52039164	-1.4243014	0.07153729	0.407697	0.994	492
DRUG METABOLISM - OTHER ENZYMES	14	-0.5871684	-1.4048282	0.09171076	0.42445695	0.996	666
FLAGELLAR ASSEMBLY	35	-0.45457843	-1.3208255	0.12006079	0.6341267	1	729
LYSINE BIOSYNTHESIS	25	-0.4951895	-1.3151962	0.14397496	0.60909146	1	741
ANDROGEN AND ESTROGEN METABOLISM	11	-0.57731664	-1.3000101	0.17169811	0.6158737	1	794
FOLATE METABOLISM	12	-0.54748577	-1.2655607	0.21005386	0.68556213	1	363
METHANE METABOLISM	18	-0.5024282	-1.2648497	0.19861831	0.6494248	1	810
ABC TRANSPORTERS	321	-0.3275306	-1.249427	0.06364617	0.6619842	1	866
STYRENE DEGRADATION	6	-0.6281749	-1.2441754	0.23917526	0.6441404	1	219
LIMONENE AND PINENE DEGRADATION	32	-0.43972996	-1.2422354	0.18238021	0.61913323	1	989
GERANIOL DEGRADATION	14	-0.5295156	-1.240223	0.20659722	0.5965208	1	831
VITAMIN B6 METABOLISM	6	-0.5864421	-1.1687661	0.31106472	0.7651475	1	1450
METABOLISM OF XENOBIOTICS BY CYTOCHROME P450	29	-0.41422892	-1.160724	0.26881722	0.7557623	1	825
FRUCTOSE AND MANNOSE METABOLISM	41	-0.38003865	-1.1413983	0.27878788	0.7833825	1	461
RIBOFLAVIN METABOLISM	12	-0.49545217	-1.1260152	0.32310468	0.7968303	1	894
NUCLEOTIDE EXCISION REPAIR	12	-0.4794516	-1.1141729	0.34859154	0.8006188	1	454
DRUG METABOLISM - CYTOCHROME P450	30	-0.39166218	-1.1057098	0.32854864	0.7943146	1	825
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	46	-0.35381234	-1.07973	0.34401113	0.83849335	1	1118
3-CLOROACRYLIC ACID DEGRADATION	24	-0.40019196	-1.0659258	0.3911672	0.8464	1	1234
TYROSINE METABOLISM	59	-0.32306537	-1.0040994	0.4525862	0.98457396	1	825
FATTY ACID METABOLISM	47	-0.323822	-1.0005534	0.4647676	0.96312004	1	831
AMINOSUGARS METABOLISM	20	-0.3857358	-0.99932396	0.4366883	0.93701476	1	997
PROPANOATE METABOLISM	44	-0.31891719	-0.96291184	0.50812405	1	1	371
OTHER GLYCAN DEGRADATION	11	-0.42253494	-0.9562939	0.52690166	0.99388367	1	973
ALKALOID BIOSYNTHESIS	10	-0.4329342	-0.94797957	0.5528169	0.9879427	1	826
ASCORBATE AND ALDARATE METABOLISM	18	-0.37354124	-0.9258594	0.5422078	1	1	264
CYANOAMINO ACID METABOLISM	9	-0.43461937	-0.9229667	0.5793499	0.995161	1	219
BUTANOATE METABOLISM	57	-0.29214633	-0.916634	0.5988539	0.98310447	1	461
GLUTAMATE METABOLISM	38	-0.31209156	-0.90735435	0.5962099	0.979508	1	722
GLYCOSAMINOGLYCAN DEGRADATION	6	-0.468536	-0.90728587	0.5690608	0.95580286	1	914
GLYCOSPHINGOLIPID BIOSYNTHESIS-GANGLIOSERIES	6	-0.468536	-0.89337	0.6137255	0.9644078	1	914
GALACTOSE METABOLISM	27	-0.32103926	-0.88497186	0.61676645	0.9599848	1	166
GLYOXYLATE AND DICARBOXYLATE METABOLISM	30	-0.30569443	-0.86890906	0.6486486	0.9734536	1	943
TYPE III SECRETION SYSTEM	16	-0.34930053	-0.85963774	0.64089346	0.972111	1	961
TYPE II SECRETION SYSTEM	16	-0.34930053	-0.85357374	0.6551155	0.9629879	1	961
ONE CARBON POOL BY FOLATE	17	-0.3493803	-0.8504096	0.6738769	0.9494058	1	1247
GLYCOSPHINGOLIPID BIOSYNTHESIS-GLOBOSERIES	6	-0.4359655	-0.8504079	0.65363127	0.9296264	1	3407
FLUORENE DEGRADATION	6	-0.44575316	-0.85018533	0.6576402	0.9110368	1	3348
FLUOROBENZOATE DEGRADATION	6	-0.44575316	-0.8478095	0.67054266	0.89731723	1	3348
CAPROLACTAM DEGRADATION	15	-0.34744483	-0.8402696	0.68350166	0.89361304	1	831
HISTIDINE METABOLISM	40	-0.28689796	-0.8393847	0.6983895	0.87797797	1	989
PANTOTHENATE AND COA BIOSYNTHESIS	22	-0.3226508	-0.83374846	0.69327074	0.8709437	1	341
TRYPTOPHAN METABOLISM	30	-0.30164918	-0.82902145	0.6964561	0.86258006	1	989
NON-HOMOLOGOUS END-JOINING	5	-0.44399175	-0.82561815	0.6802218	0.8526224	1	1143
BACTERIAL CHEMOTAXIS	48	-0.25436762	-0.7812901	0.80938834	0.90709907	1	672
ATRAZINE DEGRADATION	7	-0.3788497	-0.7548386	0.7781885	0.93069506	1	442
PENTOSE PHOSPHATE PATHWAY	32	-0.26441684	-0.7535908	0.8134796	0.91641665	1	621
AMINOACYL-TRNA BIOSYNTHESIS	29	-0.27412793	-0.753548	0.81447965	0.90093005	1	722
BETA-ALANINE METABOLISM	25	-0.2678966	-0.72533166	0.85	0.9231411	1	690

Tabla 4. Vías enriquecidas en la vida libre (VL), utilizando el programa GSEA.

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
RIBOSOME	55	0.7373833	2.7180972	0	0	0	1095
CITRATE CYCLE (TCA CYCLE)	26	0.5816109	1.8074149	0.005420054	0.054760806	0.094	1290
NOVOBIOCIN BIOSYNTHESIS	8	0.75435656	1.6566304	0.008474576	0.18189207	0.384	396
LIPOPOLYSACCHARIDE BIOSYNTHESIS	9	0.68011403	1.5385605	0.037199125	0.3817155	0.758	562
2,4-DICHLOROBENZOATE DEGRADATION	6	0.750578	1.5283483	0.037113402	0.3262661	0.78	758
REDUCTIVE CARBOXYLATE CYCLE (CO2 FIXATION)	16	0.55143857	1.4863954	0.051886793	0.37446165	0.877	1290
PYRUVATE METABOLISM	54	0.3866519	1.4122666	0.03322259	0.51469713	0.965	1245
ARGININE AND PROLINE METABOLISM	28	0.42893183	1.3445666	0.09090909	0.6683948	0.992	1793
OXIDATIVE PHOSPHORYLATION	79	0.33223552	1.3319175	0.042145595	0.6315477	0.993	1195
CARBON FIXATION IN PHOTOSYNTHETIC ORGANISMS	22	0.4505697	1.3093505	0.09703504	0.63664246	0.998	545
PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS	32	0.3909712	1.2939966	0.1017964	0.6240281	0.998	1112
BIOSYNTHESIS OF UNSATURATED FATTY ACIDS	11	0.5281451	1.2823544	0.18396227	0.6107024	0.999	156
GLUTATHIONE METABOLISM	30	0.39610562	1.2702525	0.12320917	0.59912527	0.999	1731
NICOTINATE AND NICOTIAMIDE METABOLISM	18	0.4456458	1.2560974	0.15641026	0.5945713	1	2562
SYNTHESIS AND DEGRADATION OF KETONE BODIES	5	0.6389406	1.2423265	0.20253165	0.59258515	1	1526
TYPE IV SECRETION SYSTEM	33	0.38161507	1.2403693	0.16081871	0.5605002	1	1428
C5-BRANCHED DIBASIC ACID METABOLISM	6	0.6043682	1.2293692	0.2430279	0.5565726	1	1935
GLYCOLISIS/GLUCONEOGENESIS	53	0.32816643	1.1891112	0.18333334	0.62841654	1	474
BENZOATE DEGRADATION VIA HYDROXYLATION	24	0.37864706	1.1496865	0.25319692	0.7008016	1	1281
AMINOPHOSPHONATE METABOLISM	13	0.4373464	1.1467425	0.30622008	0.6742179	1	2497
PEPTIDOGLYCAN BIOSYNTHESIS	19	0.39641324	1.1467282	0.2809524	0.64211226	1	700
PROTEIN EXPORT	16	0.40665257	1.0962995	0.3187348	0.74052364	1	728
TWO-COMPONENT SYSTEM	87	0.2656305	1.077775	0.27419356	0.75923085	1	503
PURINE METABOLISM	93	0.25926203	1.059452	0.31924883	0.7728404	1	1676
ASPARTATE AND ALANINE METABOLISM	25	0.34601223	1.0485196	0.37931034	0.76891685	1	1872
SELENOAMINO ACID METABOLISM	19	0.37315688	1.0448605	0.39948454	0.7476728	1	277
NUCLEOTIDE SUGARS METABOLISM	25	0.3336088	1.0285635	0.41558442	0.7637218	1	214
THIAMINE METABOLISM	17	0.37190387	1.0259641	0.41457286	0.7427672	1	904
PORPHYRIN AND CHLOROPHYLL METABOLISM	37	0.2980681	1.0163352	0.4115942	0.74120665	1	1228
BENZOATE DEGRADATION VIA COA LIGATION	28	0.31405127	1.0150857	0.431694	0.7193581	1	1149
GLYCEROPHOSPHOLIPID METABOLISM	26	0.32172042	1.0051224	0.41643837	0.71802706	1	1292
METHIONINE METABOLISM	16	0.35837722	0.98569816	0.4556962	0.73872083	1	868
PENTOSE AND GLUCORONATE INTERCONVERSIONS	24	0.3240395	0.9589139	0.5307692	0.7748427	1	488
CYSTEINE METABOLISM	11	0.40210015	0.9504742	0.5189573	0.7709681	1	868
PYRIMIDINE METABOLISM	54	0.25428367	0.93072104	0.58518517	0.79050124	1	1535
SULFUR METABOLISM	11	0.36921978	0.9033856	0.5695067	0.82676506	1	791
VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS	26	0.2837566	0.8754076	0.63852245	0.86132616	1	1185
LYSINE DEGRADATION	27	0.26888123	0.84976196	0.71181554	0.89081335	1	814
BASE EXCISION REPAIR	18	0.2860513	0.7946183	0.77017117	0.97088945	1	905
BIOSYNTHESIS OF STEROIDS	9	0.34817168	0.7847947	0.7318182	0.96390104	1	1874
PHENYLALANINE METABOLISM	34	0.23643818	0.7755914	0.8397626	0.9544672	1	1048
DNA REPLICATION	18	0.2720168	0.7497948	0.8193069	0.9679183	1	423
ALKALOID BIOSYNTHESIS II	5	0.38932535	0.73793143	0.7945493	0.96161914	1	479
UREA CYCLE AND METABOLISM OF AMINO GROUPS	34	0.19075058	0.6393997	0.97626114	1	1	1485
ETHYLBENZENE DEGRADATION	5	0.29663822	0.58079326	0.9154639	1	1	912
GLYCEROLIPID METABOLISM	23	0.1895742	0.5738342	0.97002727	1	1	474
D-ALANINE METABOLISM	9	0.22989249	0.5196803	0.9653061	1	1	1944
MISMATCH REPAIR	24	0.15501557	0.4689435	0.99497485	1	1	1486
STREPTOMICIN BIOSYNTHESIS	15	0.1750314	0.4673556	1	0.99365413	1	1174

La interpretación de los datos obtenidos por GSEA inician con el “Enrichment Score” (ES), el cual básicamente representa el grado en el cual un grupo de genes de cierta clase funcional este sobrerrepresentado en el tope o en el final de la lista previamente ordenada. Entre más alejado de cero, el grupo está mayormente enriquecido; el máximo valor es 1. El ambiente de GSEA también es gráfico, en la figura 7 se presentan los grupos mayormente enriquecidos en vida libre.

Table: Snapshot of enrichment results



Continúa siguiente página

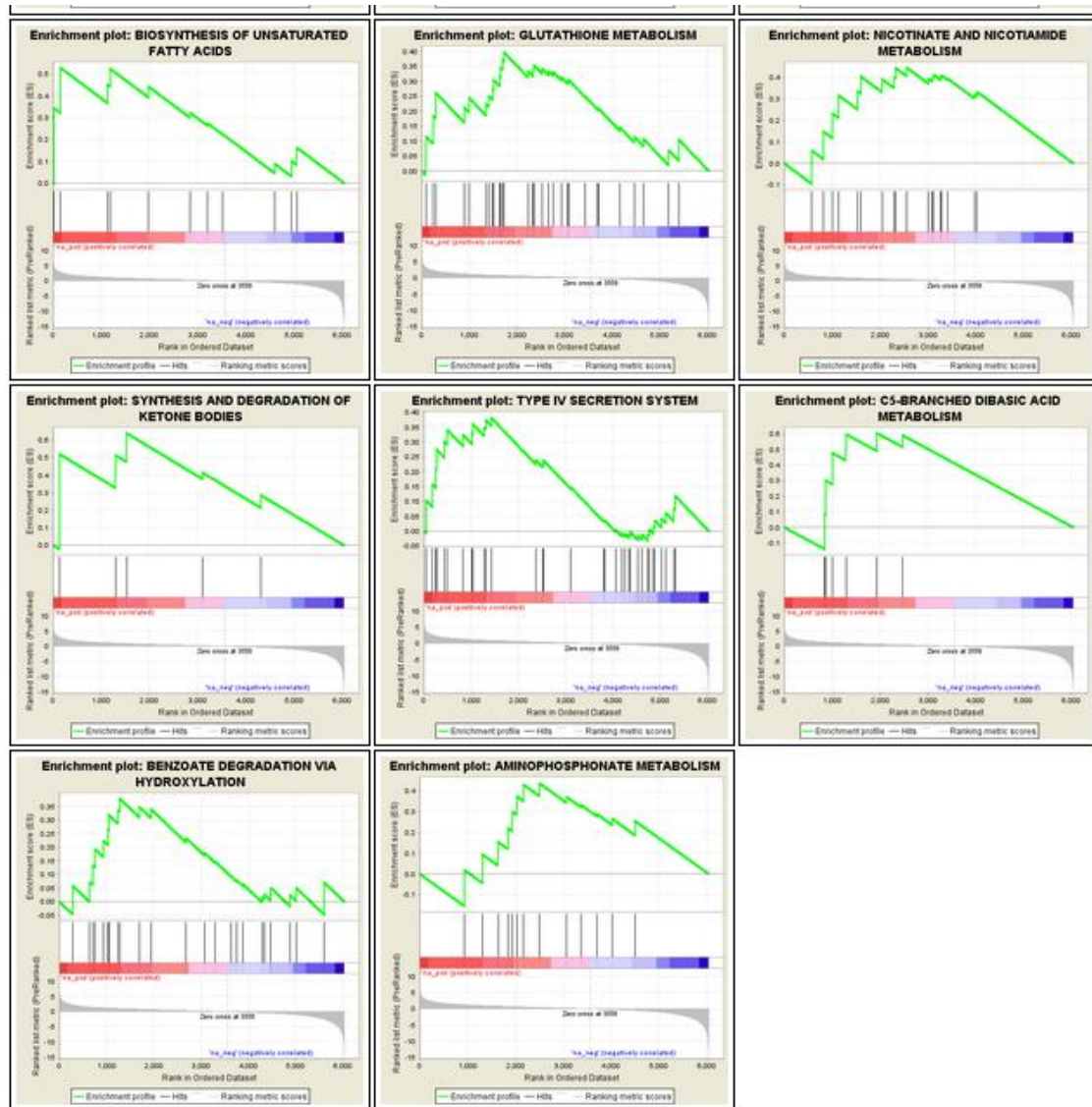


Figura 7. Representación gráfica de los grupos enriquecidos en vida libre (VL). La parte roja de la barra horizontal representa la expresión y la parte azul de la barra representa la represión, respectivamente (genes ordenados por nivel de expresión). Las líneas verticales sobre la barra representan la distribución de los genes de una misma vía.

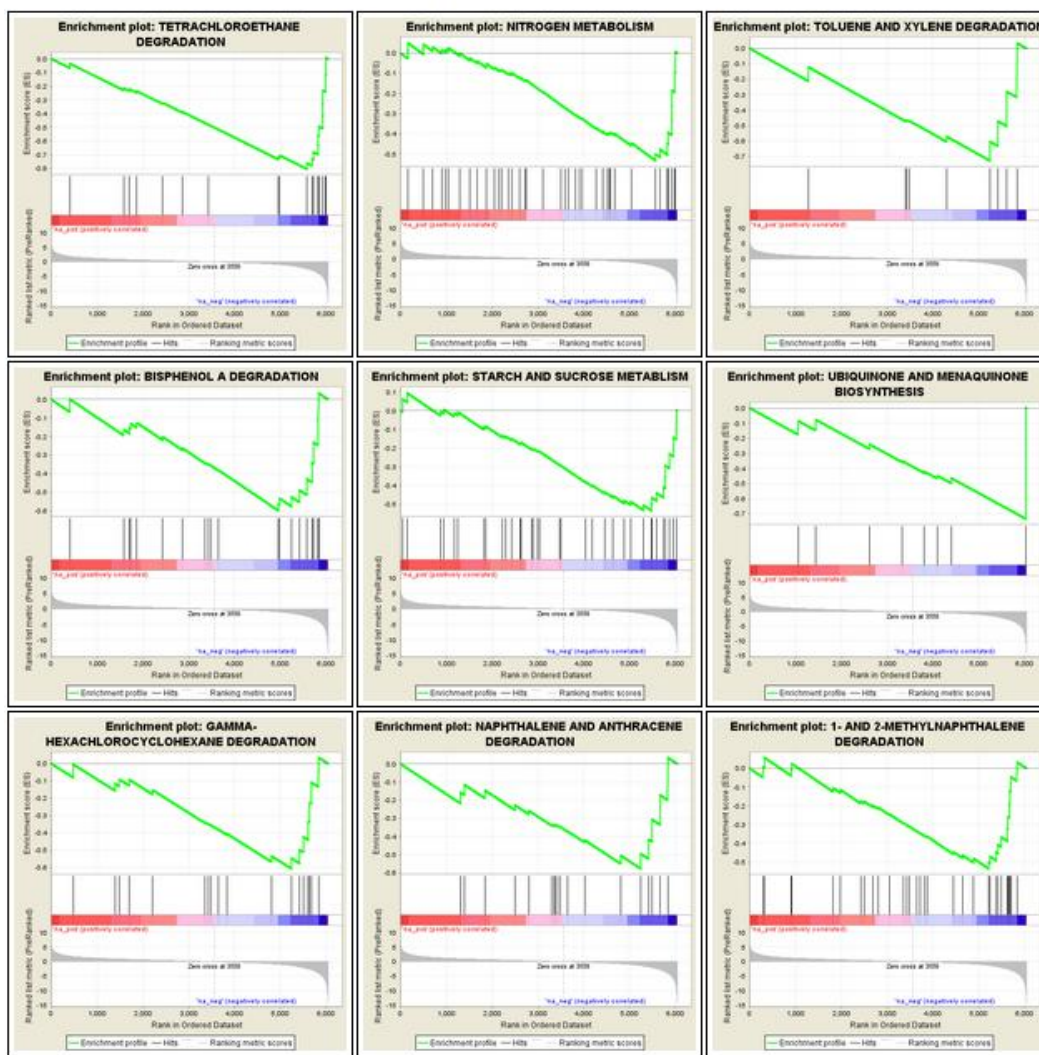
Otro parámetro que utiliza GSEA para otorgar peso al resultado es el valor de NES (Normalized Enrichment Score); el cual es el primer filtro de los resultados del enriquecimiento de las series de genes. Para esta normalización GSEA pondera las diferencias entre el tamaño de la series en correlación con las series de genes expresados. NES está basado en los “scores” de las series enriquecidas para todas las permutaciones de los datos; esto es, que al cambiar el método de permutación, el número de las

permutaciones o el tamaño de la serie de datos afectan el valor de NES. La fórmula que utiliza GSEA para obtener el valor de NES, es la siguiente:

$$NES = \frac{\text{actual ES}}{\text{mean(ESs against all permutations of the dataset)}}$$

El tercer valor a considerar es el FDR (False Discovery Rate), es la probabilidad estimada de que un NES dado represente un falso positivo. Entre más cercano sea este valor a cero el dato tiene mayor certeza; siendo el máximo valor de 1. Los grupos que consideramos enriquecidos en la fijación de nitrógeno a los 18 ddi se presentan de manera gráfica en la figura 8.

Table: Snapshot of enrichment results



Continúa siguiente página

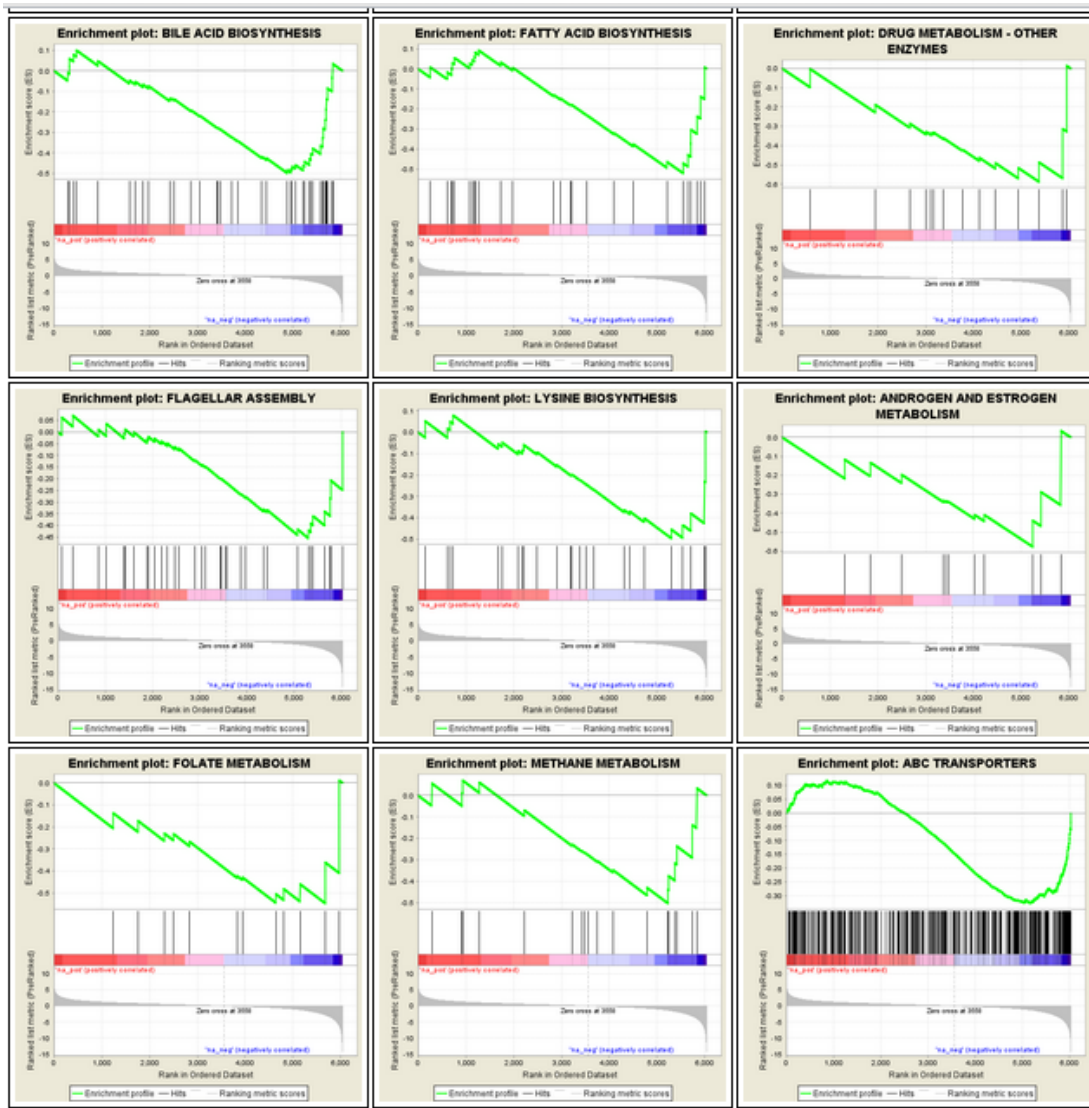


Figura 8. Representación gráfica de los grupos enriquecidos a los 18 ddi. La parte roja de la barra horizontal representa la expresión y la parte azul representa la represión, respectivamente (genes ordenados por nivel de expresión). Las líneas verticales sobre la barra de calor representan la distribución de los genes de una misma vía.

Los resultados de este análisis validaron los obtenidos en el análisis de COG's, además se rescataron grupos como los de metabolismo de Nitrógeno que está enriquecido significativamente en la condición de simbiosis. Un dato que llama la atención es que nuevamente el COG D-Motilidad celular, (en GSEA el grupo rotulado como Ensamble de Flagelo) también se enriquece en la condición de simbiosis. Un antecedente fue reportado en *R. etli* utilizando nódulos de 21 ddi ya que se detectó la expresión del gen *flaF* que

La distribución de los genes sobreexpresados en ambas condiciones y ubicados en vías metabólicas pueden observarse en la tabla 5 y Fig. 10.

Tabla 5. Datos totales de los genes ubicados con el programa “KEGG Pathway Color”, vida libre números totales (VLNum) y porcentaje (VL(%)), simbiosis en números totales (SymNum) y porcentaje (Sym(%)). La última columna es la relación entre VL y Sym en Log2, utilizando los porcentajes para obtener la relación.

Grupo KEGG Pathway	VLNum	SymNum	VL(%)	Sym(%)	VL/Sym(Log2)
ABC transporters	36	81	5.22	9.68	-0.9
Alanine, aspartate and glutamate metabolism	9	6	1.31	0.72	0.9
Aminoacyl tRNA biosynthesis	6	4	0.87	0.48	0.9
Aminobenzoate degradation	2	10	0.29	1.19	-2
Amino sugar and nucleotide sugar metabolism	14	4	2.03	0.48	2.1
Arginine and proline metabolism	21	12	3.05	1.43	1.1
Ascorbate and aldarate metabolism	8	9	1.16	1.08	0.1
Atrazine degradation	0	2	0	0.24	-7.9
Bacterial chemotaxis	0	29	0	3.46	-11.8
Bacterial secretion system	14	2	2.03	0.24	3.1
Base excision repair	2	4	0.29	0.48	-0.7
Benzoate degradation	6	3	0.87	0.36	1.3
betaAlanine metabolism	7	14	1.02	1.67	-0.7
Biosynthesis of ansamycins	3	0	0.44	0	8.8
Biosynthesis of secondary metabolites	68	72	9.87	8.6	0.2
Biosynthesis of unsaturated fatty acids	2	1	0.29	0.12	1.3
Bisphenol degradation	10	12	1.45	1.43	0
Butanoate metabolism	16	18	2.32	2.15	0.1
C5Branched dibasic acid metabolism	1	1	0.15	0.12	0.3
Caprolactam degradation	0	4	0	0.48	-8.9
Carotenoid biosynthesis	0	1	0	0.12	-6.9
Chloroalkane and chloroalkene degradation	17	32	2.47	3.82	-0.6
Chlorocyclohexane and chlorobenzene degradation	1	8	0.15	0.96	-2.7
Citrate cycle	13	0	1.89	0	10.9
Cyanoamino acid metabolism	0	2	0	0.24	-7.9
Cysteine and methionine metabolism	4	2	0.58	0.24	1.3
DAlanine metabolism	2	0	0.29	0	8.2
DGlutamine and Dglutamate metabolism	0	1	0	0.12	-6.9
Dioxin degradation	1	0	0.15	0	7.2
DNA replication	2	0	0.29	0	8.2
Fatty acid biosynthesis	0	8	0	0.96	-9.9
Fatty acid metabolism	14	24	2.03	2.87	-0.5
Flagellar assembly	6	15	0.87	1.79	-1
Fluorobenzoate degradation	0	4	0	0.48	-8.9
Folate biosynthesis	1	3	0.15	0.36	-1.3
Fructose and mannose metabolism	10	11	1.45	1.31	0.1
Galactose metabolism	11	5	1.6	0.6	1.4
Geraniol degradation	0	4	0	0.48	-8.9
Glutathione metabolism	16	14	2.32	1.67	0.5

Continúa siguiente página

Grupo KEGG Pathway	VNum	SymNum	VL(%)	Sym(%)	VL/Sym(Log2)
Glycerolipid metabolism	8	7	1.16	0.84	0.5
Glycerophospholipid metabolism	9	7	1.31	0.84	0.6
Glycine, serine and threonine metabolism	9	7	1.31	0.84	0.6
Glycolysis / Gluconeogenesis	20	19	2.9	2.27	0.4
Glyoxylate and dicarboxylate metabolism	10	7	1.45	0.84	0.8
Histidine metabolism	13	16	1.89	1.91	0
Homologous recombination	2	2	0.29	0.24	0.3
Inositol phosphate metabolism	0	4	0	0.48	-8.9
Limonene and pinene degradation	7	15	1.02	1.79	-0.8
Lipopolysaccharide biosynthesis	3	0	0.44	0	8.8
Lysine biosynthesis	1	12	0.15	1.43	-3.3
Lysine degradation	12	10	1.74	1.19	0.5
Methane metabolism	8	4	1.16	0.48	1.3
Mismatch repair	2	4	0.29	0.48	-0.7
Naphthalene degradation	2	13	0.29	1.55	-2.4
Nicotinate and nicotinamide metabolism	2	1	0.29	0.12	1.3
Nitrogen metabolism	18	23	2.61	2.75	-0.1
Nitrotoluene degradation	0	2	0	0.24	-7.9
Nonhomologous endjoining	3	4	0.44	0.48	-0.1
Novobiocin biosynthesis	6	0	0.87	0	9.8
Nucleotide excision repair	0	5	0	0.6	-9.2
One carbon pool by folate	0	6	0	0.72	-9.5
Oxidative phosphorylation	20	17	2.9	2.03	0.5
Pantothenate and CoA biosynthesis	0	5	0	0.6	-9.2
Pentose and glucuronate interconversions	13	10	1.89	1.19	0.7
Pentose phosphate pathway	4	5	0.58	0.6	0
Peptidoglycan biosynthesis	2	2	0.29	0.24	0.3
Phenylalanine metabolism	13	3	1.89	0.36	2.4
Phenylalanine, tyrosine and tryptophan biosynthesis	7	2	1.02	0.24	2.1
Phosphotransferase system	2	0	0.29	0	8.2
Polycyclic aromatic hydrocarbon degradation	0	4	0	0.48	-8.9
Porphyrin and chlorophyll metabolism	9	7	1.31	0.84	0.6
Propanoate metabolism	12	16	1.74	1.91	-0.1
Protein export	0	1	0	0.12	-6.9
Purine metabolism	7	23	1.02	2.75	-1.4
Pyrimidine metabolism	4	8	0.58	0.96	-0.7
Pyruvate metabolism	18	8	2.61	0.96	1.5
Riboflavin metabolism	0	4	0	0.48	-8.9
Ribosome	28	0	4.06	0	12
RNA degradation	7	3	1.02	0.36	1.5
RNA polymerase	0	1	0	0.12	-6.9
Selenocompound metabolism	1	3	0.15	0.36	-1.3
Starch and sucrose metabolism	2	13	0.29	1.55	-2.4
Streptomycin biosynthesis	0	4	0	0.48	-8.9
Styrene degradation	0	2	0	0.24	-7.9

Continúa siguiente página

Grupo KEGG Pathway	VLNum	SymNum	VL(%)	Sym(%)	VL/Sym(Log2)
Sulfur metabolism	0	1	0	0.12	-6.9
Sulfur relay system	2	2	0.29	0.24	0.3
Synthesis and degradation of ketone bodies	4	0	0.58	0	9.2
Terpenoid backbone biosynthesis	4	1	0.58	0.12	2.3
Thiamine metabolism	5	2	0.73	0.24	1.6
Toluene degradation	1	6	0.15	0.72	-2.3
Tryptophan metabolism	11	12	1.6	1.43	0.2
Two-component-system	26	42	3.77	5.02	-0.4
Tyrosine metabolism	9	14	1.31	1.67	-0.4
Ubiquinone and other terpenoid, quinone biosynthesis	0	1	0	0.12	-6.9
Valine, leucine and isoleucine biosynthesis	2	5	0.29	0.6	-1
Valine, leucine and isoleucine degradation	18	17	2.61	2.03	0.4
Vitamin B6 metabolism	0	3	0	0.36	-8.5

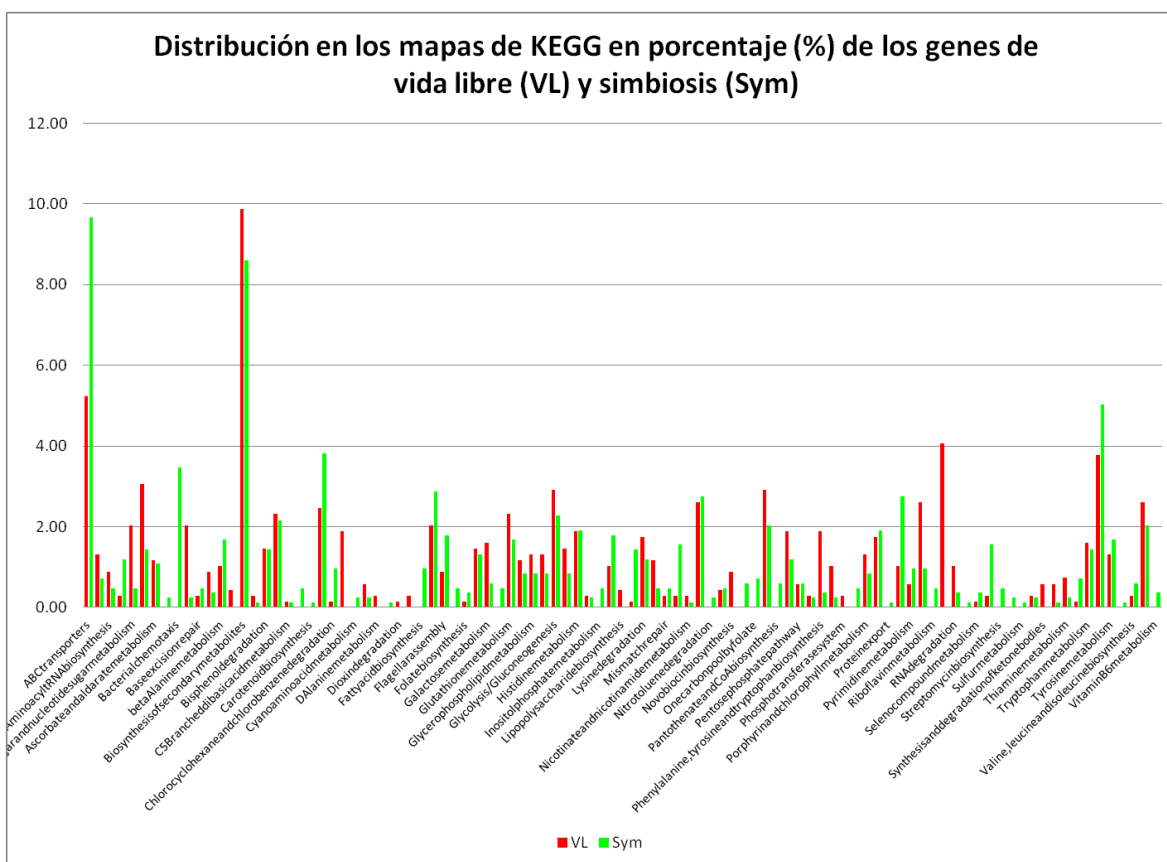


Figura 10. Distribución de los genes sobreexpresados en las vías de KEGG en porcentaje, las barras rojas representan vida libre (VL) y las verdes representan simbiosis 18 ddi (Sym).

Con la finalidad de identificar los mapas que presentan un mayor contraste entre la vida libre (VL) y la simbiosis 18 ddi (Sym), utilizamos la distribución en porcentaje (%) (Fig. 10 y

tabla 5) de los grupos de KEGG, con esto se obtuvo la relación entre VL/Sym (log2) (Fig. 11).

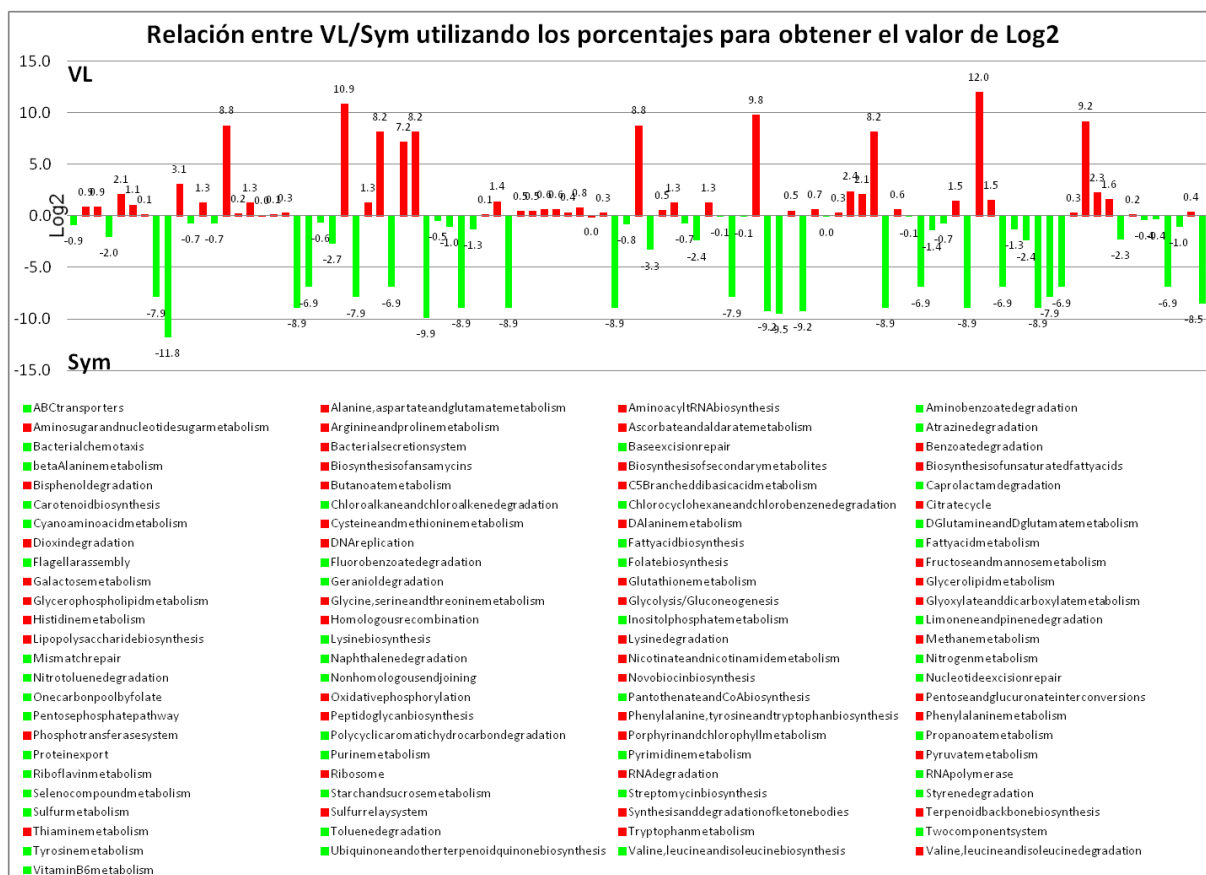


Figura 11. Relación de la distribución en KEGG entre los genes propios o con tendencia o sobreexpresión en la vida libre (Barras rojas (VL)) respecto a los grupos formados por los genes de simbiosis (Barras verdes (Sym)), utilizando los porcentajes para obtener la relación (log2).

Por último, cada elemento fue etiquetado en su respectiva vía, con los siguientes colores; rojo para vida libre, verde para lo propios de simbiosis y purpura para los comunes (dos o más genes que cubren el mismo punto del mapa, donde uno se ubica en el listado de simbiosis y su probable homólogo en vida libre o viceversa).

Los mapas metabólicos más contrastantes en KEGG en lo que respecta a simbiosis corresponden a Quimiotaxis bacterial con 29 elementos en simbiosis por cero en vida libre, la biosíntesis de ácidos grasos con 8 elementos en simbiosis por cero en vida libre (Fig. 12 inciso a), reservorio para un folato de carbono con 6 elementos en simbiosis por

cero en vida libre, la biosíntesis de pantotenato y CoA; escisión y reparación de nucleótidos con 5 elementos cada uno en simbiosis y cero en vida libre. Biosíntesis de lisina con 12 elementos en simbiosis por 1 en vida libre, degradación de clorociclohexano y clorobenceno con 8 elementos en simbiosis por 1 en vida libre, la degradación de naftaleno y el metabolismo de almidón y sacarosa con 13 elementos en simbiosis por 2 en vida libre (Fig. 12 inciso b) (Tabla 5).

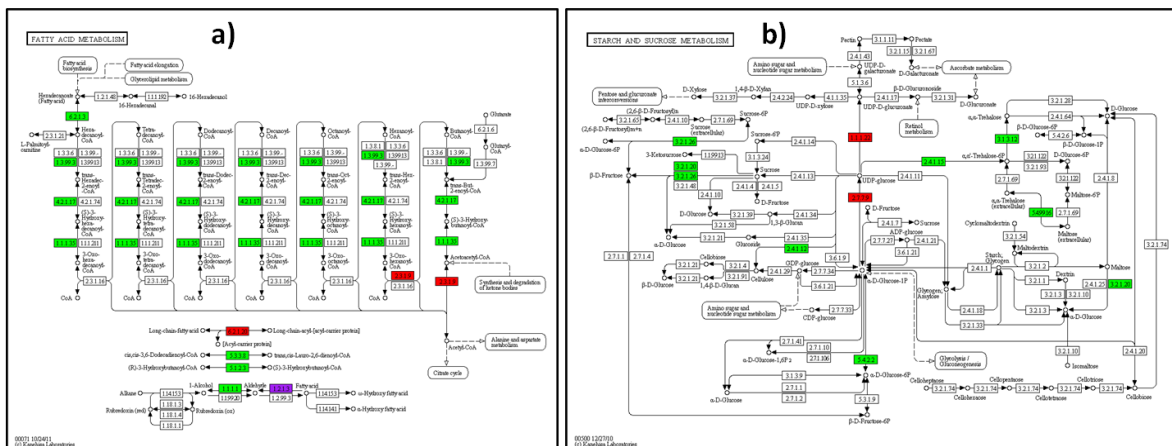


Figura 12. Ejemplos de representación grafica de los genes en los mapas de “KEGG pathway color”, en verde los genes expresados en simbiosis 18 ddi, en purpura genes comunes y en rojo los expresados en vida libre; a) Metabolismo de ácidos grasos, b) Metabolismo de almidón y sacarosa.

Por lo que respecta a vida libre, las vías metabólicas que presentan un mayor número de elementos en comparación con la simbiosis son:

Ribosoma, los genes que codifican para proteínas ribosomales tiene 28 elementos en vida libre por cero en simbiosis (Fig. 13 inciso a); el ciclo del citrato (TCA) con 13 elementos en vida libre y cero en simbiosis (Fig. 13 inciso b); biosíntesis de novobiocina, con 6 en vida libre y cero en simbiosis; biosíntesis y degradación de cuerpos cetónicos y Sistema de secreción bacteriana con 14 de vida libre y 2 en simbiosis; biosíntesis de fenilalanina, tirosina y triptófano con 13 elementos en vida libre por 3 en simbiosis; degradación de RNA con 7 elementos en vida libre por 3 en simbiosis; metabolismo de amino-azúcares y nucleótido azúcares con 14 en vida libre por 4 en simbiosis (Tabla 5).

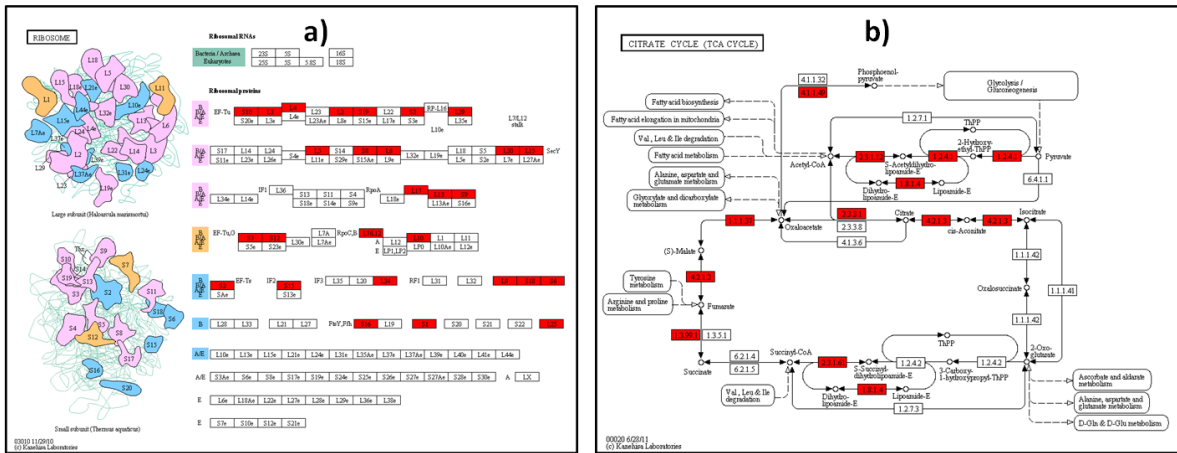


Figura 13. Ejemplos de representación gráfica de los genes en los mapas de “KEGG pathway color”, en verde los genes expresados en simbiosis 18 ddi, en purpura genes comunes y en rojo los expresados en vida libre, a) Ribosoma, b) Ciclo del citrato o TCA.

11.3. Análisis de la expresión global en diferentes etapas de la fijación de nitrógeno y una condición de vida libre

Antes, durante y al final de la interacción *Rhizobium-Phaseolus*, se presentan diferentes tipos de restricciones o barreras, estas pueden ser del tipo físico o bien de disponibilidad de fuentes de nutrientes, así como pH, oxígeno, radicales libres, entre otros; las señales provenientes de la planta impactan sobre grupos de genes de la bacteria que se expresan dependiendo de las etapas en la que se localiza la interacción. Estas se pueden dividir considerando a la actividad de la nitrogenasa (determinada mediante la reducción de acetileno) como el parámetro fundamental para definir estas etapas, en donde se observa que entre los 9 y 11 ddi hay una muy baja fijación de nitrógeno, la cual aumenta a los 18 ddi y alcanza su pico máximo aproximadamente a los 25 ddi, disminuyendo a partir de los 28 ddi y presentando a los 35 ddi niveles muy similares a los observados a los 11 ddi. Con la finalidad de obtener a detalle los elementos genéticos que participan en cada una de estas etapas, realizamos un análisis transcriptómico en cinco puntos de la fijación de nitrógeno. Primero en la etapa inicial (11ddi), posteriormente en una etapa intermedia o de adaptación y preparación (18 ddi), así como en el punto máximo de la fijación de nitrógeno (25 ddi) y un tiempo cercano (28 ddi) con la finalidad de observar genes que señalaran una senescencia o presenescencia y finalmente el punto más tardío en donde

logramos obtener una concentración e integridad adecuada del RNA que fue a los 35 ddi. Por la baja actividad de nitrogenasa detectada consideramos nos sugiere el inicio del proceso de senescencia del nódulo. Finalmente, los datos de simbiosis se analizaron y junto con los datos de vida libre (MM 6h) se compararon para obtener una serie de perfiles que detallaremos en los siguientes párrafos.

El análisis de las diferentes etapas de la fijación de nitrógeno se realizó a partir de los datos recabados en los diferentes tiempos de simbiosis, estos se normalizaron junto con los datos de las hibridaciones de la vida libre (ver la sección de materiales y métodos), los datos de 18 ddi contra vida libre se tomaron de los que se presentan en el artículo de Resendis-Antonio *et al.*, 2011 y los datos de 11 ddi en el artículo de Salazar *et al.*, 2010. También se integraron datos de 25, 28 y 35 ddi. De estas comparaciones se seleccionaron los genes más expresados en cada tiempo, a partir de estos datos se realizó un análisis de agrupamiento o “cluster” tipo “K-means”, y se aplicó un agrupamiento jerárquico. Por otro lado, los promedios de las réplicas fueron utilizados para una comparación individual por GSEA (enriquecimientos de genes por vías metabólicas). A cada gen de interés se le analizó la región promotora para localizar sitios de regulación, se hizo un análisis de la distribución de los genes de cada perfil en los 6 plásmidos y el cromosoma, se analizaron los datos en KEGG para determinar que mapas eran los que presentaba cada perfil y por último, cada perfil se separó por COG’s y se obtuvieron porcentajes de distribución de cada clase funcional. Análisis de este tipo se han realizado en *S. meliloti* (Capela *et al.*, 2006), en *B. japonicum* (Pessi *et al.*, 2007) y en *R. leguminosarum* (2009), pero en ninguno de los tres modelos se han analizado tiempos tan lejanos como a los 35 ddi, el análisis más lejano lo realizaron a los 31 ddi, obteniendo prácticamente los mismos resultados que a los 21 ddi (Pessi *et al.*, 2007). Adicionalmente, es importante mencionar que los nódulos que forma *S. meliloti* con la alfalfa y los de *R. leguminosarum* con *Pisum sativum* son nódulos indeterminados por lo que las poblaciones bacterianas o de bacteroides son heterogéneas, es decir, el mismo nódulo contiene todas las etapas de la fijación, y las principales diferencias las localizan en los tiempos tempranos, 5 y 7 ddi. En tanto que *R. etli* induce nódulos determinados en plantas de frijol (*P. vulgaris*). Por lo anterior en esta

parte de la tesis presentamos un análisis a detalle de los genes que participan en cada etapa de la fijación de nitrógeno entre *R. etli* y *P. vulgaris* incluyendo la etapa senescente del nódulo.

11.3.1. Características generales de cada perfil

Para este análisis se seleccionaron sólo aquellos genes con 3 veces de cambio o más y se obtuvo un valor de \log_2 . Bajo este criterio se obtuvieron en total 898 genes, a partir de estos valores y mediante un análisis de agrupamiento tipo “K-Means” se obtuvieron 10 perfiles de expresión, que presentan las siguientes características: Perfil 1, genes que se expresan a los 11 ddi y presentan una disminución en la expresión a los 18 ddi (en el resto de las condiciones analizadas se observa que los genes de este perfil están reprimidos), en total son 48 elementos que representan el 5.34 % de los 898. Por ubicación 27 se encuentran en cromosoma (56%), en el plásmido C se ubican 6 elementos (13%) (Fig. 14 inciso b). Perfil 2, genes que se expresan a los 18 ddi y a los 35 ddi aunque en este último punto la expresión es menor, constituido por 201 elementos que representan el 22.38%, de los cuales 130 (65%) pertenecen a cromosoma; en este perfil los 28 genes presentes en el plásmido E (13 %) y los 20 del plásmido F (10%) son los siguientes en abundancia (Fig. 15 inciso b). Perfil 3, constituido por genes que se expresan exclusivamente en vida libre (6 hrs MM), son 178 elementos que representan el 19.82% de los 898, de los cuales 135 (75%) pertenecen a cromosoma, el plásmido F con 18 elementos es el siguiente en porcentaje con el 9% (Fig. 16 inciso b). Perfil 4, a este pertenecen los genes que se expresan exclusivamente a los 35 ddi, 75 genes que representan el 8.35% del total de los genes, de los cuales 46 están ubicados en cromosoma (62%) y 9 en el plásmido F (12%) (Fig. 17 inciso b). El perfil 5 corresponde a genes expresados en los 35 ddi y vida libre, aunque a los 18 ddi a estos genes se observa una menor expresión respecto a los 35 ddi y vida libre, 60 elementos que representan el 6.68% del total, 37 genes se ubican en el cromosoma (63%) y 6 del plásmido F (10%). En este perfil el plásmido B alcanza su máximo porcentaje de elementos con el 7% (Fig. 18 inciso b). Perfil 6, genes que se expresan a los 18 y 35 ddi con un pequeño pico de expresión a los 11 ddi, en total son 70 genes representando el 7.79%, de estos 46 pertenecen a cromosoma (66%) y 11 al plásmido E

(16%) (Fig. 19 inciso b). El perfil 7, constituido por genes expresados a los 25 ddi y vida libre con un pequeño pico de expresión a los 28 ddi, 49 elementos que constituyen el 5.45% del total, de los cuales 40 son de cromosoma (82%) (Fig. 20 inciso b). El perfil 8 son genes de 18 ddi y reprimidos a los 25 y 28 ddi, a diferencia del perfil 2, los genes en el punto de VL presentan una expresión mayor al perfil 2 (18 ddi), en total son 79 genes que representan el 8.79%, de los cuales 50 están ubicados en cromosoma (63%) y 12 en el plásmido F (16%) (Fig. 21 inciso b). El perfil 9 son genes que se expresan a los 25 ddi con una ligera disminución en la expresión a los 28 ddi, 86 elementos que representan el 9.57% del total, de éstos 42 (49%) pertenecen al plásmido D y 32 al cromosoma (37%) (Fig. 22 inciso b). Finalmente, el perfil 10 contiene los genes compartidos entre la vida libre y 18 ddi, son 52 y representan el 5.79%, de los cuales 40 (77%) están ubicados en cromosoma y 4 (7%) del plásmido F. En este perfil el plásmido A alcanza su máximo porcentaje de elementos expresados con el 6% (Fig. 23 inciso b). La distribución de los 898 genes expresados en los diferentes replicones de *R. etli* es la siguiente: 584 son de cromosoma (65%), 26 del pA (3%), 25 del pB (3%), 28 del pC (3%), 72 del pD (8%), 76 del pE (8%), y el pF con 87 genes que representan el 10% (Fig. 24 inciso b).

En cuanto a los grupos funcionales, utilizando "Cluster Orthologous Groups" tenemos que el perfil 1, presenta en el COG "G" (transporte y metabolismo de carbohidratos) su mayor representante con el 19%, por otro lado el 44% del total incluye los COG's "R, S y X" (predicción de función general, Función desconocida y No en COG's respectivamente) (Fig. 14 inciso d). En tanto el perfil 2, el COG "K" (transcripción) con el 10% (el promedio de los 10 perfiles es de 6.2%) y el COG "L" (recombinación, replicación y reparación) con el 6.5% (el promedio de los 10 perfiles es de 2.68%), son los COG's con mayor número de elementos en este perfil (Fig. 15 inciso d). En el perfil 3 destacan el COG "J" (traducción) con el 11.2%, además el COG "C" (Producción y conversión de energía) con el 10.7%, el COG "U" (Tráfico intracelular y secreción) con el 4.5%, este es el porcentaje más alto observado en los 10 perfiles, y por último el COG "O" (Modificación postraduccion, degradación de proteínas y Chaperonas) con el 7.9% (Fig. 16 inciso d). En el perfil 4, los grupos funcionales más representados son: el COG "G" (Transporte y metabolismo de

carbohidratos) con el 13.3%, el COG "E" (Transporte y metabolismo de aminoácidos) con el 11% y el COG Q (Biosíntesis, transporte y catabolismo de metabolitos secundarios) con el 4% es el porcentaje más alto observado en los 10 perfiles (Fig. 17 inciso d). En el perfil 5 tenemos sobrerrepresentados los COGs: "O" (Modificación postraduccional, degradación de proteínas y chaperonas) con el 10% es el porcentaje más alto observado en los 10 perfiles, el COG "P" (Transporte y metabolismo de iones inorgánicos) con el 8.3% es el máximo porcentaje de este COG en los 10 perfiles y el COG "M" (Biogénesis de membrana y pared celular) con el 6.6% (Fig. 18 inciso d). En el perfil 6, el COG "G" (Transporte y metabolismo de carbohidratos) con el 24.3% presenta el porcentaje más alto de este COG en los 10 perfiles analizados, en tanto el COG "E" (Transporte y metabolismo de aminoácidos) con el 15.7% también en este perfil presenta el mayor porcentaje de los 10 perfiles y lo mismo para el COG "I" (Transporte y metabolismo de lípidos) que con el 8.6% es el más alto de los 10 perfiles (Fig. 19 inciso d). En el perfil 7, los COG's que presentan los mayores porcentajes son: el "K" (Transcripción) y "J" (Traducción) con el 10.2%, y el COG "O" (Modificación postraduccional, degradación de proteínas y chaperonas) que en este perfil con el 8.2% alcanza el segundo porcentaje más alto observado en los 10 perfiles (Fig. 20 inciso d). En el perfil 8, nuevamente el COG "K" (Transcripción) con el 10.1%, el COG "I" (Transporte y metabolismo de lípidos) con el 7.6% y el COG "M" (Biogénesis de membrana y pared celular) con el 6.3%, son los COG's de este perfil que están sobrerrepresentados respecto a los otros 10 perfiles (Fig. 21 inciso d). En el perfil 9, el COG "C" (Producción y conversión de energía) con el 12.64% es el porcentaje más alto de los 10 perfiles, el COG "H" (Transporte y metabolismo de coenzimas) con el 2.3% también es el porcentaje más alto de este COG respecto a los 10 perfiles (Fig. 22 inciso d). En el perfil 10, el COG "T" (Mecanismos de traducción de señales) con el 7.7 % presenta el máximo porcentaje de los 10 perfiles, en tanto los COG's "E" (Transporte y metabolismo de aminoácidos) con el 13.46% y el "P" (Transporte y metabolismo de iones inorgánicos) con el 7.7% también están sobrerrepresentados en este perfil (Fig. 23 inciso d). Los porcentajes totales de los 898 genes distribuidos en COG's se presentan en la figura 24 inciso c.

En cuanto a los sitios de regulación localizados observamos que responden a los siguientes reguladores: En el perfil 1, las cajas “Nod-box” son las predominantes con 21, seguida de RpoH con 5. En el perfil 2, RpoD con 37, RpoE y HipB con 19 y TrpI con 11. En el perfil 3, RpoD con 22, RpoE con 20, HipB con 18 y OxyR con 15. En el perfil 4, RpoD con 13, RpoE y HipB con 7 y Nod-Box, NifA y OxyR con 5. En el perfil 5, RpoD con 8, HipB con 6, RpoE 5 y RpoH con 4. En el perfil 6, RpoD con 12, RpoE con 9, HipB con 8 y CRP/FNR con 7. En el perfil 7, RpoE con 11, RpoD con 8, RpoN con 7 y NifA, PhoB y PurR con 6. En el perfil 8, RpoD 19, CRP/FNR con 10, HipB con 8 y MalT y OxyR con 7. En el perfil 9, RpoD con 51, OxyR con 36, RpoN con 35, RpoE con 22 y NifA con 21. En el perfil 10, RpoD y RpoE con 7, RpoN y RpoS con 6 y CRP/FNR con 5. De estos reguladores, RpoD, RpoE y HipB son de los que más sitios se localizaron en la mayoría de los 10 perfiles, los reguladores con un menor número de sitios localizados se ubicaron de acuerdo con el perfil y se pueden observar en las figuras 14-23, inciso f. El consenso de los sitios más representados se presentan en la figura 25; éstos se obtuvieron con el programa RSA-Tools (van Helden *et al.*, 2003) y la imagen del consenso con el programa “WebLogo” (Crooks *et al.*, 2004).

Al analizar los datos en KEGG con la finalidad de obtener las vías más representadas por perfil observamos lo siguiente: Perfil 1, Degradación de cloroalqueno y cloroalcano, Metabolismo de butanoato, Degradación de bisfenol y Metabolismo de fructosa y manosa todos con 8 elementos. Sistema de dos-componentes con 6 y Quimiotaxis con 5. El perfil 2, Metabolismo microbiano en diversos ambientes con 24 elementos fue el más abundante, seguido por Transportadores tipo ABC con 23, Biosíntesis de lisina con 12, Fosforilación oxidativa con 9, Ensamble flagelar y Metabolismo de sacarosa y almidón con 7 elementos cada uno. El perfil 3, Sistema de secreción bacteriano y Ribosoma con 15 elementos cada uno, Metabolismo microbiano en diversos ambientes con 13, Fosforilación oxidativa con 11, Transportadores ABC con 7 y Ciclo del citrato (ciclo de TCA) con 6 elementos son los grupos más representativos de este perfil. El perfil 4, Metabolismo de piruvato con 6, Metabolismo de sacarosa y almidón con 4, Metabolismo de glicina, serina y treonina, Biosíntesis de valina, leucina e isoleucina, Metabolismo de propanoato, Metabolismo de metano y Metabolismo microbiano en diversos ambientes

con 4 elementos cada uno. El perfil 5, Fosforilación oxidativa con 5 elementos, Metabolismo de butanoato, Biosíntesis de valina, leucina e isoleucina, Biosíntesis de Pantotenato y CoA y Metabolismo de ácido dibásico C5-ramificado con 3 elementos cada uno son los más representativos de este perfil. El perfil 6, Transportadores tipo ABC con 9 elementos y Metabolismo de glicerolfosfolípido con 3 elementos son los dos grupos representativos de este perfil. El perfil 7, Ribosoma con 4, Sistema de secreción bacteriano, Sistema de dos-componentes, Metabolismo de pirimidinas, Metabolismo de Purinas y Fosforilación oxidativa con 3 elementos cada uno son los grupos que destacan en este perfil. El perfil 8, Metabolismo microbiano en diversos ambientes con 5 elementos, Degradación de valina, leucina e isoleucina, Metabolismo de beta-Alanina, Metabolismo de arginina y prolina, Metabolismo de ácidos grasos y Metabolismo de propanoato con 4 elementos cada uno son los grupos relevantes de este perfil. El perfil 9, Metabolismo de nitrógeno, Degradación de cloroalqueno y cloroalcano, y Metabolismo microbiano en diversos ambientes con 7 elementos cada uno, además de los Transportadores tipo ABC con 4 elementos son los más representativos de este perfil. El perfil 10, Transportadores tipo ABC con 10, Sistema de dos-componentes y Quimiotaxis con 5 cada uno, además Metabolismo microbiano en diversos ambientes y Metabolismo de fructosa y manosa con 4 elementos cada uno son los representantes de este perfil.

El análisis de GSEA se realizó ordenando de mayor a menor expresión cada uno de los tiempos o condiciones analizadas, es decir, en este caso tenemos los grupos de genes que se enriquecieron a los 11, 18, 25, 28, 35 ddi y vida libre por separado, observándose que los tiempos 25 y 28 ddi resultaron muy similares. Los resultados obtenidos son: Para el tiempo 11 ddi, Genes del Metabolismo de amino-azúcares (enriquecidos positivamente) y los genes Degradación de tetracloroetano y Ribosomales (enriquecidos negativamente) (Fig. 26). En el tiempo 18 ddi, Transportadores tipo ABC, Biosíntesis de ácidos grasos y Genes del Metabolismo de amino-azúcares (enriquecidos positivamente) y los genes Ribosomales, del Ciclo del citrato (ciclo de TCA), del Ciclo de reductor de carboxilato (Fijación de CO₂), Sistema de secreción tipo IV (enriquecidos negativamente) (Fig. 27). Para el tiempo 25ddi, los genes del Metabolismo de nitrógeno y Degradación de

tetracloroetano (enriquecidos positivamente), y los genes del Metabolismo de aminoazúcares Transportadores tipo ABC y Quimiotaxis (enriquecidos negativamente) (Fig. 28). En el tiempo 35ddi, los genes del Metabolismo de piruvato y Biosíntesis de aminoacil-tRNA (enriquecidos positivamente), y los genes del Metabolismo de nitrógeno, Degradación de tetracloroetano y Ribosomales (enriquecidos negativamente) (Fig. 29). Finalmente, en Vida Libre 6 hrs, los genes Ribosomales, del Ciclo del citrato (ciclo de TCA), Fosforilación oxidativa, Sistema de secreción tipo IV y del Ciclo de reductor de carboxilato (Fijación de CO₂) son enriquecidos positivamente, y los genes del Metabolismo de nitrógeno y Degradación de tetracloroetano (enriquecidos negativamente) (Fig. 30).

11.3.2. Descripción de los genes por perfil

Para el desglose de genes pertenecientes a cada perfil, se realizó una búsqueda de la función, posible función, fenotipo simbiótico o no simbiótico (competencia, resistencia a estrés, entre otros); esto se realizó para cada uno de los genes con nombre asignado. La información de los genes hipotéticos o con función desconocida se revisaron de manera individual sólo en los considerados relevantes, por ejemplo, que formen parte del regulón RpoN-NifA (Salazar *et al.*, 2010) o bien tengan participación en la regulación de otros genes, entre otras características a considerar. La búsqueda inicial se realizó en rhizobiales y en simbiosis, si ésta no tenía éxito se amplió primero a bacterias patógenas y por último en todas las referencias sobre el gen en particular.

Perfil 1, 11 ddi

El primer perfil corresponde a los genes expresados preferencialmente a los 11 ddi (Fig. 14) y en menor medida a los 18 ddi. En este perfil destaca *cycM* que codifica para un citocromo; en simbiosis es indispensable para el sistema mantener la concentración de O₂ al mínimo puesto que la nitrogenasa es sensible a la presencia de O₂, en el sistema fijador de nitrógeno los citocromos son los encargados de mantener baja la concentración de O₂. En particular, en *R. tropici*, *cycM* incrementa su expresión cuando se mutageniza el gen que codifica para una sintasa de glicógeno (*glgA*), dicha mutante mejora la capacidad de oxidación de N, N-dimetil-p-phenylendiamine (DMPD) lo que incrementa la respiración vía

la citocromo oxidasa, esto a su vez hace que el peso seco de la planta se incremente alrededor de un 30% (Marroquí *et al.*, 2001). En *B. japonicum*, la mutante *cycM* aparentemente no produce el citocromo aa3 y tiene un fenotipo negativo de oxidasa menos (Bott *et al.*, 1991). Sin embargo, a diferencia de *B. japonicum*, la mutante *cycM* en *S. meliloti* sí produce la citocromo oxidasa aa3 y en vida libre no tiene fenotipo en cuanto a crecimiento y a motilidad (Yurgel, *et al.*, 2007).

En este perfil se observó expresado el gen *cheW* el cual codifica para una proteína de traducción de señales de quimiotaxis. En *Azorhizobium caulinodans* microsimbionte de la leguminosa *Sesbania rostrata*, donde no sólo forma nódulos en las raíces sino también en el tallo, una mutante *cheW* de *Azorhizobium* induce nódulos inefectivos en el tallo de *Sesbania*, aunque los cultivos bacterianos se colocaron directamente en los tallos lo que hace poco probable que la quimiotaxis y motilidad desempeñen un papel de movimiento hacia los sitios de infección. Por lo tanto, CheW pudiera tener un papel en la entrada y colonización dentro del tallo (Susuki *et al.*, 2007).

También se observó expresado el gen *kdpB*, el cual codifica para una ATPasa transportadora de potasio (K⁺) subunidad B. A diferencia de los sistemas Trk y Kup que presentan una moderada afinidad por el potasio y en general son de expresión constitutiva, el sistema Kdp es un sistema inducible con una alta afinidad y especificidad por el K⁺, se expresa cuando las necesidades de K⁺ no son satisfechas por los otros sistemas de transporte jugando un papel vital cuando la concentración y por lo tanto disponibilidad de este ion es muy baja. En *S. meliloti* mutantes en los sistemas Trk, Kup y Kpd son capaces de inducir nódulos fijadores en plantas de alfalfa, las cinéticas de nodulación entre las mutantes y la cepa silvestre son similares; no obstante, la doble mutante *trk-kup* desarrolla un menor número de nódulos que la silvestre a los 11 ddi y en el caso de la triple mutante esta alcanza un similar número de nódulos de la cepa silvestre hasta los 22 ddi (Domínguez-Ferreras *et al.*, 2009).

En *E. coli* los genes que participan en la vía catabólica de L-idonato son regulados por represión catabólica. La gluconato cinasa (*idnK*) cataliza transferencia del fosforil del ATP

al gluconato. El producto resultante es gluconato-6-fosfato el cual es un importante precursor del metabolismo del gluconato. IdnK actúa como dímero compuesto de dos subunidades idénticas. Una cepa de *S. meliloti* con el pSymA deletado, no puede utilizar gluconato como única fuente de carbono, el gen *idnK* se localiza en este plásmido, en tanto, un gen similar, *gntK* se localiza en el pSymB. En *E. coli* tanto IdnK o GntK son suficientes para catabolizar gluconato, en *S. meliloti* el gen *idnK* esta adyacente a los genes *idnD* e *idnO*, los cuales podrían estar involucrados en el catabolismo del L-idoonato (Barnett *et al.*, 2001).

En este perfil también observamos la expresión del gen *glgX*, el cual codifica para una glicosil hidrolasa. En *R. tropici* este gen se localiza cerca de los genes *nodC* y *nodG*, en esta bacteria la expresión de los tres genes es baja cuando se utilizan en fase exponencial los inductores naringenina y exudados de frijol, aunque la expresión aumenta, a partir de los 5 minutos de cultivo, cuando las células son colocadas en exudados de semillas; posteriormente la expresión decae rápidamente. La máxima expresión de estos genes en presencia de naringenina se presenta 8 hrs después de la inducción (Oliveira *et al.*, 2010).

En *A. tumefaciens* la mutante en el gen *exoR* es incapaz de formar biopelículas (biofilms), el producto de este gen es una proteína periplasmática y la mutante genera una cepa hipermucoide debido a la síntesis de exopolisacáridos del tipo de succinoglucanos. Adicionalmente, la mutante compromete la capacidad de la bacteria de adherirse a superficies abióticas así como a tejidos de plantas. Además, la expresión de los genes flagelares disminuye y en consecuencia pierde la motilidad flagelar, sin embargo a pesar de lo anterior los autores indican que el fenotipo de la disminución de la adhesión no debe achacarse completamente a la sobreproducción de succinoglucanos o a la pérdida de motilidad debido a la complejidad del fenómeno (Tomlinson *et al.*, 2010). En *R. leguminosarum* la mutante en medio mínimo desprovisto de una fuente de nitrógeno presenta un fenotipo similar: producen 3 veces más exopolisacáridos que la cepa silvestre, y la mutante también es sensible a pH ácido (Reeve *et al.*, 1997). En *S. meliloti*, la mutante en el gen *exoR* es incapaz de colonizar nódulos aunque sí es capaz de iniciar la infección (formar hilos de infección) pero no llega a producir nódulos; además, esta mutante genera

pseudorevertantes que afectan disminuyendo la síntesis de exopolisacáridos. Estas pseudorevertantes si producen nódulos y fijan nitrógeno. Por último, también en *S. meliloti* ExoR en presencia de amonio no regula la síntesis de exopolisacáridos (Ozga *et al.*, 1994). En esta misma vía observamos expresado al gen *exsB*, que codifica para un regulador transcripcional involucrado en la biosíntesis de succinoglicanos. En *S. meliloti*, en comparación a la cepa silvestre, la mutante en dicho gen monocistrónico produce tres veces más exopolisacáridos (EPS I). En contraste, un elevado número de copias del gen *exsB* tiene como resultado una disminución en la concentración de EPS I hasta de un 20% respecto a la cepa silvestre, indicando que el producto de *exsB* puede tener un efecto negativo en la biosíntesis de EPS I (Becker *et al.*, 1994).

En este perfil también se observa la expresión del gen *glpD*. *E. coli* como muchas otras bacterias, tiene la capacidad de pasar de una fase estacionaria a una exponencial si las condiciones lo permiten, pero no todas las bacterias de la población mantienen esta capacidad y aunque pocas, las bacterias que no proliferan tienen la capacidad para contener contra varios bactericidas, lo que genera una población persistente. En *E. coli*, una mutante en el gen *glpD*, el cual codifica para una deshidrogenasa glicerol-3-fosfato, incrementa la población de células persistentes en dependencia de la edad del inóculo. Esto no sucede con otras mutantes, como *sucB* e *yfgA*, las cuales hacen decrecer la población de persistentes, independientemente de la edad del inóculo (Luidalepp *et al.*, 2011). Por otro lado, en *R. leguminosarum*, mutantes incapaces de crecer en glicerol son deficientes en la competencia para la nodulación, sugiriendo que el catabolismo del glicerol confiere a las bacterias una ventaja en la rizósfera o en el hilo de infección (Ding *et al.*, 2012).

Perfil 1: 11 y 18 ddi.

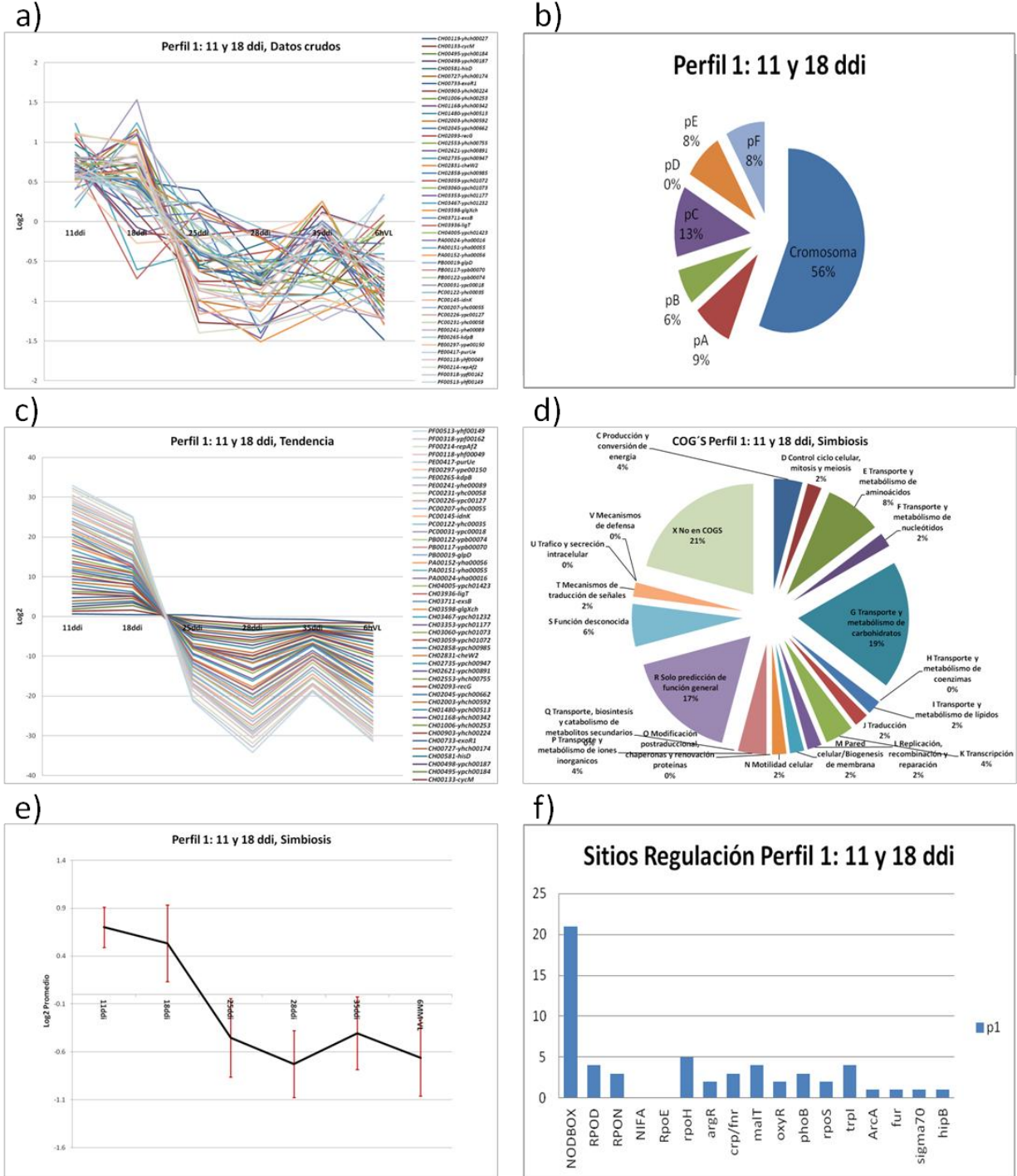


Figura 14. Perfil de expresión de los genes propios o expresados preferencialmente a los 11 y 18 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y, f) sitios de regulación localizados.

El gen *hisD* codifica para una histidinol deshidrogenasa. HisD tiene una actividad bifuncional y participa en los últimos dos pasos de la biosíntesis de la histidina. En *R. etli* CFN42 hay una sola copia de este gen. En *Brassica olearacea* se localiza en plástidos, presenta un péptido de tránsito y puede complementar el correspondiente gen de *E. coli* (Alifano *et al.*, 1996).

El gen *purU* codifica para una deformilasa formiltetrahidrofolato que produce formato a partir de formil-tetrahidrofolato, el cual es la fuente de formato para PurT en la biosíntesis “*de novo*” de purinas. En *E. coli* la mutante en *purU* disminuye la infección sobre violetas africanas (Okinaka *et al.*, 2002).

El gen *ligT* codifica para una RNA ligasa, su función principal se lleva a cabo durante el procesamiento del tRNA, pero además se han descubierto otras funciones para las RNA ligasas tales como la reparación del RNA y participa en eventos no convencionales de “*mRNA splicing*” (Popow *et al.*, 2012).

El gen *recG* codifica para una DNA helicasa dependiente de ATP. En *R. etli* RecG junto con RuvB y RadA, están involucrados en la migración de empalme tipo Holliday, las proteínas del sistema tienen un papel en la sensibilidad a daño de DNA, en particular protege contra cortes durante la exposición a luz UV, ácido nalidíxico y metil metanosulfonato, tanto en cadena sencilla o doble. La mutante *recG* aumenta los eventos de recombinación, lo que indica que RecG limita la plasticidad intra e intergenómica (Martínez-Salazar *et al.*, 2009).

El gen *thuG* codifica para una proteína transportadora de trehalosa y maltosa del tipo ABC. El gen *thuG* alcanza su máxima expresión a los 11 ddi y se mantiene a los 18 ddi. En *S. meliloti*, la mutante *thuE* es incapaz de crecer en trehalosa y presenta un retardo en el crecimiento en maltosa; además, la mutante *thuE* es indistinguible de la cepa silvestre cuando se crece en glucosa y sacarosa. Ensayos de colonización de la raíz, nodulación y fijación de nitrógeno sugieren que la toma de disacáridos puede ser clave para la colonización de raíces de alfalfa, pero que no son importantes para la nodulación ni la fijación de nitrógeno *per se* (Jensen *et al.*, 2002).

Un resumen de los genes expresados en este tiempo (perfil 1, fig. 14), nos indica que éstos están involucrados en la permeabilización de la membrana, suponemos que la finalidad es para que el intercambio de señales, nutrientes, y otros, se presenten de manera efectiva, evitando así la terminación temprana de la interacción entre la bacteria y la planta. Por ejemplo, el gen *exoR* que está involucrado en la síntesis de exopolisacáridos (succinoglucanos) mantiene una baja concentración de estos en el bacteroide o en las bacterias que están en proceso de diferenciación, las mutantes en estos genes producen nódulos pero no fijan nitrógeno. Nuestros datos indican que estos genes son necesarios en etapas tempranas pero posteriores a la infección y formación del nódulo. En este perfil se expresa el gen *exsB* cuyo producto es un regulador negativo de la síntesis de succinoglucanos. Un ejemplo más es la expresión de los genes *thuG* y CH04005 que codifican para permeasas, otro grupo de genes que codifican para permeasas se observan en el perfil 6 (18, 35 y 11 ddi). Adicionalmente, en el análisis de KEGG utilizando GSEA, el metabolismo de amino azúcares es el único grupo que se enriquece en este tiempo (Fig. 26). Los reguladores principales son los Nod involucrados en la regulación de los genes de nodulación y RpoH que está implicado en la respuesta a estrés. Además de los reguladores ExsB y ExoR, el otro regulador transcripcional presente en este perfil es el gen hipotético CH03353 que es un miembro de la familia LysR, a la cual pertenecen los genes *nodD*, sus productos son reguladores de la expresión de los genes de nodulación (Kobayashi *et al.*, 2004). Los COG's mejor representados son el G-Transporte y metabolismo de carbohidratos que contiene en porcentaje casi el doble de elementos que el promedio de todos los perfiles (19% contra el 10%). El COG D-Control del ciclo celular, mitosis y meiosis que también aumenta al doble con respecto al porcentaje de todos los perfiles (2% contra el 1%). Por el otro lado está el COG J-Traducción, que disminuye más del doble en este perfil (2% contra el 5%) y el COG C-Producción y conversión de energía, que disminuye del 6% al 4% (Comparar figuras 14d y 24c).

Perfil 2, 18 ddi

Este es el perfil más numeroso respecto al número de genes que lo conforman y por lo tanto complejo debido a la diversidad de funciones presentes en este perfil.

Algunos elementos contenidos en este perfil son los siguientes:

El gen *aarF* el cual codifica para una proteína de la biosíntesis de ubiquinona. La ubiquinona (UQ) es considerada bajo condiciones de vida libre y en la simbiosis como un mediador en la transferencia de electrones que generan ATP en *R. etli* (Xie *et al.*, 2011).

El gen *accA*, el cual codifica para una proteína de unión periplásmica para agrocinopinas A + B. En *A. tumefaciens* y *A. rhizogenes*, una inserción en el gen *accA* elimina la quimiotaxis. El análisis completo de la región *acc* confirma que *accA* es el único gen requerido del plásmido Ti para la quimiotaxis para agrocinopinas A + B. Las opinas son compuestos de carbono producidas por los callos y los pelos radiculares de las raíces inducidas por *A. tumefaciens* y *A. rhizogenes*. Estos novedosos productos, intermediarios metabólicos de las plantas, son usados como fuente de nutrientes por cepas de *Agrobacterium* para inducir el crecimiento (Kim and Farrand, 1998).

El gen *aceC*, en *Xanthomonas campestris* codifica para una alfa-cellobiosil-difosfoprenil, D-Manosa transferasa, junto con *aceB* participan catalizando el segundo y tercer paso de la vía biosintética del acetano en *X. campestris* (biosíntesis del exopolisacárido acetano) (Griffin *et al.*, 1996).

El gen *afuA* codifica para un transportador de iones, se localiza en membrana y se ha determinado como marcador en el proceso de infección por brucelosis en alces de la región de Wyoming (Lowry *et al.*, 2010). En *Actinobacillus actinomycetemcomitans* el gen *afuA* es regulado por FurR (Haraszthy *et al.*, 2006).

El gen *aroA* codifica para una enzima (3-fosfoshikimato 1-carboxivinil transferasa) que participa en la biosíntesis de aminoácidos aromáticos. En *Xenorhabdus nematophila* este gen es requerido para la colonización mutualista del nemátodo *Steinernema carpocapsae*, es una de las 15 mutantes, de las 3000 analizadas, que disminuyen al menos 15 veces la interacción entre *X. nematophila* y *S. carpocapsae* (Heungens *et al.*, 2002).

El gen *bhbA* codifica para una metil-malonil CoA mutasa. En *S. meliloti* Charles y col., buscando cósmidos que restauraran la capacidad de la bacteria para utilizar poli-3-hidroxi-butirato (PHB) o acetoacetato como única fuente de carbono, identificaron un locus que contenía los genes *bhbA-D*, por otro lado ha sido reportada que la mutante *bdhA* es deficiente en la actividad de 3-hidroxi-butirato deshidrogenasa (Charles *et al.*, 1997). La vitamina cobalamina es requerida para la interacción entre *S. meliloti* y su hospedero (alfalfa), en *S. meliloti* se conocen 3 enzimas que requieren a la cobalamina como cofactor, entre estas *bhbA*. La mutante en este último gen no afecta la simbiosis (Taga and Walker 2010).

El gen *bkdB* codifica para una alfa-ceto deshidrogenasa ácida de cadena ramificada subunidad E2, la cual cataliza junto con otras enzimas, la degradación de cadenas ramificadas de aminoácidos. En *B. subtilis* es inducible por la disminución de temperatura y junto con otros genes de este tipo permiten asegurar e incrementar la fluidez de la membrana en bajas temperaturas, además se sabe que la estabilidad de estos RNA mensajeros aumentan en estas condiciones (Nickel *et al.*, 2004).

El gen *braF* es un transportador de aminoácidos (Hoshino *et al.*, 1992). En *R. leguminosarum* los genes *bra* son esenciales para una efectiva simbiosis, las mutantes en *aap* y en *bra* tienen nódulos pequeños y en menor número, además tienen una menor cantidad de DNA respecto a la silvestre (Prell *et al.*, 2009).

El gen *crtB* codifica para una fitoeno sintasa, la cual participa para la síntesis de carotenoides. Estos funcionan como antioxidantes, utilizando geranylgeranyl pirofosfato para producir fitoeno (Sandmann and Misawa, 1992).

El gen *cyaFch2* codifica para una adenilato ciclasa que cataliza la reacción 3'5'-AMP cíclico (AMPC) de ATP. En *Xanthomonas campestris* el gen *cyaC* se expresa sólo en presencia de peróxido orgánico (Téllez *et al.*, 2002). En este perfil, el gen hipotético con ID CH02560 codifica también para una adenilato ciclasa, en el perfil 6 (incluye genes expresados de 11, 18 y 35 ddi), pero también con un pico máximo a los 18 ddi el gen hipotético con ID

PE00177 codifica para una adenilato ciclasa, lo que indica que es en este punto es en donde se presenta la máxima expresión de las adenilatos ciclasas.

El gen *cycX*, en *B. japonicum*, codifica para una proteína transportadora de grupos hemo, involucrada en la biogénesis del citocromo-C, que junto con *CycM* y *CycV* son esenciales para la formación de prácticamente todos los citocromos tipo C de la célula. La proteína *CycX* está unida a membrana y mutantes en estos genes tienen un fenotipo pleiotrópico, tales como defectos en la respiración por nitrato, oxidación de H₂, y en la respiración microaeróbica de la simbiosis (Ramseier *et al.*, 1991). En el perfil 4 (35 ddi) el gen *cycJ* alcanza su máxima expresión.

El gen *dadR* codifica para un regulador transcripcional catabólico de alanina, participando en el catabolismo de alanina en *A. tumefaciens* (De Costa *et al.*, 2001).

El gen *dapAb* codifica para una dihidrodipicolinato sintasa. En *S. meliloti* se expresa constitutivamente y está relacionado con la biosíntesis de lisina (García-Rodríguez *et al.*, 2000).

El gen *dapB2* codifica para una dihidrodipicolinato reductasa, la cual está involucrada en la biosíntesis de lisina. En *S. meliloti* GR4, la mutante *dapB* nodula alfalfa al mismo nivel que la cepa silvestre, pero la mutante no es auxótrofa de lisina. Los autores especulan con la presencia de un *dapB* adicional (García-Rodríguez *et al.*, 2000); en ese sentido, *R. etli* CFN42 tiene anotados 2 genes *dapB*, ambos localizados en el cromosoma.

Los genes *dppCch2* y *dppDf*, codifican para transportadores de dipéptidos. En *R. leguminosarum*, el operón *dpp* es regulado por sigma 70, mutantes en estos genes no tienen un fenotipo simbiótico en habas. Las mutantes *dpp* son severamente afectadas en la importación del ácido delta-aminolevulinico (ALA) necesario para la síntesis de hemo, la toma de ALA antagoniza al agregar dipéptidos, lo que indica que estos dos tipos de moléculas utilizan el mismo transportador. Por lo que los autores sugieren que ALA no es proporcionada por la planta, ni que otro sistema de toma de ALA funcione en bacteroides y por lo tanto señalan que los bacteroides producen suficiente ALA (Carter *et al.*, 2002).

Perfil 2: 18 y 35 ddi.

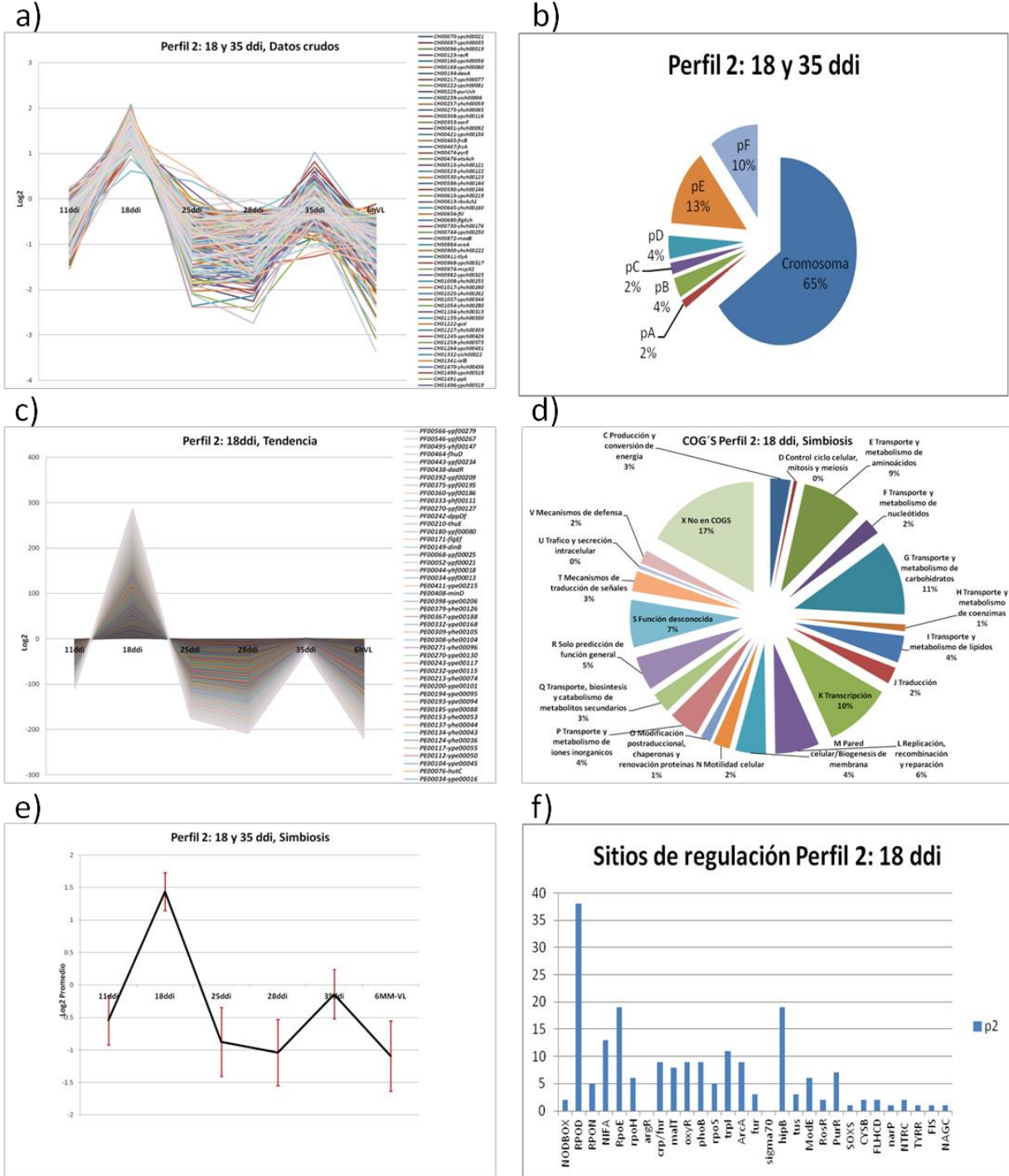


Figura 15. Perfil de expresión de los genes propios o expresados preferencialmente a los 18 ddi y con un ligero repunte de expresión a los 35 ddi, a) datos crudos, b) distribución por replicación, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *fabF* codifica para una 3-oxoacil-(proteína acarreadora de acilo) sintasa II. El producto de este gen participa en la biosíntesis de ácidos grasos saturados y no saturados (Zhu *et al.*, 2009). El gen *fabH* que codifica para una 3-oxoacil-(proteína acarreadora de acilo) sintasa III se localiza en el perfil 8 (18, 35 ddi y MM).

El gen *fhuD* de *R. leguminosarum*, cuyo producto participa en la toma de sideróforos como la vicibactina, se localiza en membrana. La mutante en el operón *fhuDCB* aparentemente no tiene fenotipo simbiótico y se regula por concentración de *hierro* (Stevens *et al.*, 1999). El gen *fhuB* se localiza en el perfil 7 (25 ddi y MM).

El gen *flgK* cuyo producto participa en la formación del flagelo, el cual definimos que en vida libre (microaerobiosis) y en simbiosis es regulado por sigma 54 y probablemente por NifA en simbiosis (Salazar *et al.*, 2010). En este perfil también se expresan los genes *flil* el cual codifica para una ATPasa y *flgE* codifica para una proteína del gancho flagelar, los tres genes participan en la motilidad celular.

El gen *frcA* codifica para una proteína de unión a ATP, citoplasmática tipo ABC. La cual en *S. meliloti* es la única con una alta afinidad (Km de 6uM) para la toma de fructosa. Así mismo, el gen *frcB* codifica para una proteína periplasmática de unión al sustrato, tipo ABC, su punto máximo de expresión es a los 18 ddi. El sistema Frc tiene una alta afinidad por manosa y ribosa, que son competidores de la fructosa. Sin embargo, estos genes se inducen sólo por la concentración de fructosa o manitol; las mutantes no se ven afectadas en la nodulación ni en la fijación de nitrógeno. No obstante la proteína FrcB se expresa en bacteroides aislados de nódulos de alfalfa infectados con *S. meliloti* (Lambert *et al.*, 2001).

El gen *gcd* codifica para una glucosa deshidrogenasa (pirroloquinolina-quinona). En *S. meliloti* RCR2011 y en *R. tropici* CIAT899 se ha propuesto que Gcd o GDH otorga ventajas energéticas y de competencia en condiciones específicas como pudieran ser la limitación de nutrientes o el estrés (Bernardelli *et al.*, 2001).

El gen *glmU* codifica para la proteína UDP-N-acetilglucosamina pirofosforilasa, su función es catalizar la conversión de N-acetilglucosamina 1-fosfato en UDP-N-acetilglucosamina, la

cual es la llave metabólica en la síntesis de lipopolisacáridos, peptidoglicanos y ácido siálico (Ullrich and van Putten, 1995).

El gen *hutC* codifica para un represor transcripcional de los genes para la utilización de histidina (genes *hut*). En el perfil 1 (11ddi) se detecta la expresión del gen *hisD* que participa en la biosíntesis de histidina. HutC presenta dos funciones: la primera es de coactivador de los genes *virB* y la segunda función es reprimiendo los genes *hut*. En *Brucella* se propone que probablemente este sea la ruta por la que la bacteria sensa su propio estado metabólico y el desarrollo de las respuestas adaptativas del hospedero librando las defensas intracelulares (Sieira *et al.*, 2010).

El gen *ilvDc* codifica para una ácido-dihidroxi dehidratasa. La mutante en *ilvD* de *S. meliloti* es auxótrofa de isoleucina y valina, aunque el fenotipo simbiótico de ésta es muy similar a la silvestre (Hassani *et al.*, 2002). No obstante, en un estudio posterior también realizado en *S. meliloti*, indicaron que la mutante en *ilvD* tiene un fenotipo simbiótico, concretamente en la infección (Pobigaylo *et al.*, 2008).

El gen *iolB* codifica para una proteína del catabolismo del mio-inositol, la mutante no tiene fenotipo simbiótico, el peso seco de plantas y la reducción de acetileno son iguales a la cepa silvestre, pero durante el proceso de infección en experimentos de competencia la mutante es desplazada por la silvestre (Fry *et al.*, 2001).

El gen *minD* codifica para una proteína inhibidora de división celular, MinD es una ATPasa asociada a membrana que tiene un papel fundamental seleccionando el punto medio de la célula sobre otros sitios a través de la activación y regulación de MinC y MinE. El gen *minE* lo observamos expresado a los 35 ddi (perfil 4).

El gen *moaB* codifica para una proteína de molibdeno con importancia en la catálisis de reacciones redox en el ciclo del nitrógeno, en particular en el paso final de la síntesis del cofactor Molibdeno, catalizando la adenilación de MTP (Terina de unión a metal) que es una vía dependiente de Mg²⁺ y ATP. La actividad de MoaB es de adenilil transferasa y es

esencial para la inserción de metal en la maduración de los cofactores tungsteno y molibdeno (Bever *et al.*, 2008).

El gen *ntrY* codifica para un regulador (sensor) del tipo de sistema de dos componentes involucrado con la regulación del nitrógeno. La mutante *ntrY* presenta un fenotipo simbiótico deteriorado, presenta una deficiente nodulación y una fijación de nitrógeno muy baja, esto analizado en *R. tropici* (Nogales *et al.*, 2002). Por otro lado, en *R. etli* CFN42 la proteína NtrX fue detectada a los 18 ddi (Resendis-Antonio *et al.*, 2011). El sistema de dos componentes funciona con la interacción entre NtrY y NtrX, el primero funciona como sensor y NtrX como ejecutor.

El gen *nuoB*, codifica para la subunidad B de una NADH deshidrogenasa. En *E. coli* es regulada por O₂ y nitrato vía ArcA, NarL, FNR u IHF e incluso por DctD (Bongaerts *et al.*, 1995).

El gen *otsA*, codifica para una trehalosa-6-fosfato sintasa. La trehalosa es un disacárido que es acumulado por muchos organismos para protegerse durante un estrés fisiológico. En *B. japonicum*, *otsA* se induce por los estreses de salinidad y desecación. Plantas de soya inoculadas con la mutante *otsA* producen sólo algunos nódulos maduros y una gran cantidad de nódulos inmaduros, esto con relación a los producidos por la cepa silvestre. El fenotipo simbiótico indica que OtsA y la acumulación de trehalosa tienen un papel relevante en la vida libre en la protección a la salinidad y durante la simbiosis (Sugamara *et al.*, 2010). En *R. leguminosarum* la doble mutante *otsA* y *treY* fueron menos competitivas en la capacidad de nodular. En el mismo estudio indican que la regulación es postranscripcional, dependiendo de la disponibilidad de la trehalosa (McIntyre *et al.*, 2007).

El gen *phoA* codifica para una fosfatasa alcalina. El fosfato inorgánico es uno de los componentes esenciales para los organismos vivos y está involucrado en un gran número de funciones celulares. A pesar de su relativa abundancia en la naturaleza, el fosfato es un factor limitante para el crecimiento. Para superar la escasez de fosfato en el medio, los

microorganismos producen fosfatasa extracelulares que rescatan el fosfato del medio. Además, se sabe que el fosfato regula negativamente la biosíntesis de antibióticos y otros metabolitos secundarios. En la región promotora de *phoA* localizamos un sitio de unión a PhoB; adicionalmente, en un análisis proteómico de *R. etli* durante la simbiosis con *P. vulgaris*, se identificó la proteína PhoU (regulador transcripcional para la toma de fosfato) (Resendis-Antonio *et al.*, 2011).

El gen *pobR* codifica para un regulador transcripcional, PobR, el cual responde al metabolito p-hidroxibenzoato y activa la transcripción del gen *pobA*, cuyo producto convierte p-hidroxibenzoato a protocatechuato. Este compuesto difiere de p-hidroxibenzoato porque el protocatechuato contiene un átomo de oxígeno adicional (Kok *et al.*, 1998).

El gen *ppk* codifica para una cinasa polifosfato, la cual en muchas bacterias es la enzima principal para la síntesis de polifosfato inorgánico (poli P) a partir de ATP. La mutante en *ppk* de *P. aeruginosa* PA01 es deficiente en la motilidad celular (defecto en el “swarming”) sobre placas de agar semisólido. Debido a la importancia de la motilidad bacteriana, PPK es de crucial relevancia para la infección sistemática al hospedero, para la formación del biofilm, la simbiosis y la virulencia (Rashid y Kornberg, 2000).

Los genes *purC* y *purU* codifican para una fosforibosilaminoimidazol-succinocarboxamida sintasa y una formiltetrahidrofolato deformilasa, respectivamente. El gen *purU* se localiza en el perfil 1 (11ddi). El gen *purU* produce formato a partir de formil-tetrahidrofolato, el cual es la fuente de formato para PurT en la biosíntesis “*de novo*” de purinas; adicionalmente, tiene un papel en el metabolismo de carbono, se activa por metionina y se inhibe por glicina. En *E. coli* la mutante en *purU* disminuye la infección sobre violetas africanas (Okinaka *et al.*, 2002).

El gen *pyrE* codifica para una orotato ribosiltransferasa. En *S. meliloti* una mutante *pyrE* retarda la nodulación (Pobigaylo *et al.*, 2008); previamente, Vineetha y colaboradores, reportaron en 2001 que la mutante *pyrF/pyrE* produce nódulos pero no fija nitrógeno. Los

nódulos producidos por la mutante *pyrF/pyrE* presentan características estructurales más avanzadas que los nódulos de la mutante *pyrC*, esto podría indicar que algunos intermediarios de la biosíntesis de las pirimidinas tienen un papel sobre la transformación del bacteroide y del desarrollo del nódulo (Vineetha *et al.*, 2001).

El gen *rbsA* codifica para un transportador de ribosa de tipo ABC, RbsA está involucrado en el acoplamiento de la energía (ATP), para el transporte de ribosa (Buckel *et al.*, 1986).

El gen *sqdD* codifica para una glicosiltransferasa (biosíntesis de sulfolípidos) y se localiza comúnmente en las membranas de tilacoides de bacterias fotosintéticas o de plantas. Adicionalmente, este tipo de proteínas se localizaron en bacterias no fotosintéticas que producen nódulos y fijan nitrógeno como, *S. meliloti*. La mutante en el gen *sqdB* no presenta un fenotipo simbiótico evidente, la mutante produce nódulos y fija nitrógeno de manera similar a la silvestre; sin embargo, los autores proponen que aún falta por analizar los lípidos de membrana, entre ellos los sulfolípidos, porque el producto de *sqdD* está involucrado en la biosíntesis de al menos uno de estos sulfolípidos (Weissenmayer *et al.*, 2000).

El gen *thuE* codifica para una proteína transportadora de trehalosa y maltosa del tipo ABC. En el perfil 1 observamos al gen *thuG* el cual presenta su máxima expresión a los 11 ddi pero se mantiene a los 18 ddi. En el artículo de Resendis-Antonio *et al.*, 2011, se reportó la identificación de la proteína ThuE, también en este tiempo (18 ddi). En *S. meliloti* una mutante *thuE* pierde la capacidad de crecer en trehalosa o maltosa, pero crece de manera similar a la cepa silvestre en glucosa o sacarosa. El gen *thuE* se induce sólo por trehalosa pero no por celobiosa, glucosa, maltopentosa, maltosa, manitol o sacarosa. La mutante no presenta fenotipo simbiótico pero es incapaz de formar nódulos cuando se inocula junto con la cepa silvestre (Jensen *et al.*, 2002).

El gen *uvrB* codifica para una proteína de reparación de DNA, este tipo de genes se han caracterizado en ambientes de desecación. La mutante en *uvrB* no presenta fenotipo simbiótico pero el DNA presenta daño si la bacteria es expuesta a la luz UV, o al compuesto 4-NQO. La mutante también es sensible a la desecación (Humann *et al.*, 2008).

En resumen, este perfil es el más numeroso e incluso se complementa con el perfil 8, donde también se observa la máxima expresión a los 18 ddi, pero diferente al perfil 2, donde la condición de vida libre (6h-VL) es el punto más bajo de expresión. En tanto, en el perfil 8, la condición de vida libre tiene el segundo punto más alto de expresión en el perfil (ver figuras 15 y 21 incisos a, c, e, además de la figura 20 inciso a). En este perfil se presentan genes cuyos productos participan en el mecanismo de defensa de la bacteria, genes que codifican para el intercambio de señales y de utilización de fuentes de carbono, así como algunos involucrados en la motilidad. Los fenotipos de mutantes en varios genes de este perfil son capaces de nodular pero presentan defectos en la competencia para la nodulación, así como una disminución en la tasa de fijación de nitrógeno. En el análisis mediante GSEA se observan enriquecidos los grupos de transportadores tipo ABC, el grupo de ácidos grasos y en menor medida el de los amino azúcares (Fig. 27). En cuanto a la regulación este perfil también es diverso. La familia más abundante es la GntR que participan como represores involucrados en una diversidad de procesos biológicos. En *S. meliloti* mutantes en los genes *gtrA* y *gtrB*, tienen una baja motilidad celular y también son deficientes en la simbiosis donde los nódulos son pequeños, redondos, ligeramente rosas y se localizan en las raíces laterales. Además, a las tres a cuatro semanas posteriores a la inoculación, las plantas inoculadas con las mutantes lucen atrofiadas con una disminución en la actividad de la nitrogenasa, pero con un remarcado incremento en la cantidad de nódulos respecto a la cepa silvestre. Por último, las mutantes reducen su capacidad de competencia para la nodulación (Wang *et al.*, 2008). El factor sigma que presenta un número importante de sitios en los genes de este perfil es RpoE, con 19 sitios localizados. En este perfil se expresa una de las cuatro copias para *rpoE* (PF00052). Además, en este perfil ya localizamos un número importante de sitios para RpoN (5) y NifA (13) pero ningún elemento de los observados comparte ambos sitios. Finalmente también están presentes sitios para HipB con 19, FNR con 10 y para el factor sigma RpoH con 6. En este perfil se expresan dos genes que codifican para reguladores de la familia LuxR, miembros de esta familia regulan genes cuyos productos están involucrados en la quimiotaxis. Este perfil es uno de los más diversos en reguladores transcripcionales, y

prueba de ello se observa al comparar el COG K-Transcripción, donde el porcentaje general de los 10 perfiles es del 6%, en contraste en el perfil 2 llega al 10%, un porcentaje similar al perfil 8 (18 ddi) y al perfil 7 (VL y 25 ddi). Otro COG que se modifica es el L-Replicación, recombinación y reparación, que presenta un incremento en porcentaje de componentes de un 3% en todos los perfiles analizados a un 6% en este perfil. El COG V-Mecanismos de defensa, el porcentaje promedio de los 10 perfiles es de 1%, a diferencia del perfil 2 donde se incrementa al 2%. El COG C-Producción y conversión de energía, disminuye del 6% al 3% en este perfil, al igual que el COG O-Modificación postraduccional, chaperonas y renovación de proteínas, que disminuye del 4% al 1%.

Perfil 3 Vida libre

Los genes presentes en el perfil 3, tienen una expresión alta en vida libre y una represión en los 5 puntos analizados de la simbiosis. Su descripción es la siguiente:

Los genes *aapM* y *aapJ* codifican para transportadores tipo ABC para L-amino ácidos, codificando para una permeasa y una proteína de unión a sustrato respectivamente. Durante la simbiosis en nódulos indeterminados de habas los bacteroides reducen su abundancia y presentan un bajo número de cromosomas (Prell *et al.*, 2010), esto reduce la transcripción de vías que participan en la biosíntesis de cadenas de aminoácidos ramificados por lo que la bacteria para su provisión depende del huésped; este fenómeno se conoce como “auxotrofia simbiótica”. En *R. leguminosarum* en asociación con *P. vulgaris*, las mutantes en estos genes producen un fenotipo de limitación o “starvation” de nitrógeno. En asociación con plantas de haba se reduce el número de nódulos y la fijación de nitrógeno; sin embargo, una doble mutante en *S. meliloti* 2011 no presenta fenotipo simbiótico en asociación con plantas de alfalfa (*Medicago sativa*). Sólo una pequeña cantidad de este tipo de aminoácidos es necesaria para revertir la auxotrofia simbiótica (Prell *et al.*, 2010). Esto explicaría en parte la presencia de estos genes en el perfil 3 (VL), ya que por un lado es el huésped el proveedor de este tipo de aminoácidos aunado a la baja cantidad de estos requerida para revertir la auxotrofia simbiótica. No obstante, en los perfiles 2 (18ddi) se localiza el gen *braF1* y en el perfil 10 (18 ddi y VL) se

expresa el gen *braC1*. En *R. etli*, las proteínas BraC2 y AapJ fueron detectadas en bacteroides de nódulos de 18 ddi (Resendis-Antonio *et al.*, 2011). Entre los genes regulados por la chaperona Hfq, 34 de ellos codifican para transportadores tipo ABC, entre estos los genes *bra* y *aap* (Mulley *et al.*, 2011).

El gen *aas* codifica para una aciltransferasa acil glicerol fosfoetanol amina, la cual participa en el transporte y metabolismo de lípidos, se localiza en membrana y posee una alta afinidad de unión a la proteína acarreadora de acilos o ACP que se encuentra de manera soluble (Cooper *et al.*, 1989). En este mismo perfil observamos la expresión del gen *acpP*, el cual codifica para una proteína acarreadora de acilos. Este tipo de proteínas funciona durante el crecimiento de cadenas de acilos como acarreador y como donador de cadenas de acilo durante la transferencia a las moléculas blanco; en rhizobios, están involucradas en la biosíntesis y transferencia de ácidos grasos comunes. En estos organismos se han reportado 3 tipos especializados de ACP's, las tres son requeridas para la biosíntesis de moléculas de superficie celular y tienen un papel relevante en el establecimiento de la relación simbiótica (Geiger y López-Lara, 2002).

El gen *acnA* codifica para una aconitasa, una enzima que participa en el ciclo de TCA. En *B. japonicum* el gen se expresa 10 veces menos en un crecimiento aeróbico respecto a un crecimiento anaeróbico. En una mutante *acnA* en vida libre, la actividad de la aconitasa disminuye en más del 70%, pero la habilidad de la bacteria de infectar y nodular plantas de soya no se ve afectada por lo que se plantea una copia adicional (Thöny-Meyer y Künzler, 1996).

Los genes *atpC*, *atpD*, *atpA*, *atpE* y *atpF* codifican para ATP sintetisas FOF1 subunidades épsilon, beta, alfa, C y B', respectivamente. Comparados con la expresión en fase exponencial, los genes involucrados en la producción de energía disminuyen drásticamente la expresión en bacteroides y durante la fase estacionaria en *R. etli* (Vercruyse *et al.*, 2011).

Los genes *azuPd* y *azuPf*, codifican para pseudoazurinas de cobre azul que probablemente participan como donadores de electrones para la nitrito reductasa. Estos genes podrían estar regulados por NnrR (Gómez-Hernández *et al.*, 2011).

El gen *chvE* codifica para un transportador tipo ABC para xilosa. En *Agrobacterium* hay una interacción física entre VirA y ChvE en el periplasma, lo cual enlaza a una señal en el citoplasma generada por compuestos fenólicos (Shimoda *et al.*, 1993). Otro gen de transporte de xilosa, *gguA*, codifica para un transportador tipo ABC para xilosa de unión a ATP y se localiza en el perfil 10 (18 y MM).

El gen *cinI* codifica para una N-Acil lactona sintasa (AHL). Se ha determinado que la expresión de este gen es dependiente de crecimiento, al eliminar la región promotora de *cinR*, éste se sobreexpresa. Las mutantes en *cinI* y en *cinR* tiene un efecto negativo en la fijación de nitrógeno ya que esta disminuye entre un 60 y 70%, comparada con la cepa silvestre de *R. etli*; además, mediante microscopia electrónica, se ha observado que los bacteroides de la mutante *cinI* no se diferencian completamente en comparación con los bacteroides de la cepa silvestre (Daniels *et al.*, 2002).

El gen *clpP* codifica para una proteasa subunidad proteolítica dependiente de ATP. En *S. meliloti*, ClpP es una de las 15 proteínas que modifican su expresión en cultivos con un pH ácido (Reeve *et al.*, 2004). La proteína ClpP disminuye su expresión ante el pH ácido, lo que sugeriría que el pH en los nódulos *R. etli-P. vulgaris* es también ácido.

El genes *cyoB* codifica para una citocromo O ubiquinol oxidasa, subunidad I, este tipo de citocromos se expresa bajo condiciones microaeróbicas y se localiza en el pE. La mutante *cyoA* (del perfil 8) que codifica para una citocromo O ubiquinol oxidasa, subunidad II, de *R. etli* CE3, crece adecuadamente en medio rico y medio mínimo en condiciones aeróbicas pero sufre una severa reducción en el crecimiento en medio mínimo en condiciones microaeróbicas (Landeta *et al.*, 2011). En *E. coli* la proteína CyoB tiene una baja afinidad por el oxígeno y su expresión depende de pH (Egawa *et al.*, 2011).

El gen *cyaCe* codifica para una adenilato ciclasa que cataliza la reacción 3'5'-AMP cíclico (AMPC) de ATP. En *X. campestris* el gen *cyaC* se expresa sólo en presencia de peróxido orgánico (Tellez *et al.*, 2002). La expresión de las adenilato ciclasas *cyaFch2*, CH02560 y PE00177 tienen su punto máximo de expresión a los 18 ddi.

El gen *dapF* codifica para una epimerasa diaminopimelato, ésta participa en la última parte de la síntesis de la lisina. El producto de DapF es el Meso-diaminopimelato, el cual está involucrado en la síntesis de peptidoglicanos como uno de los dos amino ácidos no proteínogénicos en los puentes de los pentapéptidos (Neef *et al.*, 2011).

El gen *dnaK* codifica para una chaperona, este tipo de proteínas confieren resistencia o tolerancia al estrés por temperatura. En *R. etli* en condiciones de crecimiento aeróbico (medio rico, medio mínimo con biotina o medio mínimo con tiamina) o en condiciones fermentativas (medio mínimo sin suplementos), DnaK junto con GroEL y AniA fueron las proteínas que se mantuvieron constantes en expresión a lo largo de las diferentes condiciones evaluadas (Encarnación *et al.*, 2003).

El gen *exoN* codifica para una uridililtransferasa UTP-glucosa-1-fosfato. Los productos codificados por los genes *exo* están involucrados en la síntesis de succinoglicanos. La mutante *exoN* disminuye la producción de succinoglicanos pero es eficiente durante la simbiosis. La localización de la proteína ExoN es probablemente citoplasmática (Glucksmann *et al.*, 1993).

El gen *fixHf* del plásmido F codifica para una bomba de cationes de membrana y se regula por FnrR. En el perfil 7 (25 ddi, MM) se expresa el gen *fixNd*, localizado en el plásmido simbiótico.

El gen *flaCch1* codifica para una flagelina C. En *R. leguminosarum* los 7 genes que codifican para flagelinas se transcriben de manera individual. La mutante *flaC* produce filamentos flagelares cortos y como consecuencia reducen la motilidad, el “swimming” y el “swarming” (Tambalo *et al.*, 2010).

El gen *flaF* codifica para un regulador de la biosíntesis de flagelo, funciona como un activador en la traducción de flagelinas, y pudiera ser requerido para el ensamble o secreción del filamento. En *R. etli*, en la mutante *rsh*, el gen *flaF* baja su expresión de manera significativa comparado con la cepa silvestre. En tanto el gen *flgD* que codifica para la proteína de modificación de la varilla del cuerpo basal de flagelo, al igual que *flaF* se expresan en la cepa silvestre contrastando con la mutante *rsh* (Vercruyssen *et al.*, 2011).

El gen *fpr* codifica para una reductasa ferredoxina-NADP+, es parte del COG C-Producción y conversión de energía, uno de los COG's que contiene un mayor número de elementos en este perfil.

Los genes *ftsW* y *ftsZ* codifican para proteínas de división celular. Se sugiere que FtsW cumple con dos funciones: reclutar el sitio de división y facilitar la síntesis septal de peptidoglicano a través de la transpeptidasa FtsI. En tanto, FtsZ es polimerizado para formar el anillo Z (Mercer y Weiss, 2002).

El gen *fumC* codifica para una fumarato hidratasa. Al igual que otros genes del ciclo de TCA se sobreexpresa en VL. En *R. etli* en nódulos de 21 ddi mediante proteómica se identificó la proteína FumC junto con 7 proteínas más que participan en el ciclo de TCA (Resendis-António *et al.*, 2011). FumC cataliza la conversión de fumarato a L-malato, la proteína es estable al calor por lo cual no requiere hierro para su actividad. La expresión en *E. coli* bajo condiciones anaeróbicas es baja (Park y Gunsalus, 1995).

El gen *fusA* codifica para un factor de elongación, *fusA* es parte de los genes esenciales o "Housekeeping genes", su secuencia se ha utilizado para construir filogenias en bacterias.

El gen *glnII* codifica para una glutamino sintetasa II, ésta tiene un papel fundamental en el metabolismo de nitrógeno, participa catalizando la reacción de glutamato y amonio produciendo glutamina. En *R. leguminosarum* la transcripción de *glnII* es monocistrónica, y en la región promotora presenta un sitio de unión a RpoN y otro para NtrC, su regulación es dependiente de la disponibilidad de nitrógeno (Patriarca *et al.*, 1992). En *R. etli* identificamos en su región promotora un sitio de unión para RpoN y otro para RpoE.

En este perfil observamos la expresión de los genes *groELch2* y *groELf* los cuales codifican para chaperonas, además del gen *groESch1* (el cual codifica para una co-chaperona). La expresión de las chaperonas y co-chaperonas de *R. etli* alcanza su máximo nivel en VL. Las chaperonas están involucradas en el plegamiento correcto de las proteínas, además en el nódulo pudieran participar en el plegamiento de la nitrogenasa, en contender contra el estrés, en el ensamble correcto de los reguladores transcripcionales al DNA y en la translocación de proteínas entre la bacteria y la planta.

El gen *hbdA* codifica para una 3-hidroxi-butiril-CoA deshidrogenasa. En *R. sphaeroides* participa en el metabolismo del polihidroxi-butirato (PHB) catalizando una reacción dependiente de NAD (Fales *et al.*, 2001).

Los genes *hemNd1* y *hemNf* codifican para unas coproporfirinógenos III oxidasas, involucradas en la biosíntesis del grupo hemo en bacterias anaeróbicas facultativas. En *B. japonicum* existen 2 genes *hemN*, en condiciones aeróbicas no se detectan cambios significativos en la expresión, pero en condiciones microaeróbicas o anaeróbicas ambas se expresan más de 20 veces. Estos genes son regulados por FixK, y en condiciones anaeróbicas *hemN1* es regulada parcialmente por NifA. La doble mutante *hemN1-hemN2* crece de manera normal en medio rico en condiciones aeróbicas. En microaerobiosis y en simbiosis la mutante con un efecto adverso mayor es *hemN2* (Fischer *et al.*, 2001). En el genoma de *R. etli* además de los genes *hemNd1* y *hemNf*, se localiza en el plásmido simbiótico el gen *hemNd2*, que forma parte del regulón NifA-RpoN (Salazar *et al.*, 2010).

Otro gene de este perfil es *hflK* el cual codifica para una proteasa serina hidrolasa transmembranal subunidad K. En *P. aeruginosa* la proteólisis es controlada por la proteasa de membrana FtsH, además es el mayor determinante de resistencia. HflK es un factor accesorio de la proteasa FtsH. Una mutante en *hflK* incrementa la sensibilidad a tobramicina (Hinz *et al.*, 2011).

El gen *ibpA* codifica para una proteína pequeña de choque térmico (sHsp); en *B. japonicum*, estos genes son regulados por ROSE y un elemento de regulación negativa "cis-acting". La presencia de este tipo de proteínas es diversa tanto en cepas de

Bradyrhizobium como de *Rhizobium*, en contraste con otras bacterias en las cuales sólo detectan un promedio de 2 sHsp's (Münchbach *et al.*, 1999).

En este perfil se ubica el gen *lon*, el cual codifica para una proteasa; además, el gen *hrm* que codifica para una proteína de unión a DNA de nucleóide. El gen *nrfA* codifica para una proteína de unión a RNA del tipo Hfq, y el gen *usg* codifica para una fosforibosiltransferasa.

El gen *infB* codifica para un factor del inicio de la traducción IF2. Para el inicio de la síntesis de proteínas se requiere de un complejo conformado por tres factores de iniciación, IF1, IF2 e IF3. El factor IF2 es una proteína de unión a GTP y está involucrado en la selección y unión de la fMet-tRNA (fMet) a la subunidad 30S del ribosoma. El gen *infB* de cloroplasto de frijol complementa la mutante *infB* de *E. coli* (Campos *et al.*, 2001).

El gen *lpsA* codifica para una peptidasa señal de lipoproteína. LpsA es la encargada de adicionar una hexosa a una heptosa distal. Una de las principales características que contrastan entre las bacterias en vida libre y las bacterias en simbiosis es la cantidad de lipopolisacáridos (LPS); en simbiosis la cantidad de LPS disminuye drásticamente (Palma, 2009).

El gen *lpxD* codifica para una UDP-3-O-[3-hidroxiimino] glucosamina N-aciltransferasa, participa en el tercer paso de la biosíntesis del lípido A. En *E. coli* la mutante en *lpxD* es vulnerable a antibióticos; debido a que se permeabiliza la membrana externa. La mutante también tiene un decremento en la síntesis de LPS (Roy y Coleman., 1994).

El gen *mraZ* codifica para una proteína de división celular, en este perfil se localizan también los genes *ftsW* y *ftsZ* que codifican para proteínas de división celular. Cuando las bacterias se diferencian a bacterioide pierden la capacidad de división, por lo tanto los genes de división están reprimidos en la simbiosis.

Perfil 3: Vida Libre (VL).

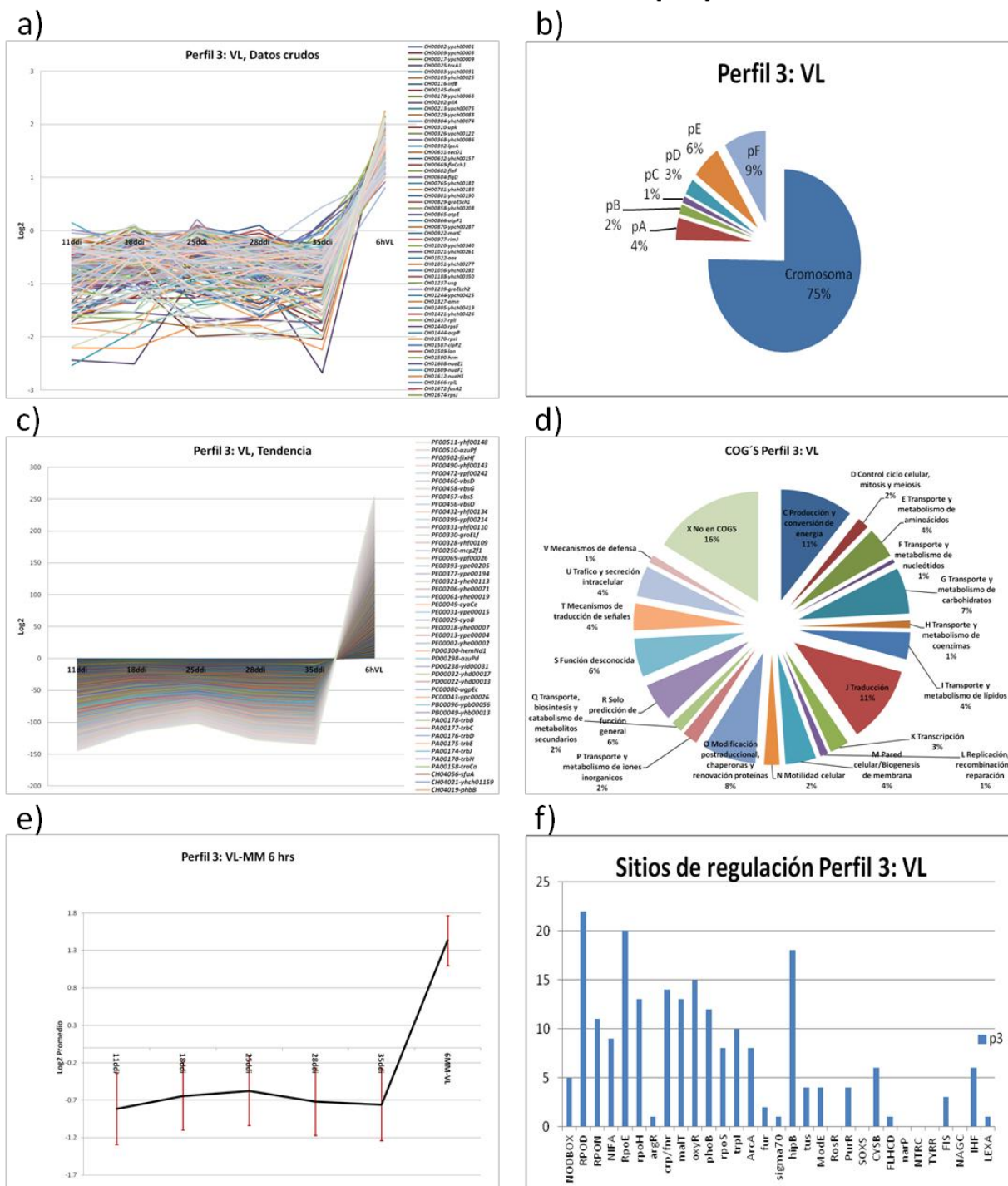


Figura 16. Perfil de expresión de los genes propios o expresados preferencialmente en vida libre (MM 6h), a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *noeJ* codifica para una manosa-1-fosfato guanililtransferasa, la cual está involucrada en la biosíntesis de lipopolisacáridos. En *R. tropici*, al caracterizar nueve mutantes con menor tolerancia a la sal, caracterizaron al gen *noeJ*. Al inocular plantas de *P. vulgaris* con la mutante *noeJ* se observó que se disminuye ostensiblemente la nodulación (Nogales *et al.*, 2002).

Los genes *nuoE1*, *nuoF1* y *nuoH1* codifican para una NADH deshidrogenasa subunidades E, F y H respectivamente. Los productos del locus forman parte de la bomba de protones NADH: quinona oxidoreductasa, la cual acopla la transferencia de electrones de NADH a ubiquinona con la translocación de protones a través de la membrana. Los genes *nuoEFH1* son expresados en vida libre, en tanto, los genes *nuoEB2* se expresan en simbiosis a los 18 ddi (perfiles 2 y 8).

El gen *omp* codifica para una lipoproteína de membrana externa. En *B. melitensis* Omp31 forma oligómeros resistentes en bajas temperaturas a la desnaturalización por SDS, lo que es una característica de varias porinas de bacterias (Vizcaíno *et al.*, 1996).

El gen *pdhA1* codifica para piruvato deshidrogenasa subunidad beta, el complejo de la piruvato deshidrogenasa cataliza la descarboxilación oxidativa del piruvato para producir acetil-CoA, lo que une la glicólisis con el ciclo de Krebs (Cabanes *et al.*, 2000). El gen *pdhA2* alcanza su máxima expresión en el perfil 7 (25 ddi y MM).

El gen *phbB* o *phaB* codifica para una reductasa acetoacetil-CoA, la cual es la segunda enzima de la vía de síntesis de PHB. Los genes *phaA*, *phaB* y *phaC* de *S. meliloti* son suficientes para la producción directa de PHB en una cepa de *E. coli* (Tombolini *et al.*, 1995). En *R. etli*, una mutante *aniA* presenta una drástica alteración de la expresión proteica, que incluye la ausencia de la proteína PhaB o PhbB, lo cual sugiere que AniA podría ser un regulador positivo de *phaB* (Encarnación *et al.*, 2002). El gen con ID CH00017 codifica para una depolimerasa de PHB intracelular y también se localiza en este perfil.

El gen *pilA* codifica para un componente de pili tipo IV, subunidad pilin. En el endófito de pasto, *Azoarcus* sp. BH72, PilA es esencial para la colonización de la planta y es regulado

por densidad de población; además, junto con PilB son indispensables para la adhesión celular (Hauberg-Lotte *et al.*, 2012).

El gen *plyA1* codifica para una polisacaridasa. En el perfil 10 (MM y 18 ddi) se ubica *plyA2*. Las proteínas PlyA, al igual que PlyB, son conocidas como glicanasas y degradan EPS y caboximetil celulosa. En *R. leguminosarum* *plyA* se expresa a muy bajo niveles en cultivos en vida libre. En simbiosis una doble mutante *plyA-plyB* no afecta la fijación de nitrógeno y tiene un efecto insignificante en la nodulación (Finnie *et al.*, 1998). También en *R. leguminosarum* observaron que para la activación de las glicanasas extracelulares (PlyA y PlyB) se requiere de EPS (Zorreguieta *et al.*, 2000).

Los genes *ppiA* y *ppiB* codifican para peptidilprolil cis-trans isomerasas péptido señal A y B respectivamente. En *Lactococcus lactis* el gen *ppiA* se expresa constitutivamente en condiciones normales y en estrés por choque térmico y peróxido. La mutante en *ppiA* no presenta un fenotipo obvio en las condiciones analizadas incluyendo estrés, a excepción de un ligero aumento en la sensibilidad a peróxido (Trémillon *et al.*, 2012). El gen *ppiD2* se ubica en el perfil 10 (MM y 18 ddi).

El gen *ptsP* codifica para la enzima I del sistema de fosfotransferasa, en *R. leguminosarum* (Rlv3841) la mutante en *ptsP* presenta un fenotipo pleiotrópico, formando colonias secas y con poco crecimiento en dicarboxilatos y nitrógeno orgánico. Además, la mutante *ptsP* presenta una muy baja actividad de un amplio rango de transportadores dependientes de ATP tipo ABC (Prell *et al.*, 2012).

Los genes *rimJ*, *rplI*, *rplN*, *rplP*, *rplQ*, *rplY*, *rpmD*, *rpsB*, *rpsF*, *rpsH*, *rpsI*, *rpsJ*, *rpsN*, *rpsQ* y *rpsS*, en el mismo orden codifican para una proteína ribosomal alanina acetiltransferasa, proteína ribosomal 50S L9, proteína ribosomal 50S L14, proteína ribosomal 50S L16, proteína ribosomal 50S L17, proteína ribosomal 50S L25/proteína de estrés general Ctc, proteína ribosomal 50S L30, proteína ribosomal 30S S2, proteína ribosomal 30S S2, proteína ribosomal 30S S6, proteína ribosomal 30S S9, proteína ribosomal 30S S10, proteína ribosomal 30S S14, proteína ribosomal 30S S17, y para una proteína ribosomal 30S S19, respectivamente. Los genes que codifican para proteínas ribosomales localizados

en otros perfiles, también presentan su máxima expresión en la condición de vida libre. El nivel de la expresión de genes que codifican para proteínas asociadas con la maquinaria de traducción está íntimamente relacionado con la velocidad de crecimiento. En bacteroides de *R. leguminosarum* y de *B. japonicum*, se observa que cerca de la mitad de genes que codifican para proteínas ribosomales son reprimidos cuando se comparan con bacterias de un cultivo en fase exponencial (Pessi *et al.*, 2007; Karunakaran *et al.*, 2009).

El gen *secD1* codifica para una translocasa bifuncional de preproteínas subunidad SecD/SecF. En *Rhizobium* sp. NGR234, 132 genes están relacionados con los procesos de secreción. Este rhizobio presenta el más amplio espectro de nodulación de leguminosas, en parte se debe a la cantidad de sistemas de secreción que presenta (Schmeisser *et al.*, 2009). El gen *yajC* codifica para una translocasa de preproteínas, la cual junto con SedDF controla el movimiento de la preproteínas a lo largo de la membrana (Duong y Wickner, 1997).

El gen *sfuA* codifica para una proteína tipo ABC para el transporte de hierro, la expresión del gen *sfuB* se observa en el perfil 5, y su máximo valor de expresión se observa en VL-MM. Los productos de la región de los genes *sfu* de *S. marcescens* confirieron a una cepa de *E. coli* k-12 incapaz de sintetizar sideróforos, la capacidad de crecer en un medio limitado de hierro (Angerer *et al.*, 1990).

Los genes *sucB* y *sucC* codifican para una dihidrolipoamida acetil transferasa y para una succinil-CoA sintetasa subunidad beta, respectivamente. En *R. leguminosarum* la mutante *sucC* evita la expresión de *sucAB*, la mutante *sucB* aumenta los niveles de actividad de succinil-CoA y de malato deshidrogenasa (Walshaw *et al.*, 1997). El gen *mdh* codifica para una malato deshidrogenasa. En *R. leguminosarum* se sugiere el que los genes *mdh-sucCDAB* forman parte de un operón.

El gen *matC* codifica para una proteína acarreadora de dicarboxilatos. El malonato es un ácido dicarboxílico de tres carbonos que es un inhibidor competitivo de la succinato deshidrogenasa. En *R. leguminosarum* bv trifolii la mutante es incapaz de llevar a cabo simbiosis con trébol (Kim *et al.*, 2002).

El gen *tcrX* es un regulador de respuesta de dos-componentes. En *R. etli* el gen *tcrX* se induce por H₂O₂, NaCl y por paraquat, pero no por sacarosa; además, se expresa en fase estacionaria en condiciones microaeróbicas. La expresión de *tcrX* es dependiente de RpoE4. La mutante *tcrX* no presenta fenotipo al cultivarse bajo estrés osmótico ni oxidativo (Martínez-Salazar *et al.*, 2009).

El gen *tesB* codifica para una acil-CoA tioesterasa II. TesB cataliza la hidrólisis de palmitoil-CoA a CoA y palmitato, en cepas de *E. coli* que presentan una larga delección en el gen *tesB* crecen normalmente pero presentan un bajo nivel de actividad de tioesterasa (Naggert *et al.*, 1991)

El gen *tkt* codifica para una transcetolasa, la cual cataliza la transferencia de carbono. En *S. meliloti* una cepa con un defecto en el gen *tkt-2*, es sensible a H₂O₂ y a menadiona (Davies y Walker, 2007).

El gen *tolB* codifica para la proteína de translocación TolB, la cual esta involucrada en la estabilidad de la membrana externa. TolB interactua con Pal cerca de la membrana externa. En *E. coli* defectos en cualquier elemento del sistema TolQR-TolB genera un defecto en la membrana externa, aumenta drásticamente la formación de vesículas, causando una salida de las proteínas periplasmáticas y las células son hipersensibles a agentes nocivos (Lazzaroni *et al.*, 1999).

Los genes *vbsD*, *vbsG*, *vbsS* y *vbsO*, en ese orden, codifican para una proteína de flujo en membrana interna, una poliquétido sintasa, una péptido sintasa no-ribosomal para la biosíntesis de vicibactina y para una 6-monooxigenasa L-lisina (hidroxilasa lisina N(6)), respectivamente. En *R. leguminosarum*, los genes *vbs* están involucrados en la síntesis del sideróforo vicibactina, y su expresión depende del factor sigma extracitoplasmático Rpol. Ninguno de los genes *vbs* se requiere para la fijación simbiótica de nitrógeno (Carter *et al.*, 2002).

Los genes *traCa*, *trbH*, *trbJ*, *trbE*, *trbD*, *trbC* y *trbB*, codifican para proteínas para la transferencia conjugativa C, H, J, E, D, C y B respectivamente. En *R. leguminosarum* los

genes *trb* están involucrados en la transferencia del pSim (pRL1J1); una mutación en el gen *trbL* (ubicado en el perfil 4 (35 ddi) bloquea la transferencia del plásmido simbiótico (Wilkinson *et al.*, 2002). En *R. leguminosarum* adyacentes a *tral-trbBCDEJKLFGHI* se localizan los genes *bisR* y *traR*. BisR induce la expresión de *traR*, en tanto TraR induce la expresión del operón *tral-trb*; la expresión es dependiente de “quórum sensing” a través de la concentración de homoserinas lactonas (Danino *et al.*, 2003).

El gen *trxA1* codifica para una tiorredoxina. En *S. enterica* serovar Typhimurium, TrxA participa en el sistema de secreción tipo III para la virulencia y replicación intracelular. La mutante *trxA* no presenta virulencia en ratones, TrxA es necesaria para la inducción del sistema de patogenicidad de manera intracelular (Negrea *et al.*, 2009).

El gen *tyrA* codifica para una corismato mutasa, involucrada en la síntesis de aminoácidos aromáticos, catalizando la reacción de corismato a pefenato. En la relación endosimbiótica entre *Blochmannia floridanus* y hormigas carpintero, la expresión de los genes de las vías biosintéticas para aminoácidos aromáticos se incrementa al inicio de la interacción (Zientz *et al.*, 2006).

El gen *ugpEc* codifica para un transportador tipo ABC de sn-glicerol-3-fosfato. En *E. coli* éste es regulado de manera negativa por PhoR y PhoST y de manera positiva por PhoB, PhoM y PhoR. Cepas de *E. coli* capaces de utilizar glicerol-2-fosfato como única fuente de fosfato pierden la capacidad cuando se mutageniza el operón *ugpBAECQ* (Yang *et al.*, 2009).

En este perfil son notorios los genes que codifican para proteínas ribosomales, factores de traducción y otros “housekeeping genes”, además de genes cuyos productos participan en la generación de energía, genes de división celular y los genes *trb* para la transferencia conjugativa. Los grupos enriquecidos que corresponden a este perfil son: ribosomas, componentes del TCA, fosforilación oxidativa, el sistema de transporte tipo IV y el ciclo reductivo del carboxilato (fijación de CO₂) (Fig. 30). Los reguladores de la familia AraC que son activadores transcripcionales son los que predominan con 2 elementos de 4 localizados, en cuanto a sitios de unión a DNA identificados RpoD y RpoE con 22 y 20

elementos los más abundantes seguidos por HipB y OxyR con 15. Varios de los sitios consenso para los cuales se realizó la búsqueda, en este perfil aumentan. El COG J-Traducción aumenta un poco más de dos veces en este perfil respecto al promedio total de los 10 perfiles (11% contra 5%), además del COG O-Modificación postraducciona, chaperonas y renovación de proteínas aumenta del 4% en promedio de los 10 perfiles al 8%, el COG C-Producción y conversión de energía con el 11% en este perfil contra el 6% en promedio de los 10 perfiles, por último en este perfil el COG U-Tráfico y secreción intracelular (genes *trb*) alcanza su máximo porcentaje con el 4%. El porcentaje promedio de todos los perfiles en los elementos que componen este COG es del 1%. Los COG's que disminuyen en porcentaje de elementos en este perfil respecto al promedio total de los 10 perfiles son: COG E-Transporte y metabolismo de aminoácidos disminuye de un 8% en los 10 perfiles al 4%, el COG F-Transporte y metabolismo de nucleótidos del 4% en promedio disminuye al 2% en este perfil, y el COG P-Transporte y metabolismo de iones inorgánicos disminuye de un 4% al 2% en este perfil (Fig. 16, inciso d, comparar con fig. 24, inciso c).

Perfil 4, 35 ddi

Este perfil, está conformado por los genes expresados a los 35 ddi el punto más lejano de la simbiosis analizado en este trabajo.

El gen *atpB*, codifica para una ATP sintetasa subunidad A. En la cianobacteria *Trichodesmium* IMS101 el elevar la concentración de CO₂ y la cantidad e intensidad de la luz, produce un decremento en la concentración de varias proteínas clave en la fijación de nitrógeno y carbono como NifH, PsbA y PsaC, mientras que las concentraciones de AtpB y RbcL no sufren cambios significativos. Levitan y col., proponen que el mantener los niveles de proteínas o incluso reducir la biosíntesis de proteínas pudiera liberar energía para utilizarla en el incremento de la fijación de N₂ y el crecimiento en un medio con elevada pCO₂ y luz (Levitan, *et al.*, 2011). El gen *atpI* codifica para una ATP sintetasa subunidad I. Como otras enzimas, la actividad de la ATP sintetasa es reversible; grandes cantidades de ATP generan un cambio en el gradiente de protones transmembranal. Este es utilizado por las bacterias en fermentación que no presentan un transporte en la cadena de electrones,

e hidrolizan ATP, el cual es utilizado por el flagelo y el transporte de nutrientes dentro de la célula. Los genes *atpI* y *atpB*, tienen los ID continuos CH00863 y CH00864, respectivamente, lo que probablemente indique que están en operón.

El gen *bgIS* codifica para una endo-beta-1,3-1,4-glucanasa o beta-glucosidasa, las termitas en simbiosis con bacterias como *Dyella* sp., *Chryseobacterium* sp., y *Bacillus* sp., son capaces de digerir celulosa en su sistema digestivo. La actividad en el aparato digestivo de *Reticulitermes speratus* de la endo-beta 1,4-glucanasa, en 16 cepas bacterianas aisladas, fue baja en todos los casos. Chau y col., 2010 suponen que esta baja actividad es debido a que es realizada por las enzimas de la termita y de los protozoarios en el sistema digestivo, por lo tanto las glucanasas de las bacterias sólo tienen que funcionar para el sistema de degradación bacteriano (Chau *et al.*, 2010).

El gen *birA* codifica para una biotina [acetil-CoA carboxilasa]-sintetasa. La biotina funciona como una coenzima para las 5 caboxilasas dependientes de biotina. La BirA ligasa, que es la contraparte procariotica de la acetil-CoA carboxilasa, juega un papel critico en la señalización celular y en el remodelamiento de la cromatina durante la regulación de la biosíntesis de de biotina en procariotes (Kobza *et al.*, 2008)

El gen *cycJ* codifica para una proteína para la biosíntesis de citocromo-C, junto con los genes *cycHKL* es esencial para la biogénesis de este citocromo. CycJ esta constituido por 164 aminoácidos, estudios de localización lo ubican en el espacio periplasmático, y que probablemente está anclada a la membrana citoplasmática vía un dominio hidrofóbico amino terminal (Ritz *et al.*, 1995). En *R. etli*, la mutante en *cycHJKL* tiene la capacidad íntegra de nodulación pero es incapaz de fijar de nitrógeno (Tabche *et al.*, 1998).

El gen *dppCch*, codifica para una permeasa de tipo ABC transportadora de dipéptidos. Mutantes en *R. leguminosarum* fueron severamente afectadas en el importe del ácido delta-aminolevulinico (ALA), un precursor hemo. La toma de ALA es antagonizada en la presencia de dipéptidos, lo que indica que ambas moléculas utilizan el mismo transporte. En el modelo *V. faba-R. leguminosarum*, mutaciones en *dppABCDF* no presentan un fenotipo simbiótico, lo que indica que el ALA necesario para la síntesis de hemo no es

suplementada por la planta, tampoco se ha identificado otro sistema de transporte que funcione en los bacteroides, lo que sugiere que los bacteroides producen suficiente ALA. Por otro lado, los transcritos de *dppA* y *dppB* contienen inversos repetidos que pudieran estabilizar el RNA mensajero, el promotor para el operón *dppABCDF* es de tipo sigma70 (Carter *et al.*, 2002).

El gen *fdsD* codifica para la proteína formiato deshidrogenasa NAD-dependiente subunidad delta. En condiciones anaeróbicas y en ausencia de un agente alternativo de oxidación, la regeneración de NAD⁺ es acoplada a la fermentación utilizando NADH para reducir intermediarios metabólicos. Por lo que un incremento en la disponibilidad de NADH produciría un efecto en la distribución metabólica. En *E. coli*, la sobreexpresión de una *fdh* NAD(+)-dependiente incrementa al doble la producción de NADH de 2 a 4 moles de NADH/mol por glucosa consumida, incrementando la densidad celular y provocando un cambio significativo en la concentración final de ambos modelos anaeróbico y aeróbico. Bajo condiciones anaeróbicas, la producción de los metabolitos reducidos fue favorecida, como se evidencia con un dramático incremento en la relación entre etanol/acetato. Pero lo más interesante, es la observación durante el crecimiento aeróbico, en donde el incremento en la disponibilidad de NADH induce un cambio a fermentación aún en la presencia de oxígeno (Berríos-Rivera *et al.*, 2002).

El gen *fmt* codifica para una metionil-tRNA formiltransferasa, su función es modificar el grupo amino libre del aminoacil del metionil-tRNA (fMet) el cual es importante en el inicio de la traducción. La inactivación de este gen en *E. coli* tiene un efecto severo en el crecimiento (Meinzel *et al.*, 1993).

El gen *hisS* codifica para una histidinil-tRNA sintetasa, funciona en la traducción de proteínas y es la responsable del anclaje del correspondiente tRNA, esta proteína es dependiente de ATP (Freedman *et al.*, 1985).

Perfil 4: 35 ddi.

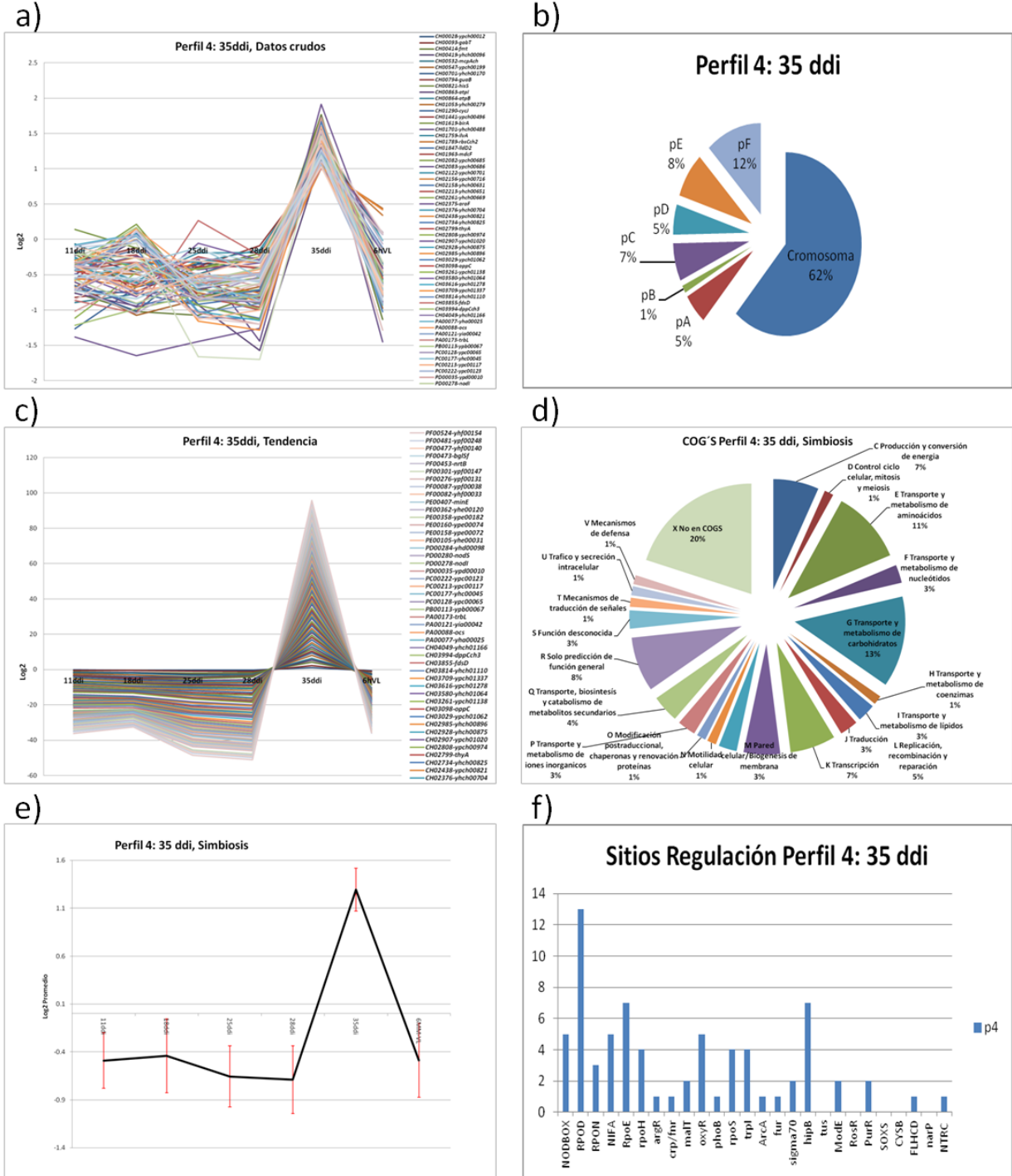


Figura 17. Perfil de expresión de los genes propios o expresados preferencialmente a los 35 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *gabT*, codifica para una aminobutirato aminotransferasa (GABA). En *R. leguminosarum* mutantes en *gabT* aún son capaces de crecer en GABA como única fuente de carbono y nitrógeno. Bajo condiciones de vida libre *gabT* es inducido por GABA y reprimido por succinato, la regulación transcripcional es mediada por GabR que actúa como represor. Durante la simbiosis *gabT* es altamente expresado en la zona simbiótica III aunque la mutante *gabT* aún es efectiva durante la fijación de nitrógeno (Prell *et al.*, 2002). Dos vías para el metabolismo de GABA fueron identificados en *R. leguminosarum*, en una de ellas participa GabT y en la otra vía GabD. Mutantes en *gabT*, *opaA* y *gabD* no pueden usar GABA pero fijan nitrógeno, los autores proponen que si bien el catabolismo de GABA no es esencial para fijar nitrógeno, este puede tener un papel en la generación de energía y en el paso del brazo de descarboxilación del ciclo del ácido tricarboxílico (Prell *et al.*, 2002).

El gen *guaB* codifica para una inositol-5-monofosfato deshidrogenasa. En *R. tropici* realizando una búsqueda de cepas sensibles a altas temperaturas observaron que la mutante en *guaB* no crece en condiciones de altas temperaturas. La mutante no requiere guanina para crecer debido a una vía alternativa a través de la deshidrogenasa de xantina. Además de la sensibilidad al calor, la mutante *guaB* presenta un fenotipo simbiótico formando nódulos carentes de rhizobios (Riccillo *et al.*, 2000). En un artículo posterior los autores utilizaron los promotores para *nodA*, *bacA* y *nifH* fusionados al gen *guaB*, siendo el promotor para *nodA* el que recobró el fenotipo silvestre; lo que indicaría que GuaB se requiere en las etapas iniciales de la interacción. La mutante *guaB* en *S. meliloti* induce nódulos normales en alfalfa y en *M. truncatula*, y analizando la mutante *guaB* de *R. tropici* en *L. leucocephala* ésta no presenta fenotipo simbiótico, por lo que sugieren que *guaB* es requerido para la simbiosis pero sólo en el caso de plantas que forman nódulos determinados (Collavino *et al.*, 2005).

El gen *ilvA* codifica para una treonina deshidratasa, la cual es esencial para la biosíntesis de aminoácidos, en la asociación entre bacterias del género *Buchnera* y afidos, la treonina deshidratasa se sintetiza en el áfido, ya que *Buchnera* no la codifica en su genoma. La

biosíntesis de aminoácidos esenciales por parte del hospedero sugiere que esto le permite controlar el metabolismo de la bacteria (Wilson *et al.*, 2010).

El gen *lldD* codifica para una deshidrogenasa (citocromo) L-lactato. En *E. coli*, en presencia de lactato el gen es inducido y regulado por LldR, el cual actúa como un activador o represor. En ausencia de lactato LldR se une a los sitios O1 (-105 a -89) y O2 (+22 a +38), probablemente formando una horquilla de DNA, lo que evita la transcripción. En presencia de lactato, este se une a LlrR, lo que produce un cambio que puede disgregar la orquilla del DNA, esto permite la formación del complejo abierto (Aguilera *et al.*, 2008).

El gen *mcpA* codifica para una proteína receptora de estímulos quimiotácticos aceptora de grupos metilos. En la bacteria intracelular *Burkholderia* se ha demostrado que ésta se mueve a través del micelio extrarradicular de las estructura de los hongos y esto ocurre en la raíz de la planta hospedera, probablemente en respuesta a un estímulo quimiotáctil (Bianciotto *et al.*, 1996). La motilidad de la bacteria se puede detectar hacia compuestos orgánicos del medio, muchos de los cuales pueden ser utilizados como fuentes de carbono y nitrógeno. La mayoría de las *Burkholderia* se localizan en vacuolas que contienen una gran cantidad de proteínas. Cuando inicia el proceso de esporulación y la disponibilidad de proteínas disminuye dramáticamente, probablemente la bacteria utiliza a McpA, para buscar nuevas fuentes de energía ya sea en el hongo o en la planta, moviéndose a través de ellos (Minerdi *et al.*, 2002).

El gen *mdcF* codifica para una proteína transportadora de malonato. El malonato es un ácido dicarboxílico compuesto por tres carbonos cuya función e importancia durante la simbiosis ha sido comprobado; incluso, debido a su abundancia en el nódulo se ha propuesto que el malonato pudiera ser la mayor fuente de carbono en el bacteroide. Mutantes en genes que codifican para enzimas del metabolismo de malonato en vida libre no presentan fenotipo en esta etapa pero en simbiosis se observa una marcada reducción del crecimiento del nódulo, a primera vista los nódulos lucen similares a los de la cepa silvestre, pero al hacer un corte, los nódulos de la mutante (en este caso *matB*) presentan

zonas senescentes y bacteroides empaquetados con vacuolas y gránulos de almidón (Kim, 2002).

En este perfil también observamos la expresión del gen *minE* el cual codifica para un factor de división celular de especificidad topológica y pertenece al COG D.

El gen *nodI* codifica para una subunidad de transporte de factores de nodulación. La mutante en *nodI* de *Oxytropis arctobia* presenta un fenotipo simbiótico Nod+ pero con un menor número de nódulos respecto a la silvestre en la planta *Onobrychis viciifolia*. Además, los nódulos son 10 veces más pequeños, aunque esta reducción no fue observada en *Astragalus cicer* (Cloutier *et al.*, 1996).

El gen *nodS* codifica para una metil transferasa. La mutante *nodS* en *R. tropici* CIAT899 es incapaz de producir nódulos (Waelkens *et al.*, 1995).

El gen *nrtB* codifica para una permeasa tipo ABC transportadora de nitrato. El nitrato en los suelos es la forma dominante de nitrógeno inorgánico, y puede ser fácilmente asimilado por bacterias, hongos y plantas. En la bacteria *Aspergillus nidulans* se determinó que los productos de los genes *nrtA* (CrnA) y *nrtB* (CrnB) involucrados en el transporte de nitrato tienen un efecto indirecto sobre la activación de NirA, que en esta bacteria es el regulador maestro de los genes involucrados en el metabolismo del nitrato (Schinko *et al.*, 2010).

El gen *ocs* codifica para octopina deshidrogenasa. *A. tumefaciens* induce tumores en plantas y en respuesta éstas liberan octopinas que son utilizadas por la bacteria para colonizar al hospedero. Varios tipos de opinas proveen fuentes de nitrógeno, fósforo y carbono, y recientemente se reportó la toma de sulfuro a través de una sulforoctopina. El sulfuro es un elemento esencial para este patógeno (Flores-Mireles *et al.*, 2012).

El gen *oppC* codifica para un transportador de oligopéptidos tipo ABC, en el perfil 10 (18ddi y VL) se expresa el gen *oppA*. Estos transportadores son requeridos para la toma de tetrapéptidos y de ciertos tripéptidos como 3Ala y bialafos (este último es un herbicida natural que inhibe a la glutamina sintetasa en plantas). En *S. meliloti*, las mutantes en este

tipo de transportadores no tienen un efecto en el establecimiento de la interacción (simbiosis) con la alfalfa; en vida libre, a diferencia de los genes *dpp* los genes *opp* se expresan en fase estacionaria (Nogales *et al.*, 2009).

El gen *rbsCch2* codifica para un transportador de ribosa tipo ABC, en el perfil 6 (genes expresados a los 18 y 35ddi) se identificó el gen *rbsCch3*. Estos genes codifican para una proteína transmembranal. En *M. loti* se caracterizaron dos compuestos, uno de ellos la rhizolectina, ésta no es utilizada como fuente de carbono ni de nitrógeno. A diferencia de las bacterias *S. meliloti*, *R. trifolii* o *A. tumefaciens* que no pueden degradar la rhizolectina, *M. loti* NZP2037 tiene la capacidad de degradarla e incluso uno de los productos de la degradación pudiese ser ribosa. En *M. loti* NZP2037 esta pentosa se ha utilizado como fuente de carbono en el medio S-30 con resultados positivos en el crecimiento (Scott *et al.*, 1987).

El gen *thyA* codifica para una timidilato sintasa; el timidilato es esencial para la construcción de bloques de DNA y se sintetiza a partir de deoxiuridilato utilizando la enzima timidilato sintasa (TS), o bien de timidina (dT) utilizando la timidina cinasa (TK). En *Mycoplasma pneumoniae* la sobre expresión de la timidina cinasa complementa la pérdida de la timidilato sintasa, aunque los autores proponen que podría existir una TS alternativa en el genoma de *M. pneumoniae* (Wang *et al.*, 2010).

El gen *trbL* codifica para una proteína de conjugación. En *R. leguminosarum* los genes *trb* están involucrados en la transferencia del pSim (pRL1JI), una mutación en el gen *trbL* bloquea la transferencia del plásmido simbiótico (Wilkinson *et al.*, 2002). El resto de los genes del operón *trb* pertenecen al perfil 3 (VL).

En resumen el perfil número 4, a pesar de no estar constituido por un amplio número de elementos, es uno de los más interesantes ya que contiene genes que se expresan más abundantemente en el punto más tardío de la fijación de nitrógeno determinado en el presente trabajo, ya que fue hasta este punto en donde se obtuvo una buena cantidad y calidad de RNA. A los 35 días, la expresión de los genes de fijación de nitrógeno así como la actividad de la nitrogenasa es la más baja de los 5 puntos de la simbiosis analizados por

lo que proponemos que la bacteria prácticamente ha dejado de fijar nitrógeno. Los genes en este punto pertenecen a diversos grupos funcionales, pero principalmente al G-Transporte y metabolismo de carbohidratos. El promedio de los elementos de cada perfil que constituyen este grupo funcional en los 10 perfiles es del 10%, y en este perfil aumenta al 13%. Destacan los genes que codifican para proteínas transportadoras de azúcares tipo ABC, además de *rbsCch2*, y el gen con ID RHE_PC00213 que codifica para una proteína que es parte del sistema de transporte de C4-dicarboxilato. También el COG E-Transporte y metabolismo de aminoácidos, aumenta del 8 al 11%, el COG Q-Transporte, biosíntesis y catabolismo de metabolitos secundarios, aumenta del 2 al 4%. En cuanto al análisis de GSEA, los grupos enriquecidos observados están involucrados en la biosíntesis de Aminoacil-tRNAs, y para el metabolismo de piruvato (Fig. 29). Destaca la presencia de transportadores para utilizar diferentes fuentes de carbono como los transportadores tipo ABC para azúcares, RbsC para utilizar ribosa, y con el producto del gen hipotético RHE_PC00213 para dicarboxílicos. También la presencia de genes que pudieran ser claves en cambios metabólicos, como por ejemplo, el gen *birA* cuyo producto en *E. coli* esta involucrado en la biosíntesis de la biotina a su vez relacionada directamente con el metabolismo del piruvato (enriquecido positivamente a los 35 ddi). Otros como el gen *mcpA* cuyo producto participa activamente en la quimiotaxis y su función ayuda a la bacteria dentro del huésped en la búsqueda de nuevas fuentes tanto de carbono, como nitrógeno, entre otros. Toda esta información sugiere que las bacterias no diferenciadas pero presentes dentro del nódulo están en este punto de la simbiosis en un proceso metabólico que les permitirá, al salir del nódulo tener la capacidad de re-colonizar el suelo. Un dato más que sustenta esto, es la expresión del gen *guaB*, cuyo producto participa activamente en la biosíntesis de nucleótidos, también a destacar es la expresión de *minE* cuyo producto participa en la división celular, el gen *minD* lo ubicamos en el perfil 2.

Después del perfil 1 (11 ddi), este es el perfil que más cajas Nod tiene, lo que sugiere que los productos de los genes de nodulación pudieran participar en el mantenimiento y desarrollo del bacteroides en tiempos tardíos, a partir de los 28 ddi las membranas del

simbiosoma comienzan a degradarse (Cermola *et al.*, 2000), por lo que suponemos que a los 35 ddi los bacteroides nuevamente entran en contacto con células de la planta sin infectar y en respuesta liberan señales propias del inicio de la interacción. Un dato por demás interesante es la expresión en este perfil del gen *ypd00010* ya que en el análisis *in silico*, en donde definimos el regulón NifA-RpoN, se identificaron ambos sitios de regulación (Salazar *et al.*, 2010); sin embargo, en los análisis de expresión mediante transcriptomas realizados en vida libre (microaerobiosis) y en simbiosis con *P. vulgaris* a los 11 ddi no se detecta un cambio en la expresión de este gen al comparar la mutante *nifA* y la cepa silvestre. Proponemos que NifA aún participa a los 35 ddi. En la búsqueda de sitios de regulación utilizando el programa “RSA-Tools” (ver material y métodos) nuevamente localizamos ambos sitios (NifA y RpoN) sobre la región promotora de *ypd00010*, pero no detectamos ningún otro sitio para hipotetizar sobre la regulación específica de este gen en este perfil, por lo que ahora lo incluimos como parte del regulón RpoN2-NifA. En cuanto a reguladores transcripcionales en este perfil, los genes con ID’s CH00547, CH02438 y CH02808 codifican para un posible regulador transcripcional, para un miembro de la familia AsnC y para un regulador miembro de la familia AraC, respectivamente. Los genes de la familia Lrp/AsnC (proteína regulatoria de respuesta a leucina/asparagina sintetasa C) son conocidos como proteínas de festín/hambruna (FFRPs por sus siglas en inglés). Estas regulan múltiples metabolismos celulares globalmente (Lrp) o específicamente (AsnC), tales como metabolismo de aminoácidos, síntesis de pili, transacciones de DNA durante la reparación de DNA y la recombinación, y también podrían estar implicados en la persistencia de las bacterias en este punto de la simbiosis.

Perfil 5 (MM y 35 ddi)

En este perfil se observaron genes expresados a los 18, 35 ddi y en VL-MM6h donde se observa la mayor expresión (Fig. 18), en total son 60 genes entre los que destacan los que a continuación se describen.

El gen *afuB* codifica para una permeasa de hierro (III) transportador tipo ABC. En *P. fluorescens* HV37 las proteínas AfuA y AfuB son determinantes para la producción de

antibióticos regulados de manera transcripcional por la presencia u ausencia de la glucosa (Gutterson *et al.*, 1988), además como se describe en el perfil 2, las proteínas Afu se localizan en membrana y son marcadores en el proceso de infección de brucelosis en alces de la región de Wyoming (Lowry *et al.*, 2010).

El gen *argS* codifica para arginil-tRNA sintetasa, funciona en la traducción de proteínas y es la responsable del anclaje del correspondiente tRNA, ésta funciona dependiente de ATP (Eriani *et al.*, 1989).

En este perfil se ubican varios genes para la maduración de citocromo. Como el gen *ccmB*, el cual codifica para una permeasa hemo transportador tipo ABC, (involucrada en la biogénesis de citocromo). Los genes *ccmI* y *ccsA* codifican para proteínas de la biogénesis de citocromo tipo-C; en tanto, el gen *coxP* codifica para una proteína citocromo-c oxidasa subunidad III. Estos genes compartidos por el tiempo 35 ddi y la condición de medio mínimo pudiesen indicar que a este tiempo en el nódulo hay un proceso de cambio en cuanto a la disponibilidad de oxígeno se refiere, aunque el fenotipo en bacterias de mutantes en *ccm* es diverso; por ejemplo pierden la capacidad de producción y utilización de sideróforos, reducen la capacidad de crecer en condiciones de bajo contenido de hierro y disminuyen su habilidad de crecer en células huésped (Cianciotto *et al.*, 2005). En *Legionella pneumophilla*, una mutante *ccmB* es incapaz de crecer en concentraciones bajas de hierro y disminuye su crecimiento intracelular en la amiba acuática *Hartmannella vermiformis* y en la línea celular de macrófagos de humano (Polesky *et al.*, 2001).

El gen *ctaB* codifica para una farnesil transferasa protohemo IX, junto con *ctaA*, codifican para enzimas requeridas para la biosíntesis de hemo A. Este es el grupo prostético para las citocromos aa3 (quinol oxidasa) y caa3 (citocromo oxidasa C). CtaB es la enzima responsable de la conversión del hemo B a hemo O, en tanto que CtaA participa en la conversión de hemo O a hemo A. En *B. subtilis*, la expresión de *ctaB* se incrementa al suplementar el medio con fuentes de carbono no fermentables, o bien en baja concentración de glucosa (1%). La represión catabólica depende de CcpA (Liu y Taber.,

1998), además requiere de los reguladores globales ResD y ResE, involucrados en la respiración aeróbica y anaeróbica (Zhang y Hulett., 2000).

El gen *dadX* codifica para una alanina racemasa. Las proteínas DadA y DadX son las responsables del catabolismo para alanina. En *R. leguminosarum*, el gen *dadX* se comparó a nivel transcripcional en diferentes fuentes de carbono, como succinato y glucosa, ambas en vida libre. Se observó que la expresión de *dadX* es 10 veces mayor en succinato, extrapolando el dato los autores indican la importancia del metabolismo de alanina en el crecimiento de *R. leguminosarum* utilizando ácidos dicarboxílicos como sustrato. En el mismo trabajo utilizando como fuente de carbono a piruvato, no se observó la inducción de *dadX* ni de *dadA*, y al comparar la expresión en bacteriodes respecto a vida libre, sólo en la comparación bacteroide-glucosa se detecta la expresión de *dadX*; en tanto, en la comparación succinato-bacteroide la expresión de este gen es mayor en vida libre, esto se debe a que tanto el gen *dadX* como los genes *dctA* y *pckA*, son fuertemente inducidos por ácidos dicarboxílicos (Karunakaran *et al.*, 2009).

El gen *eft* codifica para una flavoproteína para transferencia de electrones. Se ha determinado en simbiosis los productos de los genes *fixABCX* son los que participan en la transferencia de electrones. Estos genes también han sido descritos como indispensables para la fijación de nitrógeno. En el perfil 10 (18 y MM), detectamos la máxima expresión de *etfAch*; el producto de este gen también fue identificado mediante proteoma de bacteroides a los 18 ddi (Resendis-Antonio *et al.*, 2011).

El gen *ftsH* codifica para una metaloproteinasa involucrada en división celular. En *E. coli* esta proteína se requiere para modular la concentración de LpsX (una proteína de síntesis de lipopolisacáridos), aunque LpsX en alfaproteobacterias, como *A. tumefaciens* y *R. capsulatus*, también es degradada por la proteasa Lon (Langklotz *et al.*, 2011). En *R. etli*, en un trabajo sobre la alarmona (p)ppGpp, se detecta que ésta regula un grupo de genes que codifican para proteasas que incluye a *ftsH* y *lon* (Vercruysse *et al.*, 2011). Las proteasas contienen con el mal plegamiento y la desnaturalización de las proteínas; durante condiciones de estrés, éstas aumentan su expresión ya sea para prevenir la

agregación o bien para reciclar aminoácidos. El aumento de su expresión y de 4 peptidasas más en una mutante *rhs* de *R. leguminosarum* indican que más que un aumento en la actividad traduccional se debe a un defecto o disrupción de la homeóstasis de proteínas (Vercruyse *et al.*, 2011). En el perfil 3 (MM 6h) se localizan dos genes más que codifican para proteínas de división celular *ftsZ2* y *ftsW*, que indican que la expresión de este grupo de genes se presenta en bacterias que están en crecimiento y división constante; sin embargo, como se menciona al inicio del perfil, la máxima señal de los genes se presenta en la condición de medio mínimo.

El gen *gguC* inicialmente se caracterizó en *Agrobacterium* por su vecindad con el gen *chvE* cuyo producto tiene un papel indispensable en la virulencia de *Agrobacterium*. Los genes *gguABC* fueron propuestos para el transporte de glucosa y galactosa. Mutantes en estos genes no presentan un efecto en la virulencia de *Agrobacterium* sobre *Kalanchoe diargremontiana*; ni en el crecimiento en 1 mM de galactosa, glucosa, xilosa, ribosa, arabinosa, fucosa, o sacarosa; ni en la quimiotaxis en glucosa, galactosa, xilosa, ni arabinosa (Kemner *et al.*, 1997).

El gen *hrcT* codifica para una proteína de translocación del sistema de secreción tipo III. El sistema tipo III se ha analizado en diferentes modelos durante la simbiosis, obteniendo diferentes resultados, por ejemplo en *Sinorhizobium sp.* NGR234 la mutación en este sistema afecta la simbiosis de manera específica de acuerdo con el hospedero, ya que en *V. unguiculata* y *L. leucocephala* no tiene un efecto, en *Pachyrhizus tuberosus* tiene incluso un efecto benéfico y en un tercer grupo como en el caso de *Tephrosia vogelii* en donde se requiere para una simbiosis efectiva. En *B. japonicum* USDA110, *S. fredii* HH103 y USDA257, *M. loti* MAFF303099 y *R. etli* CNPAF512, mutantes en genes del sistema tipo III afectan la simbiosis de manera específica (Fauvert y Michiels, 2008).

Perfil 5: 18, 35 ddi y VL.

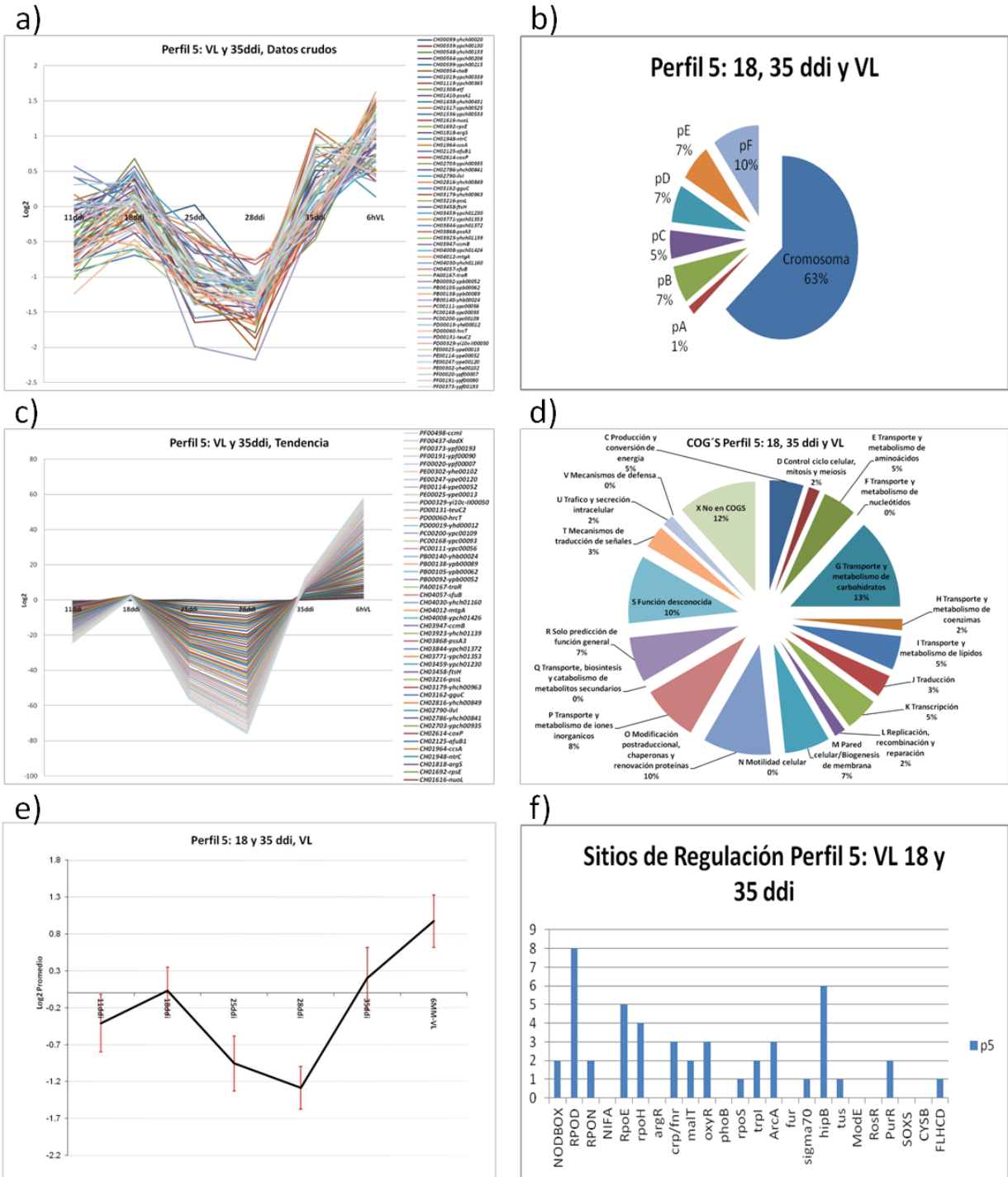


Figura 18. Perfil de expresión de los genes propios o expresados preferencialmente en vida libre (VL), 18 y 35 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *ilvI* codifica para una subunidad catalítica de acetolactato sintasa 3. En el perfil 2 también está presente el gen *ilvD2* (18 ddi) y en el perfil 4 el gen *ilvA* (35ddi). En *S. meliloti* una mutante en *ilvI* tiene una disminución en la eficiencia de la infección en alfalfa y al igual que la mutante *ilvD*, sólo producen algunos pseudonódulos (Pobigaylo *et al.*, 2008).

El gen *mtgA* codifica para una transglicosilasa biosintética monofuncional péptidoglicano. MtgA sintetiza cadenas lineares de glicano a partir de precursores unidos a lípidos y funciona durante la biosíntesis de péptidoglicano. En *Brucella abortus* la activación del gen *mtgA* es detectable a cierto nivel en macrófagos, pero no en cultivos de agar lo cual es incierto. Los autores explican que esto pudiera ser una respuesta a daño de la capa de péptidoglicanos de la bacteria en el fagosoma (Eskra *et al.*, 2001). Por otro lado, en *R. leguminosarum* y en *R. etli* (nódulos determinados), BacA es una proteína integral de membrana, y una mutación incrementa la resistencia a los antibióticos bleomicina y Bac7; además, aumenta la sensibilidad a SDS y a vancomicina. En nódulos indeterminados, una mutante en BacA después de ingresar por el hilo de infección es lisada y por lo tanto no llega a producir nódulos maduros. En ensayos de expresión de la mutante *bacA*, utilizando microarreglos, no se observan cambios respecto a la cepa silvestre esto en vida libre y en nódulos determinados; pero en bacteroides de nódulos indeterminados, los genes que codifican para componentes de membrana celular representan más de la mitad de los cambios de expresión entre la mutante *bacA* y la cepa silvestre. Entre los genes que codifican para enzimas que pueden afectar la estabilidad de la membrana y la estructura de los péptidoglicanos, se encuentra una fosfolipasa (RL2533) que reduce 3 veces su expresión y el gen que codifica para la péptidoglicano transglicosilasa (*mtgA*) que disminuye su expresión 5 veces (Karunakaran *et al.*, 2010).

El gen *ntrC* codifica para una proteína de regulación de nitrógeno (sistema de dos componentes). Si bien nuestros resultados agrupan la expresión de este gen con otros que se expresan en MM y a los 18 y 35 ddi, la máxima expresión de *ntrC* se observa a los 35 ddi (log 0.82), seguido del tiempo 18 ddi (log 0.47) y finalmente MM (log 0.13). A los 18 ddi *ntrY*, que también pertenece al sistema de dos componentes en la regulación de nitrógeno; alcanza su máxima expresión (log 1.1). NtrC funciona como un activador

transcripcional que modula la expresión de genes en respuesta de la concentración de nitrógeno, NtrB es una histidina cinasa que fosforila a NtrC bajo condiciones limitantes de nitrógeno y también puede funcionar como fosfatasa para desfosforilar NtrC en condiciones de exceso de nitrógeno. Una mutante en *ntrC* no afecta la expresión de los genes *fixNOQP*; adicionalmente, en *R. leguminosarum* una mutante en *ntrC* continúa utilizando nitrato como única fuente de nitrógeno (Moreno *et al.*, 1992) y en *S. meliloti* una mutante en *ntrC* no afecta la capacidad de fijación de nitrógeno (Szeto *et al.*, 1987).

El gen *nuoL* codifica para una NADH deshidrogenasa subunidad L. La bomba de protones NADH: ubiquinona oxidoreductasa complejo I de respiración, de la cual NuoL forma parte, acopla la transferencia de electrones de NADH a ubiquinona con la translocación de protones a través de la membrana. En *E. coli*, mediante la intervención de NuoN, el complejo está compuesto por 13 subunidades llamadas NuoA y contiene un FMN y 9 grupos hierro-sulfuro (Flemming *et al.*, 2005). El gen *nuoB*, codifica para la subunidad B de una NADH deshidrogenasa, en *E. coli* es regulada por O₂ y nitrato vía ArcA, NarL, FNR e IHF, incluso por DctD (Bongaerts *et al.*, 1995). El gen *nuoB2* presenta su máxima expresión a los 18 ddi (perfil 2). Los genes *nuoE1*, *nuoF1* y *nuoH1* presentan su máxima expresión en el perfil 3 (MM), en tanto que el gen *nuoA1* la tiene en el perfil 6 (11, 18 y 35 ddi) y el gen *nuoE2* en el perfil 8 (18, 35 ddi y MM).

El gen *pssA1* codifica para una CDP-diacilglicerol-serina O-fosfatidil transferasa o proteína de producción de exopolisacáridos. El gen *pssA3* participa en la producción de exopolisacáridos y el gen *pssL* codifica para una proteína de polimerización y/o exportación de exopolisacáridos. Los polisacáridos de superficie son de gran importancia para el establecimiento de la interacción planta-rhizobio. En *R. leguminosarum* el gen *pssA* codifica para una glucosil-transferasa, la cual lleva a cabo el primer paso en la biosíntesis de exopolisacáridos (EPS) al transferir una UDP-glucosa al lípido acarreador que se une a la membrana citoplasmática. En *R. leguminosarum* una mutante en *pssA* genera un fenotipo no mucoide y produce nódulos no fijadores en haba, trébol y arveja. El gen *pssL* codifica para una proteína que forma estructuras secundarias similares a las flipasas tipo Wzx, actuando específicamente en el desplazamiento del antígeno-O de la membrana interna

hacia la membrana externa de la célula (Skorupska *et al.*, 2006). El papel de los EPS puede ser más amplio de lo hasta ahora señalado, se propone que pudieran participar como moléculas de señalización durante la invasión, así como ser la base de los mecanismos de control para los sistemas de defensa de las plantas.

El gen *rpsE* codifica para una proteína ribosomal de la subunidad 5S (Wittmann-Liebold y Greuer, 1978). Debido al perfil donde se ubica (18, 35 ddi y MM), este dato podría indicar que los genes de la síntesis de proteínas en esta etapa inician su activación en los bacteroides; no obstante, la máxima expresión de este gen pertenece a VL-MM, al igual que la mayoría de los genes que codifican para proteínas ribosomales.

El gen *sfuB* codifica para una proteína tipo ABC para el transporte de hierro. El gen *sfuA* se localiza en el perfil 3 (MM), la expresión de *sfuB* alcanza su máximo valor en VL-MM. A diferencia de *sfuA*, *sfuB* presenta un valor de expresión positivo a los 18 ddi y un valor cercano a cero a los 35 ddi. La región de los genes *sfu* de *Serratia marcescens* confiere la capacidad de crecer en un medio limitado de hierro a la cepa *E. coli* K-12, la cual no es capaz de sintetizar sideróforos. El producto del gen *sfuB* es hidrofóbico, tiene el peso molecular de 56 kDa y es una proteína periplasmática que participa en mecanismos de transporte (Angerer *et al.*, 1990).

El gen *traR* codifica para un regulador transcripcional de los genes de conjugación, participa en el “quorum sensing” y funciona como dímero. En *A. tumefaciens* se demostró que es necesario para activar la resistencia a proteasas celulares, esto se determinó realizando ensayos de vida media de las proteínas *in vivo* e *in vitro* con o sin dímero, además con ensayos de pulso y caza etiquetando con radioisótopos (Pinto y Winans, 2009).

Los genes del perfil 5 nos indican que la bacteria está expresando a los 35 ddi genes que comparte con la vida libre donde alcanza su punto máximo y que además son genes de muy baja expresión a los 18 ddi. El grupo incluye genes relevantes para la protección y el mantenimiento de la bacteria en el nódulo, como los genes cuyos productos participan en la respiración, así como genes cuyos productos son necesarios para la utilización de

fuentes de carbono y nitrógeno. En este perfil destacan los genes cuyos productos participan en la biogénesis de membrana, como *mtgA*. Los COG's que incrementan su porcentaje en este perfil respecto al promedio de los 10 perfiles son: COG O-Modificación postraducciona, chaperonas y renovación de proteínas, que pasa del 4% al 10%; el COG M-Pared celular/Biogénesis de membrana que de un 4% presenta un 7%, el porcentaje más alto de los 10 perfiles; el COG P-Transporte y metabolismo de iones inorgánicos incrementa del 4% al 8%, también es el porcentaje más alto de los 10 perfiles; y por último el COG D-Control ciclo celular, mitosis y meiosis que incrementa del 1 al 2% (Fig. 18 inciso d). En este punto, los reguladores transcripcionales que destacan son HipB que se ha descrito como un regulador hélice vuelta hélice (HTH) tipo Cro (Black et al., 1994), involucrado en la regulación de genes para la síntesis de péptidoglicanos con el sitio consenso TATCCN8GGATA (en este perfil identificamos 6 blancos probables para este regulador) en tanto que los sigmas RpoE y RpoH con 5 y 4 blancos respectivamente, son los sobrerrepresentados en este perfil (Figuras 18 inciso f y fig. 25). En este perfil destaca TetR con 2 de los 3 reguladores identificados; los miembros de esta familia actúan como sensores químicos de la célula para monitorear el medio. Para llevar a cabo esta función, los miembros de la familia TetR protegen una diversidad de dominios unión-ligando capaces de reconocer la misma serie de compuestos como los transportadores que éstos regulan. Los reguladores TetR pueden inducir una amplia serie de compuestos estructuralmente no relacionados. La unión de éstos resulta en un cambio conformacional que es transmitido a la región de unión a DNA, causando que el represor pierda su capacidad de unión lo que permite el inicio de la transcripción. En *E. coli* y en *Campylobacter jejuni*, dos miembros de esta familia, AcrR y CmeR, regulan la expresión de genes que codifican para bombas de resistencia a drogas (Routh et al., 2009).

Perfil 6 (18 y 35ddi)

Este perfil contiene los genes expresados a los 18 y 35 ddi, y en menor medida a los 11 ddi (Fig. 19), la descripción de éstos es la siguiente:

El gen *aglG* codifica para un transportador de alfa-glucósidos tipo ABC, permeasa. La sacarosa es la fuente de carbono producida fotosintéticamente y transportada más abundantemente dentro de los nódulos en raíces de leguminosas. *R. eutropha*, por ejemplo, puede crecer en sacarosa, maltosa y trehalosa debido a sus *agl*'s (elementos para la utilización de alfa glucósidos). La mutación de los genes *agl* en *S. meliloti* sugirió que posee al menos un sistema de transporte adicional de baja afinidad para los alfa glucosidos. El fenotipo simbiótico es Fix+ y los autores proponen que es debido a una probable copia adicional (Willis and Walker *et al.*, 1999). *R. etli* no cuenta con copias adicionales del gen *aglG*.

El gen *asnP* codifica para una permeasa para L-asparagina. En una mutante *asn* de *S. meliloti* la introducción del operón *asnPAB* de *R. etli* permite a la bacteria crecer en L-asparagina como única fuente de carbono y nitrógeno, sugiriendo que el operón *asn* juega el mismo papel fisiológico en ambas bacterias (Ortuño-Olea y Durán-Vargas, 2000).

El gen *argG* codifica para una proteína arginina-succinato sintetasa. En *S. meliloti* se obtuvieron 78 mutantes auxótrofos de arginina, las cuales fueron clasificadas como auxótrofos de ornitina (*argA/argB/argC/argD/argE*) y auxótrofos de arginina (*argF/argI, argG* y *argH*). Las mutantes auxótrofos de ornitina producen nódulos inefectivos en plantas de alfalfa, en tanto que las auxótrofos de arginina inducen nódulos efectivos (Kumar *et al.*, 2003).

El gen *citEch* codifica para una citrato liasa subunidad beta. Los dominios de CitE tienen una fuerte similitud con liasas malil-CoA y en menor medida con las malato sintasas (Miller *et al.*, 2007). En *Riftia pachyptila*, simbionte facultativo no cultivable, mediante metagenómica reportaron que este simbionte utiliza sólo una parte del ciclo Calvin-Benson para la fijación de carbono y un ciclo reverso para el ciclo del TCA, que es una forma alternativa para la fijación de carbono, para lo cual utiliza una inusual ATP citrato liasa (Robidart *et al.*, 2008).

Perfil 6: 11, 18 y 35 ddi.

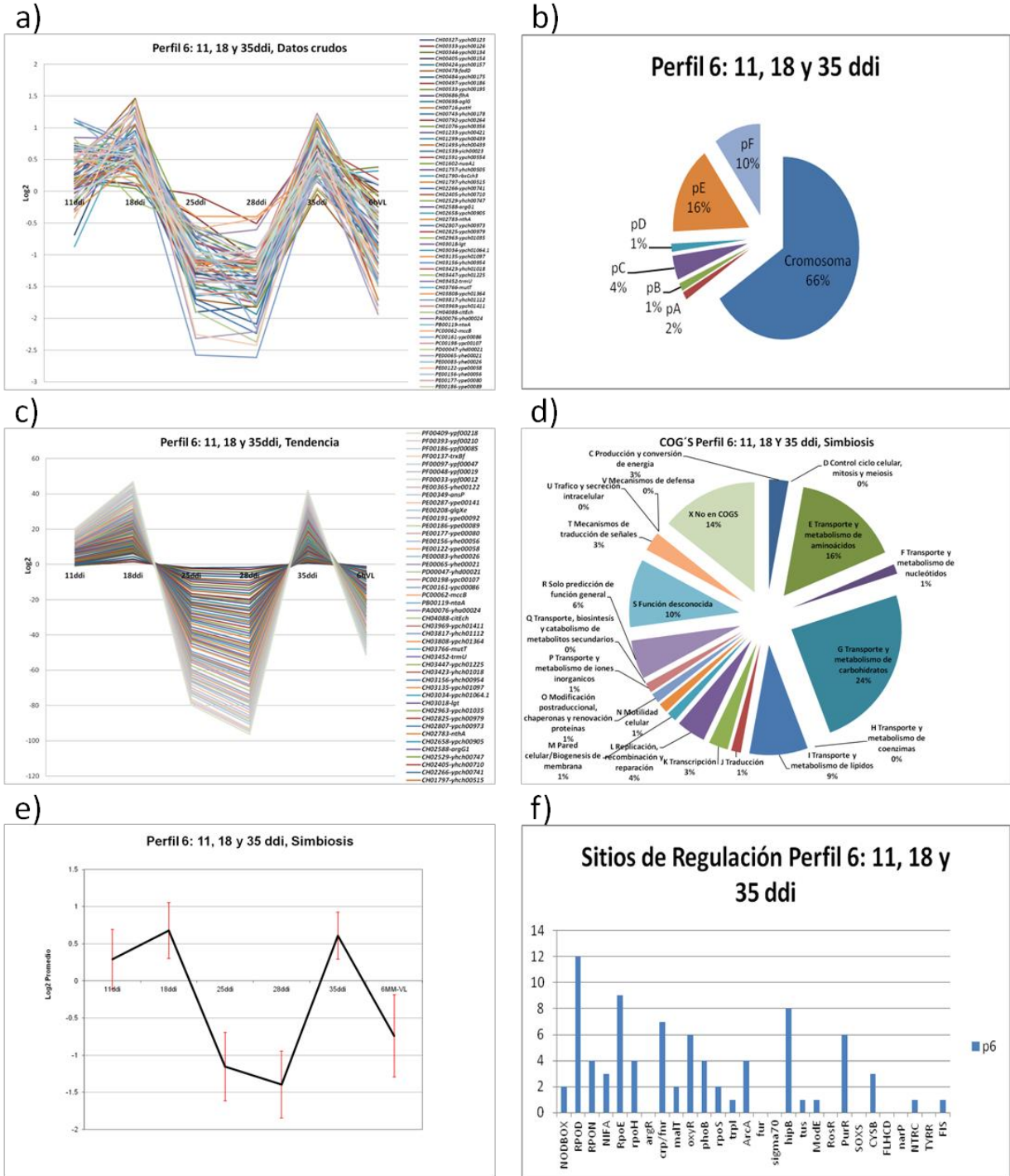


Figura 19. Perfil de expresión de los genes propios o expresados preferencialmente a los 11, 18 y 35, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *fadD* codifica para una acil-CoA de cadena larga sintetasa. El “swarming” es un fenómeno multicelular caracterizado coordinado por un rápido movimiento de las bacterias a lo largo de superficies semisólidas. La mutante *fadD* en *S. meliloti* tiene un fenotipo en el “swarming”, las mutantes son hiperflageladas y el doble de grandes que las células silvestres. La mutante es incapaz de establecer la simbiosis y al complementar con un plásmido multicopia se restaura la simbiosis así como el crecimiento en oleato y el control sobre la motilidad. Además, la mutante presenta una disminución en la expresión de genes de nodulación así como una expresión diferencial en los genes de motilidad en respuesta a diferentes condiciones (Soto *et al.*, 2002). En *S. meliloti* también se caracterizó mediante transcriptoma a la mutante *fadD*, las condiciones fueron en “swarming” y en No-“swarming”, los grupos de genes identificados como diferencialmente regulados fueron absorción de hierro, quimiotaxis, motilidad y relacionados a estrés (Nogales *et al.*, 2010).

El gen *flhA* codifica para una proteína involucrada en la biosíntesis de flagelo. Muchas de las proteínas flagelares son exportadas a través de un canal de estructura flagelar central llamado aparato de exportación de proteínas flagelares tipo III. Las proteínas FlhA y FlhB son proteínas de membrana que forman parte integral de esta estructura; éstas forman una plataforma de muelle para el paso de los componentes solubles del aparato de exportación, FliH, FliI y FliJ. Mutantes puntuales en el gen *flhA* indican que además de formar parte de la plataforma del muelle, FlhA también está involucrada en la translocación del sustrato de exportación dentro del canal central del aparato de exportación flagelar (Minamino *et al.*, 2010). En *S. meliloti* en estrés osmótico, el gen *flhA* junto con otros genes de quimiotaxis y de motilidad celular se observaron reprimidos (Domínguez-Ferreras *et al.*, 2006).

El gen *lgt* codifica para una prolipoproteína diacilglicerol transferasa. En algunas bacterias Gram-positivas las lipoproteínas juegan un papel crucial durante la virulencia. En *B. anthracis* una mutante en el gen *lgt* carece de lipoproteínas y se observa una disminución superficial de la hidrofobicidad. Asimismo las esporas, germinan ineficientemente tanto *in*

vitro como en la piel de ratón, por lo que la virulencia es atenuada. No obstante, las células vegetativas de la mutante fueron tan virulentas como la silvestre (Okugawa *et al.*, 2012).

El gen *mccB* codifica para una metilcrotonoil-CoA carboxilasa subunidad biotinilada y participa en el metabolismo y transporte de lípidos. En *R. etli* MccB fue identificada en simbiosis mediante proteómica. Los autores sugieren que el metabolismo de lípidos juegan un papel importante en el bacteroide, proveyendo de una variedad de precursores que serían componentes de membrana, Lps y coenzimas requeridos para la traducción de señales (Resendis-Antonio *et al.*, 2011). En otros modelos existe la evidencia de que estos precursores son suplementados por la planta hospedera (Braunstein *et al.*, 2001).

El gen *mutT* codifica para una hiruermutasa. En el perfil 4 (35ddi) existe otra proteína hipotética con el mismo dominio de MutT/NUDIX (CH02122), incluso en este mismo perfil se localiza el gen CH02266 que también posee un dominio MutT/NUDIX. En *E. coli*, MutT hidroliza a 8-oxodGTP evitando la incorporación de éste durante la síntesis de DNA. Otros genes de este tipo pueden estar involucrados en el saneamiento de las pozas de nucleótidos; la existencia de varias proteínas de este tipo es común, por ejemplo, en *D. radiodurans* se han identificado 22 proteínas de este tipo, en *C. crescentus* 11 (Martins-Pinheiro *et al.*, 2007), en tanto que *R. etli* cuenta con 11 genes que presentan el dominio MutT/NUDIX y un gen *mutT*.

El gen *ntaA* codifica para una nitrilo triacetato monooxigenasa componente A. El nitrilo triacetato (NTA) es un agente quelante que interactúa con metales pesados. NtaA es la primera enzima que participa en la biodegradación de NTA, el cual es fácil de degradar y varios organismos la utilizan como única fuente de carbono y nitrógeno (Xu *et al.*, 1997).

El gen *nthA* codifica para una proteína nitrilo hidratasa. Las nitrilasas catalizan la hidrólisis de componentes de nitrilo a su correspondiente ácido carboxílico y amonio. Aunque las nitrilasas están presentes en un amplio rango de organismos que incluye bacterias, hongos y plantas, su importancia biológica ha sido poco explorada. En el contexto de una

interacción planta-bacteria, las nitrilasas podrían tener un papel en la síntesis de hormonas, asimilación de nutrientes y detoxificación de nitrilos exógenos y endógenos. Las nitrilasas son producidas tanto por las plantas que estimulan el crecimiento en presencia de microorganismos así como de patógenos, y su actividad podría tener un impacto significativo sobre la interacción planta-bacteria (Howden y Preston, 2009).

El gen *nuoA1* codifica para una NADH deshidrogenasa subunidad A. Ésta forma parte de la bomba de protones NADH: ubiquinona oxidoreductasa, complejo I de respiración y acopla la transferencia de electrones de NADH a ubiquinona con la translocación de protones a través de la membrana. En *E. coli* el complejo está constituido por 13 subunidades llamadas NuoA-NuoN y contiene un FMN y 9 grupos hierro-sulfuro (Flemming *et al.*, 2005).

El gen *potH* codifica para una permeasa de espermidina/putrescina, un transportador tipo ABC. En *S. meliloti* mutantes en *aceA* y en SMc00767 (un represor involucrado en el metabolismo del acetato) tienen un pobre crecimiento en acetato, al realizar el transcriptoma de la mutante SMc000767, *potH* es uno de los 7 genes sobreexpresados incluyendo *aceA*. Los autores de la investigación sugieren que estos genes podrían participar en la utilización de acetato como fuente de carbono (Ramírez-Trujillo *et al.*, 2007).

El gen *rbsCch3* codifica para la proteína transmembranal de un transportador de ribosa tipo ABC. En el perfil 4 (35ddi) observamos al gen *rbsCch2*. En algunas rhizobiaceas, como *M. loti* cepa NZP2037, se ha observado que son capaces de utilizar ribosa como fuente de carbono en un medio llamado S-30 (Scott *et al.*, 1987).

El gen *trmU* codifica para una tRNA metiltransferasa-(5-metilaminometil-2-tiouridilato). En *S. enterica* la cisteína desulfurasa (IscS) es requerida para la formación de todos los nucleósidos (4) tiolados en tRNAs, esto ocurre principalmente de 2 maneras, y ambas requieren a IscS. En *R. etli* el gen *iscN* del plásmido simbiótico es regulado por NifA-RpoN (Dombrecht *et al.*, 2002). La síntesis de 4-tiouridina (s(4)U) y 5-metilaminometil-2-

tiouridilato ocurre por una transferencia del sulfuro de IscS por varias proteínas para el blanco del nucleósido en el tRNA. De acuerdo con esto, la inactivación de IscS debería disminuir la síntesis de los nucleosidos tiolados. Los autores sugieren que IscS tiene un sustrato intrínseco específico el cual media la movilización de sulfuro y/o la formación y mantenimiento de un grupo hierro-sulfuro requerido para la tiolación del tRNA (Lundgren and Björk, 2006).

El gen *trxBf* codifica para una reductasa tiorredoxina (NADPH). TrxB al igual que KatA, AhpCF y Bcp, confieren resistencia al estrés oxidativo. Excepto KatA, en *Staphylococcus aureus* también son requeridas para la virulencia y además todas son reguladas por el mismo factor transcripcional, PerR (Horsburgh *et al.*, 2001). Las tiorredoxinas en plantas constituyen una familia de pequeñas proteínas. En *M. truncatula*, localizaron dos isoformas para Trx específicas de leguminosas; además, su expresión es durante la simbiosis y las nombran trxs's simbióticas (Alkhalifioui *et al.*, 2008).

En este perfil destacan los genes que codifican para proteínas involucradas en la asimilación de varias fuentes de carbono y en menor medida en la síntesis y transporte de aminoácidos. Los COG's dominantes son: el COG G-Transporte y metabolismo de carbohidratos, que se incrementa de un 10% (promedio de todos los perfiles), a un 24% en este perfil, siendo el máximo porcentaje para este COG en los 10 perfiles. El COG E-Transporte y metabolismo de aminoácidos, que tiene un promedio del 8% en los 10 perfiles alcanza el 16% en este perfil, también el máximo porcentaje de los 10 perfiles. El COG I-Transporte y metabolismo de lípidos, incrementa de un 4% promedio de todos los perfiles al 8.5% en este perfil, siendo el porcentaje más alto de los 10 perfiles. Por otro lado, varios COG's sufren una drástica disminución en este perfil, por ejemplo, el COG J-Traducción tiene una caída de un 5% en los 10 perfiles a cerca del 1%; el COG O-Modificación postraducciona, chaperonas y renovación disminuye del 4% al 1% y el COG H-Transporte y metabolismo de coenzimas pasa del 1% en todos los perfiles a 0% en este perfil (Fig. 20, inciso d). En cuanto a regulación, en este perfil se localiza un regulador transcripcional con ID CH01076 y un miembro de la familia LysR con ID CH000344. El

factor sigma que destaca es RpoE con 9 blancos presentes en las regiones regulatorias de los genes de este perfil, además detectamos sitios de unión para FNR (7 sitios) y para PurR que alcanza su máximo número de elementos en los 10 perfiles analizados, con 6 blancos, los genes *purUe*, *purUch* y *purC1* se localizan en los perfiles 1 (11 y 18 ddi) y 2 (18 ddi). La ausencia de estos genes en este perfil y la presencia de otros con blancos de PurR se puede explicar a través del regulón PurR de *E. coli*, que además de regular los genes del metabolismo de las purinas tiene un efecto en un 10 % de los genes de esta bacteria (Cho *et al.*, 2011).

Perfil 7 (25 ddi y MM)

En este perfil se observaron los genes expresados tanto en vida libre como en simbiosis a los 25 ddi, destacamos los siguientes:

El gen *cspA* codifica para una pequeña proteína que se induce por choque de frío, es decir, el RNAm de 428 nucleótidos incrementa su abundancia y vida media en estas condiciones de estrés. Observaciones *in vitro* del RNA mensajero indican que éste puede unirse a múltiples copias de Hfq (proteína hexamérica que promueve una variedad de interacciones RNA-RNA). La presencia de Hfq sobre el RNA mensajero inhibe casi en su totalidad el corte de la RNasa E en un sitio cercano a los 35 nt del 3' del RNA (Hankins *et al.*, 2010). Respecto a la estabilidad del RNA mensajero y su presencia a los 25 ddi, la cual coincide con la máxima expresión de los genes *nif* y *fix* (ver perfil 9), se ha reportado que la degradación del RNA mensajero de los genes *nif* es menor cuando la bacteria fija nitrógeno. En *K. pneumoniae* en condiciones favorables de fijación de nitrógeno la vida media de los mensajeros *nif* va de los 20 a los 30 min; en cambio, la estabilidad del mensajero disminuye en un factor de 3 a 5 veces cuando se incrementa la concentración de oxígeno o se aumenta la temperatura a 37°C (Simon *et al.*, 1989).

El gen *dctA* codifica para un transportador de dicarboxilatos-C4. Durante la simbiosis el transporte de dicarboxilato juega un importante papel en la operación y efectividad de la interacción. Existe suficiente evidencia de que este tipo de compuestos son las mayores

fuerza de energía y carbono durante la fijación de nitrógeno en rizobias (Yurgel y Kahn, 2004). La proteína DctA pertenece a un subgrupo de transportadores de glutamato. La mutante en *S. meliloti* presenta un fenotipo Fix- porque una gran parte del nódulo es senescente. Su región promotora es parecida a la de los genes *nif* y presenta un probable sitio de unión a NifA (Engelke *et al.*, 1989). En nuestro análisis de búsqueda de sitios de regulación, localizamos en la región promotora de este gen un posible sitio de unión a sigma 54 y un probable sitio para OxyR.

El gen *fhuB* codifica para una permeasa de hierro-hidroxiato. En el perfil 2 (18 ddi) se localiza el gen *fhuD*. En *R. leguminosarum*, al igual que FhuD, FhuB participa en la toma de sideróforos como la vicibactina y también se localiza en membrana. La mutante en el operón *fhuDEF* aparentemente no presenta fenotipo simbiótico y se regula por la concentración de Fe, a mayor concentración mayor expresión (Stevens *et al.*, 1999). En el perfil 2 y 8 (18 ddi) se localizan dos marcos de lectura abierta identificadas con los ID's PF00270 y PF00269 que codifican para proteínas hipotéticas, las cuales han sido anotadas como transportadores tipo ABC de hierro.

El gen *fixNd* codifica para una oxidasa citocromo C. La expresión de los genes *fixN*, *fixG* y *arcA* de *R. etli* CNPAF512 es regulada estrictamente por FnrR bajo condiciones microaeróbicas y durante la simbiosis (bacteroides). Sin embargo, se ha propuesto una regulación adicional para *fixN* y *fixG* durante la simbiosis (Moris *et al.*, 2004). En *R. etli* CFN42, una doble mutante del *fnrN* de cromosoma y *fnrN* del plásmido D, demuestra que los productos de ambos genes participan en conjunto con FixL-FixKf, en la regulación (inducción) durante la microaerobiosis de los genes *fixNOQPd*. Durante la simbiosis con *P. vulgaris* la mutante sencilla o la doble también tienen un efecto y se reduce la fijación de nitrógeno (Lopez *et al.*, 2001).

El gen *grlA* codifica para una glutarredoxina, su función es contener contra el estrés oxidativo. En *R. etli* se regula positivamente por (p)ppGpp (Vercruyssen *et al.*, 2011). En *S. meliloti* se caracterizaron las 3 glutarredoxinas; la mutante en SmGRX3 no presenta fenotipo alguno, en tanto que las mutantes en SmGRX1 y SmGRX2 reducen el crecimiento

de la bacteria y tienen un efecto adverso sobre la fijación de nitrógeno. Las características de la mutante en SmGRX1 son nódulos abortivos y el rhizobio no presenta diferenciación a bacteroide. En tanto, la mutante en SmGRX2, tiene un efecto negativo en el desarrollo del nódulo y no presenta la modificación en el desarrollo del bacteroide, lo que podría indicar la importancia de este tipo de enzimas en la diferenciación de la bacteria a bacteroide (Benyamina *et al.*, 2012).

En este perfil se observaron los genes expresados en vida libre y simbiosis (25 ddi). Los genes codifican para cochaperonas de GroEL. Los genes *groEL* y *groES* se observan en 3 diferentes perfiles (perfil 3: MM, perfil 7: 25 ddi y MM y perfil 10: 18 ddi y MM), pero en todos los casos anteriores la señal máxima de expresión se presenta en MM. Las chaperonas están involucradas en el plegamiento correcto de las proteínas; además en el nódulo pudieran participar en el plegamiento de la nitrogenasa, en contender contra el estrés, en el ensamble correcto de los reguladores transcripcionales al DNA y en la translocación de proteínas entra la bacteria y la planta. En *B. japonicum* se identificó que los genes *groESL3* se co-regulan con los genes de fijación, lo que implica su probable participación en el nódulo, ya que además una mutante en *groEL* en *B. japonicum* la nodulación y la fijación de nitrógeno disminuyen a pesar de tener 4 copias adicionales de *groEL* (Fischer *et al.*, 1993). El genoma de *S. meliloti* posee 5 genes *groEL* y sólo la mutante en *groEL1* presenta un fenotipo simbiótico. En esta cepa se observa que no puede inducir completamente la expresión de los genes *nod* y forma nódulos inefectivos lo que indica que *groEL1* es fundamental para la simbiosis; no obstante, en la cepa silvestre se ha detectado en simbiosis la expresión de las otras copias de los genes *groELS* (Bittner *et al.*, 2007).

El gen *nolR* codifica para un regulador transcripcional de la familia ArsR, el cual es un regulador de los los genes de nodulación. Una mutante *nolR* en *R. leguminosarum* cepa MR114 (la cual carece del plásmido simbiótico) presenta una expresión constitutiva de los genes *nodA*, *nodF* y *nodM* (Li *et al.*, 2008). En *S. meliloti*, la mutante *nolR* nodula aunque menos eficiente que la cepa silvestre pero no afecta la fijación de nitrógeno. La expresión

del gen *nolR* es dependiente de la densidad celular e influenciada por diversos estímulos que incluyen limitación de nutrientes, pH y oxígeno. Bajo condiciones de laboratorio, *NolR* se requiere para un óptimo crecimiento en medio sólido, para la óptima sobrevivencia en un medio mínimo deficiente en carbono y para después de un reto por choque térmico (Chen *et al.*, 2005). En *R. etli* CFN42, al analizar el transcriptoma en simbiosis de la cepa silvestre comparada con la mutante *nifA*, no se observó la expresión del gen *nodA*, aun cuando en la región promotora presenta los sitios de unión para RpoN y NifA. También presenta un sitio de unión para *NolR*, lo que podría sugerir que esta última regula de manera negativa la expresión de *nodA*. La expresión del gen *nolRch* se observa tanto en la cepa silvestre como en la mutante *nifA* a los 11ddi (Salazar *et al.*, 2010). El gen *nolRch* tiene una copia extra en el plásmido F, *nolRf*, del cual no observamos expresión en simbiosis ni en MM.

El gen *nrdE* codifica para una ribonucleótido-difosfato reductasa subunidad alfa. La vitamina B(12) o cobalamina es un cofactor crítico para animales y protistas, pero la biosíntesis está limitada a procariotes. Para el establecimiento de la relación simbiótica entre *S. meliloti* y *M. sativa* (alfalfa) la bacteria requiere de cobalamina. Investigando la función de tres enzimas dependientes de cobalamina, entre ellas *nrdJ* que codifica para una ribonucleótido reductasa dependiente de cobalamina (RNR), al evaluar una cepa de *S. meliloti* que sobreexpresa una RNR independiente de cobalamina (clase Ia) de *E. coli*, se observó que ésta se ve afectada en simbiosis. Mediante micrografías se observa que las bacterias de esta cepa pueden inducir nódulos pero que son incapaces de diferenciarse a bacteroides y en cierto momento son lisadas en el citoplasma de la planta. Este defecto puede deberse a la inactivación de la RNR por especies reactivas de oxígeno o bien por una inadecuada habilidad de contender con el oxígeno en el nódulo, o una combinación de ambas (Taga y Walker, 2010).

Perfil 7: 25, 28 ddi y VL.



Figura 20. Perfil de expresión de los genes propios o expresados preferencialmente en vida libre (VL) y 25, 28 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución COG'S, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *pdhA2* codifica para piruvato deshidrogenasa subunidad beta. El gen *pdhA1* alcanza su máxima expresión en el perfil 3 (MM). Por otro lado, PdhA2 fue identificada en *R. etli* en simbiosis mediante proteómica (Resendis-Antonio *et al.*, 2011). El complejo de la piruvato deshidrogenasa cataliza la decarboxilación oxidativa del piruvato para producir acetil-CoA, lo que une la glicólisis con el ciclo de Krebs. En *S. meliloti* observaron que *pdhA* tiene una mayor expresión en bacteroides de 3 semanas después de inocular alfalfa, comparando con la expresión en vida libre en condiciones aeróbicas, (Cabanes *et al.*, 2000). En *S. meliloti*, buscando mutantes con la actividad de PDH disminuida, identificaron a la cepa *S. meliloti* PD3; la caracterización de la mutante revela que la inserción del Tn5 se localiza en el 3' del gen *ada* cuyo producto es una arilesterasa. Dicha mutante presenta un fenotipo simbiótico al reducir la fijación de nitrógeno, lo que lleva a los autores a concluir que la actividad de la PDH es importante para la simbiosis entre *S. meliloti* y alfalfa (Soto *et al.*, 2001).

El gen *ropA* codifica para una proteína de membrana externa conocida como porina. De hecho, los genes *ropAch1* y *ropAch2* se localizan en este perfil. En total *R. etli* CFN42 tiene 4 genes *ropa*: 3 en cromosoma y 1 en el pE. En *R. leguminosarum* bv. *viciae* se localizaron dos *ropA* de tamaño (gen) y peso molecular (proteína) diferente, con una identidad a nivel de aa entre ambos de 91.8%. Las regiones promotoras contienen un sitio de unión a IHF, sin embargo, no presentaron datos de su posible función (Roest *et al.*, 1995). En un artículo reciente, los autores retoman el estudio de los genes *rop* en *R. leguminosarum*, los renombran como *ropB*, *ropB2* y *ropB3*, cada miembro fue mutagenizado y ninguno presenta fenotipo simbiótico; además, su expresión en simbiosis fue insignificante. En cuanto a vida libre la función de los genes *ropB2* y *ropB3*, aún es criptica, a diferencia de la mutante *ropB*, la cual es más sensible a los detergentes, antibióticos hidrofóbicos y ácidos débiles. Por lo tanto sugieren que la función de RopB es dar estabilidad a la membrana externa (Foreman *et al.*, 2010).

El gen *rosR* codifica para un regulador transcripcional. En *R. leguminosarum* bv. *Trifolii*, RosR está involucrado en la regulación positiva en la síntesis de los EPS. Un análisis de la región promotora revela varios sitios inversos repetidos y motivos que pudieran ser

reconocidos por PhoB y por LysR, asociados con la dependencia de fosfato y de flavonoides, respectivamente. Al probar exudados de raíces de trébol y diferentes concentraciones de fosfato, la expresión de *rosR* se incrementa lo que indica la dependencia de ambos sitios. Sin embargo, la formación de *biofilms* y la síntesis de exopolisacáridos disminuyen cuando aumenta la concentración de fosfato y resulta contrario al efecto de la presencia de exudados (aumentan la formación de *biofilms* y la síntesis de EPS), lo que indica una compleja regulación (Janczarek y Skorupska, 2011). La mutante *rosR* exhibe un fenotipo pleiotrópico, produce tres veces menos EPS que la cepa silvestre, es más sensible a los detergentes, a antibióticos del grupo beta-lactámicos y algunos osmolitos, lo que indica un cambio en la membrana. Además, exhibe una reducción en la formación de *biofilm* sobre superficies plásticas, el cual difiere en la arquitectura, profundidad y viabilidad respecto a la cepa silvestre. La mutante tiene una considerable disminución en la adherencia y colonización de las raíces, por lo que los autores proponen su participación en las etapas tempranas de la invasión. La mutante sí forma nódulos no fijadores en trébol pero no compiten con la cepa silvestre pues son desplazados por esta (Janczarek *et al.*, 2010).

El gen *sdhC* codifica para una succinato deshidrogenasa, citocromo subunidad B-556. La importancia de los ácidos dicarboxílicos en la simbiosis ha sido ampliamente documentada. Tanto el succinato, fumarato, malato y otros son utilizados como fuentes de carbono y energía. En *Bradyrhizobium*, mutantes que pierden la capacidad de tomar C4-dicarboxilatos desarrollan nódulos que no son capaces de fijar nitrógeno, lo que demuestra el papel fundamental de estos compuestos en la simbiosis. La Sdh cataliza la oxidación de succinato a fumarato y esta reacción está acoplada a la reducción de ubiquinona a ubiquinol; en particular, SdhC funciona para fijar la subunidad catalítica en la parte interna de la membrana citoplasmática (Westenberg y Guerinot, 1999).

El gen *virB5* codifica para una proteína del sistema de secreción tipo IV (T4SS). En *A. tumefaciens* se tiene el antecedente que el *pilus-T* está formado por VirB2 como componente mayor y uno menor llamado VirB5. Se ha hipotetizado que el *pilus-T* está

involucrado en la especificidad al hospedero. Los cambios en el carboxilo terminal de VirB5, ya sea una delección de 4 aa o bien una sustitución de 4 aa (también en el carboxilo terminal) por alanina, tienen un impacto en el tamaño del *pilus-T*; estos cambios también afectan la habilidad del sistema T4SS para transferir el DNA a la planta y bacterias receptoras. Mediante inmunolocalización de VirB5 lo ubican en la punta del *pili* lo que sugiere un papel en el reconocimiento al hospedero (Aly y Baron, 2007). En el análisis transcriptómico de *R. etli* en nódulos de 21 ddi, el gen *virB5a* se expresa junto con *virB8a* y *virB7a* (Vercruyssen *et al.*, 2011). En los perfiles descritos en esta tesis el gen *virB9* se localiza en el perfil 10 (18 ddi y VL).

Este perfil contiene genes que codifican para reguladores transcripcionales, como NolR y RosR. El primero participa reprimiendo genes de nodulación, por el contrario RosR actúa de manera positiva en genes que están involucrados en la síntesis de exopolisacáridos (EPS). También se localizan en este perfil genes cuyos productos participan en la producción de energía y en la asimilación de carbono se localizan en este perfil. Destacan en este perfil genes que son claves para la producción de energía y el transporte de C4-dicarboxilatos, como lo son *sdhC*, *pdhA* y *dctA*. Los COG's sobrerrepresentados son: COG X-No en COGS que representa el 47% del perfil, el COG J-Traducción, que de un porcentaje del 5% en los 10 perfiles y se incrementa al 10% en este perfil; además, del COG K-Transcripción que llega al 10% cuando el promedio es de 6% en los 10 perfiles. El COG L-Replicación, recombinación y reparación no está representado por ningún elemento en este perfil, en el porcentaje de los 10 perfiles es de 3%. El COG que presenta el cambio más drástico es el E-Transporte y metabolismo de aminoácidos, que pasa de un 8% en los 10 perfiles a un 0% en este perfil; es decir, este COG no está representado por un sólo elemento. Lo mismo sucede con el COG I-transporte y metabolismo de lípidos, que pasa de un 4% a 0% (Fig. 21, inciso d). El factor sigma dominante es RpoE con 11 sitios, RpoN con 8 y NifA con 7. La presencia de RpoN-NifA sólo se da en la región promotora del gen *rosR*. En este perfil, el ORF con ID CH04025 representa a un miembro de la familia CarD. En *M. tuberculosis* CarD es un regulador global, cuya carencia produce un defecto en la

transcripción del RNA ribosomal; la mutante es sensible a estrés oxidativo, inanición y daño al DNA (Stallings *et al.*, 2009).

Perfil 8 (18 ddi)

En este perfil, al igual que el perfil 2, se ubican los genes expresados a los 18 ddi. A diferencia del perfil 2, donde los genes expresados en vida libre presentan el punto con mayor represión, en este perfil los genes expresados también en vida libre presentan el segundo punto mayor de expresión. Destacamos los siguientes:

El gen *afuB2* codifica para una permeasa transportador de hierro (III) tipo ABC. En *Pseudomonas fluorescens* HV37 las proteínas AfuA y AfuB son determinantes para la producción de antibióticos regulados de manera transcripcional por la presencia o ausencia de glucosa (Gutterson *et al.*, 1988). Las proteínas Afu se localizan en membrana y se han definido como marcadores en el proceso de infección de la brucelosis en alces de la región de Wyoming (Lowry *et al.*, 2010). En *Actinobacillus actinomycetemcomitans* el gen *afuA* es regulado por FurR (Haraszthy, *et al.*, 2006). Los genes *afu* se expresan en el perfil 5 (35 y MM), perfil 2 (18ddi) y perfil 8 (18,35 ddi y en MM).

El gen *aptch* codifica para una adenina fosforibosil transferasa. El gen *aptd* se expresa en el perfil 9 (25 ddi) y es parte del regulón NifA-RpoN (Salazar *et al.*, 2010). La región promotora de *aptch* también contiene ambos sitios de pegado (NifA y RpoN). La función del Apt es catalizar una reacción de salvamento que da por resultado la formación de AMP a partir de Adenina y 5-fosfo-alfa-D-ribosa 1-difosfato, siendo esto menos costoso energéticamente que la síntesis *de novo* (Hershey *et al.*, 1982).

El gen *cyoA* codifica para una citocromo O ubiquinol oxidasa, subunidad II. Este tipo de citocromos se expresa bajo condiciones microaeróbicas y se localiza en el pE. La mutante *cyoA* de *R. etli* CE3 crece adecuadamente en medio rico y medio mínimo en condiciones aeróbicas, pero sufre una severa reducción en el crecimiento en medio mínimo en condiciones microaeróbicas (Landeta *et al.*, 2011).

El gen *dapE1* codifica para una succinil-diaminopimelato desuccinilasa. En el perfil 2 (18 ddi) se expresan otros 2 genes *dap*: el *dapB2* y *dapA1*. El gen *dapF* se localiza en MM (perfil 3). El producto de *dapE1* se detectó e identificó mediante proteómica en simbiosis (Resendis-Antonio *et al.*, 2011). DapE1, así como DapB2, participan en la vía de biosíntesis de lisina.

Los genes *fabI1* y *fabI2* codifican para enoilil-(proteína acarreadora de acilo) reductasas y el gen *fabH* codifica para una 3-oxoacil-(proteína acarreadora de acilo) sintasa III. El gen que codifica para la 3-oxoacil-(proteína acarreadora de acilo) sintasa II (*fabF*) se localiza en el perfil 2 (18ddi). A pesar de estar agrupados en el perfil 8 (18, 35 y MM), los genes *fabI1* y *2*, así como *fabH*, tienen su pico máximo de expresión a los 18 ddi al igual que el gen *fabF*.

El gen *fbaB* codifica para una fructosa-bisfosfato aldolasa. Esta enzima participa en la glicólisis y gluconeogénesis. Una mutación en el único gen que codifica para FbaB de *X. oryzae*, agente causal de la infección bacteriana roya de la hoja, provoca que el patógeno sea incapaz de usar piruvato y malato para crecer y disminuye su crecimiento cuando se utiliza la fructosa como única fuente de carbono. Adicionalmente, reduce la producción de exopolisacáridos (EPS) y deteriora la virulencia y crecimiento del patógeno en el arroz. La expresión de *fbaB* es mayor en la planta que en un medio rico (Guo *et al.*, 2012). La proteína FbaB fue previamente detectada expresada en la simbiosis de *R. etli-P. vulgaris* (Resendis-Antonio *et al.*, 2011).

El gen *fdsB* codifica para la proteína formiato deshidrogenasa subunidad beta. El gen *fdsD* presenta su máxima expresión en el perfil 4 (35ddi). Las formiato deshidrogenasas oxidan formiato a CO₂ y esto puede contribuir con electrones para la cadena respiratoria. En *R. sphaeroides* un estrés por peróxido induce dos subunidades de *fds* (Zeller *et al.*, 2005).

El gen *flaCch4* codifica para una flagelina C. La quimiotaxis y la motilidad son importantes en los primeros eventos de la nodulación, pero los mecanismos de cómo se llevan a cabo han sido poco estudiados. Las cepas de *R. leguminosarum* 3891 y VF39SM cuentan con 7 flagelinas; de éstas, la mutante de *flaA* tiene un fenotipo no mótil para la cepa VF39SM y

muy poca motilidad en la cepa 3841. Las mutantes individuales de *flaC* y *flaB* presentan filamentos flagelares más cortos y con ello reducen el “swimming” y el “swarming” (Tambalo *et al.*, 2010). En el perfil 3 (MM), el gen *flaCch1* alcanza su máxima expresión.

El gen *gmk* codifica para una guanilato cinasa (GMP cinasa), la cual es una enzima esencial para la biosíntesis de GTP y dGTP. Para la conversión de IMP a GTP o ATP se requieren cuatro enzimas; entre estas la guanilato cinasa (*gmk*). En *H. pylori* a pesar de existir los genes, no fue detectada la actividad de la guanilato cinasa en lisados de bacterias; por esta razón, los autores sugieren la presencia de una vía alternativa (Marais *et al.*, 1999).

En este perfil el gen *idhA* codifica para una mio-inositol2-deshidrogenasa. El inositol y los compuestos derivados de éste proveen una fuente de nutrientes para las bacterias que poseen la habilidad para degradar tales compuestos. Las cepas de *Rhizobium* que son capaces de utilizar inositol y sus derivados son mejores colonizadoras de plantas. La proteína IdhA es la primera enzima responsable del catabolismo de inositol. Una mutante *idhA* de *S. fredii* no crece en mio-inositol como única fuente de carbono; las fusiones transcripcionales indican que *idhA* se induce por myo-inositol. La mutante en simbiosis es afectada drásticamente en la habilidad para fijar nitrógeno y produce bacteroides deteriorados. Adicionalmente, el número de bacterias recuperadas de nódulos en la mutante es al menos tres veces menor respecto a la cepa silvestre. La mutante también está afectada en la capacidad para competir con la cepa silvestre en la nodulación de soya. La relación de bacterias recuperadas de nódulos, es de diez bacterias silvestres por una bacteria de la mutante (Jiang *et al.*, 2001).

El gen *kamA* codifica para una L-lysina 2,3-aminomutasa; esta cataliza la interconversión de L-lysina y L-beta-lysina, que es el primer paso en la degradación de lisina en *Clostridium subterminale* SB4. La enzima KAM también contiene un grupo hierro-azufre y requiere de piridoxal 5'-fosfato (PLP) para su activación (Ruzicka *et al.*, 2000). En *B. subtilis* los genes *yodP* y *kamA* son sobreexpresados durante la esporulación y controlados por el factor sigma E. Las deleciones en estos genes no presentan fenotipo durante el crecimiento en medio mínimo, medio rico, o condiciones de estrés osmótico (Müller *et al.*, 2011).

El gen *mcpB* codifica para una proteína de quimiotaxis (metil-acceptor); ésta, como las proteínas de su tipo, tienen papeles relevantes en la respuesta quimiotáctica de muchas bacterias. En *R. leguminosarum* McpB presenta un dominio periplasmático largo en comparación con otras MCPs. La mutante *mcpB* está impedida en la quimiotaxis en un amplio rango de fuentes de carbono en placas de “swarm”. En ensayos de competencia, al inocular plantas de haba con la cepa silvestre junto con la mutante *mcpB*, o con la mutante *mcpC*, ambas mutantes producen menos nódulos. Los autores sugieren que las MCPs participan en las etapas tempranas de nodulación (Yost *et al.*, 1998). En este trabajo observamos que en cuatro perfiles se ubican los genes que codifican para MCPs; el gen *mcpX2* en el perfil 2 (18 ddi), *mcpZf1* en el perfil 3 (MM), *mcpAch* en el perfil 4 (35 ddi) y *mcpB* en el perfil 8 (18 ddi, y 35 ddi y VL).

El gen *mexE2* codifica para una proteína del sistema de secreción de la familia HlyD para multidrogas de difusión. En *P. aeruginosa* *mexEF* y *oprN* se sobreexpresan en la mutante *nfxC*, confiriéndole resistencia a quinolonas, cloranfenicol y trimethoprima (Köhler *et al.*, 1999).

El gen *mraY* codifica para una fosfo-N-acetilmuramoil-pentapéptido transferasa. La síntesis de péptidoglicanos es vital para el crecimiento celular y para el mantenimiento de la forma celular, tanto en bacterias Gram-positivas como Gram-negativas. Las cadenas de este polímero de glicano son fijadas por puentes entre péptidos y forman una capa extracelular que provee protección contra el estrés osmótico, así como un sólido andamio para accesorios extracelulares. Las enzimas MurA, MurB y MurE sintetizan el MurNAc-pentapéptido soluble. Comenzando con UDP-GlcNAc, MraY une esta molécula a una cadena isoprenoide, formando así la membrana asociada al precursor del lipido I. En el siguiente paso, MurG agrega UDP-GlcNAc para la síntesis del lipido II, el cual es rápidamente colocado en la membrana y conectado vía reacciones de transglicosilación y transpeptidación a proteínas de unión a penicilina, lo que finaliza con la síntesis del péptidoglicano maduro (Fay y Dworkin, 2009).

Perfil 8: 18ddi.

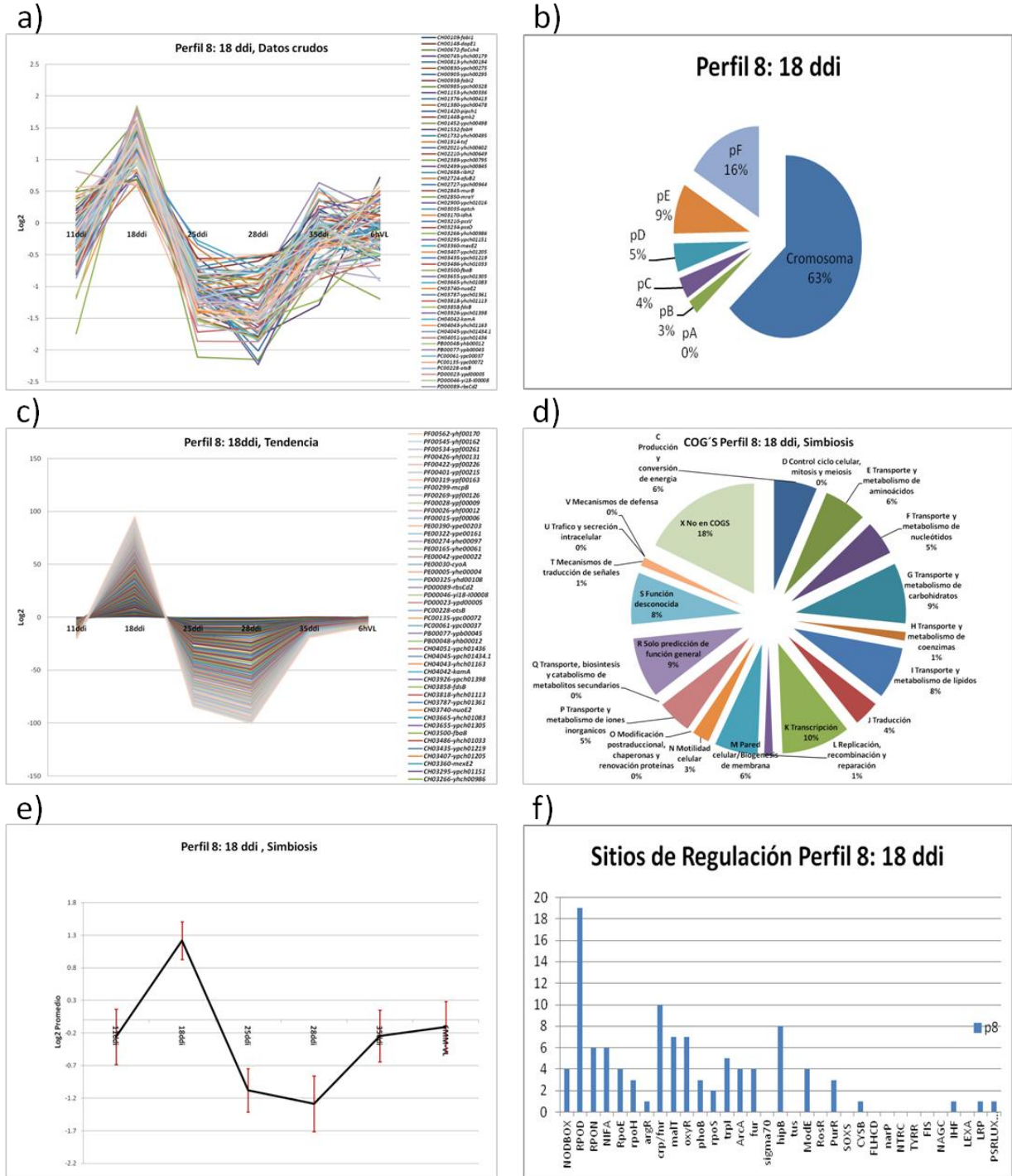


Figura 21. Perfil de expresión de los genes propios o expresados preferencialmente a los 18 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *murB* codifica para una UDP-N-acetilenoil piruvoil glucosamina reductasa. Está relacionada con MraY en la síntesis de péptidoglicano y pertenece al COG G (Biosíntesis de membrana/pared celular). Las mutantes en *murB* de *B. subtilis* son inviables, por lo que para analizar el fenotipo fueron reportadas mutantes condicionales (que mantienen la viabilidad de la célula en condiciones determinadas). Durante la esporulación aumenta la expresión del gen *murB*, y en una mutante *murB* disminuye significativamente la esporulación aún al aumentar la concentración del inductor del proceso de esporulación (Real y Henriques, 2006).

El gen *nuoE2* codifica para la cadena E de una NADH-ubiquinona oxidoreductasa. Los genes *nuo* se localizan en otros perfiles, principalmente a los 18 ddi y en MM. El gen *nuoL* codifica para la subunidad L de una NADH deshidrogenasa, que forma parte de la bomba de protones NADH: ubiquinona oxidoreductasa, también llamada complejo I, acopla la transferencia de electrones de NADH a ubiquinona con la translocación de protones a través de la membrana. En *E.coli* el complejo está constituido por 13 subunidades llamadas NuoA-NuoN y contiene un FMN y 9 grupos hierro-sulfuro (Flemming *et al.*, 2005).

El gen *otsB* codifica para la enzima trehalosa-6-fosfatasa, al igual que el gen *otsA* (trehalosa-6-fosfato sintasa) se observa su máxima expresión a los 18 ddi. La expresión de *otsB* también se presenta en menor medida a los 35 ddi. En bacterias se han definido cuatro vías para la biosíntesis de trehalosa; la más común involucra la condensación de glucosa-6-fosfato para formar trehalosa 6 fosfato a través de una trehalosa-6-fosfato sintasa (OtsA), con la subsecuente desfosforilación por OtsB, con lo que se produce la trehalosa libre (McIntyre *et al.*, 2007). En la simbiosis con *P. vulgaris* la sobreexpresión de la trehalosa-6-fosfato de *R. etli* provoca el incremento del número de nódulos, aumento de la actividad de la nitrogenasa, así como de la biomasa respecto a la cepa silvestre (Suárez *et al.*, 2008).

El gen *pipch* codifica para una prolina iminopeptidasa. En *X. oryzae* pv. *Oryzae*, agente causal de la “rotura de hoja en arroz”, el gen *pip* es blanco de regulación de OryR. Este

regulador transcripcional de tipo LuxR, interactúa con la planta de arroz a través de una molécula de señal de arroz o “rice signal molecule” (RSM). La concentración de OryR se incrementa por la concentración de RSM, la cual aumenta ante la presencia de *X. oryzae*. Los autores proponen que la característica principal de este regulador es su participación en la comunicación entre el huésped y el patógeno, a través de la expresión de genes que participan en la virulencia, como la prolina iminopeptidasa (Ferluga y Venturi, 2009).

Los genes *pssO* y *pssV*, codifican para una proteína de membrana interna involucrada en la síntesis y transporte de exopolisacáridos, y una proteína de producción de exopolisacáridos, respectivamente. En *R. leguminosarum* bv trifolii, PssO es una proteína fundamental, quizá única, para la producción de exopolisacáridos (EPS). La probable función de ésta es emular a una “chaperona” periplasmática que cubre el polímero de EPS protegiéndolo de la glicanasas, o bien cotransportándolo con el polisacárido a través del canal formado en la membrana externa (Marczak *et al.*, 2008).

El gen *rbsCd2* codifica para un transportador de ribosa tipo ABC. En el perfil 6 (18 y 35ddi) se ubica el gen *rbsCch3*, en el perfil 2 (18ddi) se observa el gen *rbsCch1* y en el perfil 4 (35ddi) se agrupo el gen *rbsCch2*. RbsC es una proteína transmembranal e hidrofóbica. En *M. loti* cepa NZP2037 a diferencia de otros rhizobios como *S. meliloti* y *R. leguminosarum*, es capaz de degradar rhizolectina, cuyo producto de degradación pudiese ser ribosa, que ha sido utilizada como fuente de carbono en el medio S-30 para cultivar esta cepa de *Mesorhizobium* (Scott *et al.*, 1987).

El gen *ribH2* codifica para una riboflavina sintasa subunidad beta. *Brucella* spp. es el agente causal de la brucelosis la cual afecta tanto humanos como otros mamíferos, *Brucella* puede sobrevivir en un nicho replicativo, el cual es caracterizado por una baja tensión de oxígeno y privado de nutrientes tales como aa y vitaminas, entre éstas, la riboflavina (vitamina B12). El metabolismo de la flavina ha sido raramente implicado en la virulencia por bacterias. Las mutantes sencilla y doble de *ribH* tienen un comportamiento diferente a la cepa silvestre, tanto en la viabilidad como en la virulencia en células de

ratón; la conclusión de los autores es que la biosíntesis de la riboflavina es esencial para la sobrevivencia en células de ratón *in vitro* o en el ratón (Bonomi *et al.*, 2010).

En este perfil los genes que destacan están involucrados en protección al estrés como *mexE2* y *otsB*; y otros genes como *mryA*, *murB*, *pssO* y *pssV* cuyos productos son parte de la membrana y confieren forma y estabilidad a la células de *R. etli*. En este perfil también se ubican genes cuyos productos participan en eventos de producción de energía como *fdsB* y *nuoE2*, así como para el intercambio de señales entre la planta y la bacteria como *pipch* y *kamA*. Los COG's que tienen un mayor número de elementos en este perfil son el K-Transcripción, al igual que en el perfil 2 se incrementa de un 6% del promedio de los 10 perfiles al 10% en este perfil. Asimismo, el COG I-Transporte y metabolismo de lípidos, incrementa su porcentaje de un 4% al 8%; además el COG F-Transporte y metabolismo de nucleótidos incrementa del 2% al 5% representando el más alto porcentaje en los 10 perfiles. El COG M-Pared celular/Biogenesis de membrana aumenta del 4% al 6%, el segundo más alto porcentaje de los 10 perfiles; y finalmente, el COG N-Motilidad celular en este punto alcanza su mayor porcentaje, con el 2.5%. Varios COG de este perfil disminuyen en el porcentaje de elementos que los constituyen, entre estos podemos mencionar el COG Q-Transporte y biosíntesis de metabolitos secundarios con una disminución del 2% al 0% y el COG T-Mecanismo de traducción de señales disminuye del 3 al 1% (Fig. 22, inciso d). Por otro lado, los principales sitios de regulación localizados pertenecen a los consensos para CRP/FNR con 10 sitios, HipB con 8, además MalT y OxyR con 7 sitios cada uno. MalT en *E. coli* tiene un efecto sobre el sistema de transporte y metabolismo de la maltosa; si se expresa constitutivamente causa una elevada expresión en una porina de membrana externa, llamada LamB que no esta relacionada con el metabolismo de la maltosa (Reimann y Wolfe, 2011).

Al igual que en el perfil 2, este perfil presenta un número elevado de reguladores entre los que destacan 2 miembros de la familia MarR (ID's CH02389 y CH02727). Los reguladores de esta familia son positivos y negativos, presentan un amplio espectro de blancos de regulación que participan en diversos mecanismos, principalmente relacionados con la virulencia; adicionalmente, los reguladores de la familia MarR compiten con otras

proteínas para la unión a la RNA polimerasa y pueden activar la transcripción (Ellison y Miller, 2006).

Perfil 9 (25 ddi)

En este perfil se observaron los genes expresados a los 25 ddi, y una ligera disminución a los 28 ddi, destaca porque se localizan los genes involucrados directamente en la fijación de nitrógeno como los *nifHDK* y *fixABCX*.

El gen *alkA* codifica para una DNA-3-metiladenina glicosidasa II. En *E. coli* se induce por la presencia de agentes alquilantes. Una fusión del gen *alkA-LacZ* de *E. coli* fue transferida a 33 diferentes bacterias, observándose actividad en casi todas, concluyendo que los sitios de regulación así como la actividad de la enzima funciona y está conservada en casi todas las bacterias, a excepción de *A. tumefaciens*, *S. meliloti* y cinco bacterias más, en el análisis no incluyeron *R. etli* (Fernandez de Henestrosa y Barbé, 1991).

El gen *aptd* codifica para una adenina fosforibosil transferasa, es parte del regulón NifA-RpoN (Salazar *et al.*, 2010). La función del Apt es catalizar una reacción de salvamento que da por resultado la formación de AMP + Difosfato a partir de Adenina + 5-fosfo-alfa-D-ribosa 1-difosfato; esto es menos costoso energéticamente que la síntesis “*de novo*”. Al igual que en nuestro análisis, en *R. etli* utilizando nódulos de 21 ddi, también fue detectada la expresión de un gen *aptD* exclusivo de simbiosis (Vecruysse *et al.*, 2011). Otro gen *aptch* localizado en cromosoma se agrupa en el perfil 8 (18 y 35 ddi, MM).

El gen *ardC* codifica para una enzima de transferencia conjugativa antirestricción. La conjugación bacteriana juega un papel relevante en la diseminación de los genes de resistencia a antibióticos en la célula. Se ha propuesto que ArdC es transportada a la célula hospedera con DNA de cadena sencilla (ssDNA), con lo que lo protege de las endonucleasas del hospedero (Belogurov *et al.*, 2000).

El gen *azIC* codifica para un transportador de cadenas ramificadas de aminoácidos (resistencia de azaleucina). En *B. subtilis* una mutante *azIC* no confiere resistencia a azoleucina, sin embargo una sobre expresión de *azIC* y *azID* en una mutante en *azIB* es

resistente a azoleucina (Belitsky *et al.*, 1997). AzlC es una proteína hidrofóbica, que probablemente funciona en el transporte de cadenas-ramificadas de aminoácidos. Probablemente está relacionada con leucina, en este perfil también se localiza el gen *leuD*. En un análisis del transcriptoma usando “Tiling arrays” en nódulos de 21 ddi, el transcrito para *azlC* es expresado de manera diferencial respecto a la fase estacionaria de la VL (Vercruyse *et al.*, 2011).

El gen *casA* codifica para una proteína de unión a calcio (calsimina). En *R. etli* CPA540 *casA* se expresa durante la colonización e infección con el hospedero. La expresión es controlada por un represor que pertenece a la familia TetR. Una mutación en *casA* afecta el desarrollo del bacteroide durante la simbiosis y la fijación de nitrógeno (Xi *et al.*, 2000).

El gen *coaA* codifica para una pantotenato cinasa. Los genes para la biosíntesis de pantotenato en *R. etli* se distribuyen en cromosoma y en los plásmidos E y F. Los genes *panCB* que se localizan en el plásmido F, son indispensables para la síntesis de pantotenano, pero su sola presencia no restaura el crecimiento en medio mínimo de una cepa de *R. etli* carente del pF (Villaseñor *et al.*, 2011). El parásito *Plasmodium*, utiliza la biosíntesis de ácidos grasos y de elongación, así como la biosíntesis de pantotenato y CoA para obtener nutrientes del ambiente del hospedero, en este caso de células de humano (Xu *et al.*, 2010).

Los genes *cpxA* y *cpxP2* codifican para citocromos P450 monooxigenasas, ambos pertenecen al regulón RpoN-NifA (Salazar *et al.*, 2010). A los citocromos P450 se les ha involucrado en el metabolismo de una gran variedad de sustancias endobióticas y xenobióticas. En *B. japonicum*, los citocromos *cyp112* y *cyp114* no se expresan en vida libre en aerobiosis, pero sí en células cultivadas en anaerobiosis y en bacteroides. La mutante en *cyp112* tiene un efecto polar sobre *cyp114* y produce nódulos efectivos en soya, lo que sugiere que los citocromos P450 (*cyp112* y *cyp114*) no están involucrados en una función esencial para la simbiosis (Tully y Keister, 1993). En *B. japonicum* el citocromo P450 también es regulado por RpoN-NifA (Hauser *et al.*, 2007).

Perfil 9: 25 y 28 ddi.

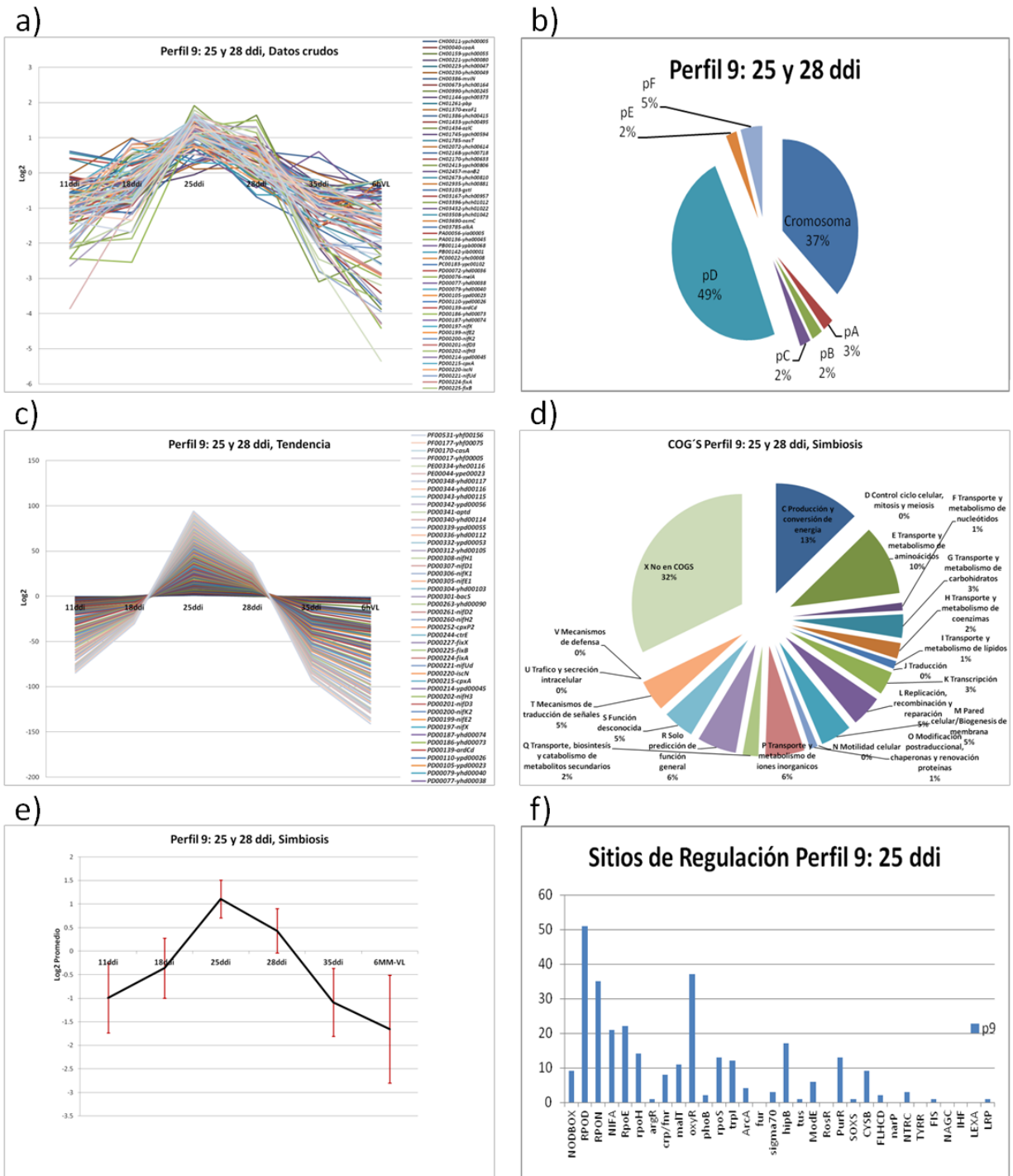


Figura 22. Perfil de expresión de los genes propios o expresados preferencialmente a los 25 y 28 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG's, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *ctrE* codifica para una farnesil transtransferasa, es parte del regulón RpoN-NifA (Salazar *et al.*, 2010). En plantas y eucariotes la farnesilación es requerida para blanco “targeting” de membrana, interacciones proteína-proteína y para la actividad biológica de proteínas regulatorias claves, como Ras y cinasas. Una farnesil transferasa de haba, se expresa en nódulos con una activa división celular. Un análisis en células en suspensión de tabaco sugiere que la farnesil transferasa es requerida para el ciclo celular de la planta, ya que la actividad enzimática precedió a cambios en la fase temprana de la actividad mitótica, quizás a partir de la modulación de la progresión a través de la fase S y la transición de la fase G1 a la fase S, lo que no sucede cuando se agrega el inhibidor específico para la farnesiltransferasa (Quian *et al.*, 1996).

El gen *exoF1*, codifica para una proteína de biosíntesis de exopolisacáridos, el punto máximo de expresión se presenta a los 28 ddi. En el perfil 1 (11ddi) se ubica el gen *exoR*, en el tiempo de 18 ddi y 35 ddi se detectaron otros genes que codifican para proteínas que participan en la síntesis de exopolisacáridos. En mutantes *exoF* en *Sinorhizobium sp.* NGR234 es suprimida la producción de exopolisacáridos (Staehein *et al.*, 2006).

El gen *gstI* codifica para una proteína inhibidora de la traducción de la glutamino sintetasa. En *R.leguminosarum* la presencia de *GstI* reduce la capacidad de asimilación de NH_4^+ . También en *R. leguminosarum*, *gstI* es regulado por el sistema Ntr, incluyendo la proteína PII, que funciona para censar el estatus del nitrógeno en las células. Con un promotor inducible, *GstI* inhibe la expresión de la glutamino sintetasa II (*glnII*) en todas las condiciones analizadas. En simbiosis, *gstI* se expresa en simbiosomas jóvenes diferenciados (SBs), pero no en simbiosomas diferenciados fijadores de nitrógeno. En nódulos jóvenes, la proteína PII modula la transcripción de *ntrC* cuyo producto regula los genes *gstI* y *glnII* (Tatè *et al.*, 2001).

El gen *leuD* codifica para una isopropil-malato isomerasa subunidad pequeña. En *S. meliloti* Rmd201 las mutantes en *leu* son capaces de nodular, pero tienen un fenotipo simbiótico Fix-; el suplementar con leucina el medio no revierte el fenotipo de auxotrofia. Estudios histológicos de nódulos producidos por la mutantes en *leu*, presentan un

desarrollo incompleto respecto a los nódulos de la cepa silvestre, aunque los nódulos de la mutante *leuB* se observan más estructurados que los nódulos de las mutantes *leuC* y *leuD* (Hassani *et al.*, 2002).

El gen *mviN* codifica para un factor de virulencia transmembranal. En *R. tropici* CIAT899 induce una severa clorosis foliar cuando se aplica a semillas de plantas dicotiledonias hospederas y no hospederas. Una mutación en *mviN* no tiene un efecto en la habilidad de *R. etli* CIAT899 de inducir la clorosis en plantas de *P. vulgaris*, tampoco tiene un efecto significativo en la nodulación y en la fijación de nitrógeno, pero sí afecta la motilidad en *R. tropici* (O'Connell *et al.*, 1998). En *S. meliloti* los genes *glnD* y *mviN* se cotranscriben y son indispensables para la bacteria, ya que mutantes en estos sólo se pueden obtener al complementar con el cósmido que posee ambos genes (Rudnick *et al.*, 2001). En *B. subtilis* existen 4 proteínas probablemente homólogas a MviN; en este trabajo indican que *mviN* codifica para una lipido flipasa II, necesaria durante la esporulación, para la síntesis de péptidoglicano de la corteza de la espora. Una cuádruple mutante no presenta un defecto en el crecimiento (Fay y Dworkin, 1999).

El gen *nasT* codifica para un regulador del tipo sistema de dos componentes. En *R. capsulatus* en una región de 17 kb se localizan los genes del sistema para la asimilación de nitrato (*nas*), en donde el producto de *nasT* es un antiterminador de la transcripción (Pino *et al.*, 2006).

El gen *osmC* codifica para una proteína de inducción osmótica; a pesar de localizarse en el perfil 9 (25 y 28 ddi), la máxima expresión se observa a los 25 ddi pero cercana respecto a los 18 ddi. Previamente fue reportado que en nódulos de *R. etli* a los 21 días se detecta la expresión de este gen (Vecruijsse *et al.*, 2011). En *E. coli*, *osmC* se induce por la presencia de sal; estudios de cristalografía de proteínas y datos adicionales en su estructura señalan que pudiera funcionar como una peroxiredoxina (Shin *et al.*, 2004).

El gen *pbp* codifica para una proteína de unión a penicilina de biosíntesis de la pared celular. En *A. tumefaciens*, PBP Atu2422 está implicada en la virulencia y transporte de GABA (Planamente *et al.*, 2010).

En este perfil se ubican los genes del operón *nifHDK* que codifican para la nitrogenasa reductasa (*nifH*), y las nitrogenasas hierro-molibdeno cadena alfa y beta (*nifD* y *nifK* respectivamente). *R. etli* cuenta con tres copias del gen *nifH*, dos copias completas y una truncada del gen *nifD*, y 2 copias del gen *nifK*. Los ocho elementos se ubican en este perfil y son parte del regulón RpoN-NifA (Salazar *et al.*, 2010). Además, en este perfil se ubican *nifE1* (truncado) y *nifE2*, *nifUd* y *nifX*, que codifican para una proteína ensambladora de la nitrogenasa (NifE), para una proteína de fijación de nitrógeno (NifU) y para la proteína de procesamiento del cofactor hierro-molibdeno (NifX). Estos genes también son parte del regulón NifA-RpoN (Salazar *et al.*, 2010).

Los genes del operón *fixABCX* codifican para proteínas involucradas en la transferencia de electrones durante la fijación de nitrógeno, su expresión se reportó como dependiente de RpoN-NifA (Salazar *et al.*, 2010).

En este perfil también se ubican: el gen *meIA* codifica para una monofenol monooxigenasa, los genes *bacS* y sus copias *yhd00073* e *yhd00114* codifican para una proteína de bacteroide S, para su expresión dependen de RpoN y NifA y sus mutantes no presentan un fenotipo simbiótico (Noel *et al.*, 2002, Salazar *et al.*, 2010).

En este perfil prácticamente se agrupan la mayoría de los genes regulados por RpoN-NifA, a pesar de corresponder a dos diferentes tiempos (25 y 28 ddi). Los genes identificados como parte del regulón RpoN-NifA presentan su máxima expresión a los 25 ddi y en el tiempo de 28 ddi se observa en éstos una reducción de la expresión de alrededor del 50%. Además de los genes descritos en este perfil con una obvia relevancia en la fijación de nitrógeno, un gen a destacar es el *ypch00593* el cual codifica para una proteína hipotética descrita como una dioxigenasa 4-hidroxifenilpiruvato. El gen que pertenece al perfil 9 (25 y 28 ddi) tiene una expresión tres veces mayor en nódulos de 28 ddi respecto a los 25 ddi. En nódulos de soja (*Glycine max*) se observa la expresión de un par de dioxigenasas 2-dependientes de oxoglutarato (2-ODDs); aunque sólo una de ellas aumenta su expresión en nódulos senescentes o expuestos a obscuridad. A esta 2-ODD la catalogan como un marcador de senescencia del nódulo (Webb *et al.*, 2008). En este perfil, a diferencia de los

otros 9, el mayor porcentaje de los genes se localizan en el plásmido D o simbiótico. Los reguladores ubicados en este perfil son dos, uno de la familia AraC y un probable regulador transcripcional. En este perfil dominan los sitios para RpoN, y para el activador transcripcional NifA, aunque RpoE y OxyR también están sobrerrepresentados. En este perfil el COG C-Producción y conversión de energía alcanza su máximo porcentaje con el 13%, el porcentaje de este COG en los 10 perfiles es del 6%. El COG H-Transporte y metabolismo de coenzimas también alcanza su punto más alto con el 2.3%, siendo el promedio de los 10 perfiles del 1%. Los COG's que disminuyen en el porcentaje de elementos que lo constituyen respecto al porcentaje promedio de los 10 perfiles son: COG G-Transporte y metabolismo de carbohidratos, disminuye del 10% al 3.5% y el COG J-Traducción, del 5% en promedio, se queda sin representante en este perfil. Los grupos enriquecidos positivamente por GSEA son el metabolismo de nitrógeno y la degradación de tetracloroetano, en tanto que los principales enriquecidos negativamente son: quimiotaxis, amino-azúcares y transportadores tipo ABC (Fig. 28).

Perfil 10 (18 ddi y MM)

En este perfil se localizan los genes expresados en vida libre y en la simbiosis a los 18 ddi, donde destacan los siguientes:

El gen *braC1* codifica para un transportador de aminoácidos tipo ABC. En los nódulos, el citosol de la planta es el principal sitio para la asimilación de amonio en aminoácidos; los genes *aap* y *bra* son transportadores de aminoácidos que operan en los bacteroides. En *R. leguminosarum* bv *viciae* mutantes en estos sistemas son incapaces de transportar aminoácidos ramificados, lo que produce un fenotipo de hambruna de nitrógeno cuando se inoculan plantas de chícharo (*P. sativum*). El fenómeno también se observa en plantas de frijol común (*P. vulgaris*), pero no en alfalfa (*M. sativa*) cuando se inocula con la doble mutante (*bra* y *aap*) de *S. meliloti*. Los autores proponen que este último fenotipo se debe a que el requerimiento de aa ramificados es en pequeñas cantidades, así se revierte la auxotrofia (Prell *et al.*, 2010). En los perfiles 2 (18 ddi) se localiza el gen *braF1*, y en el

perfil 3 (MM) los genes *aapF* y *aapJ*. En *R. etli*, las proteínas BraC2 y AapJ fueron identificadas en bacteroides de 18 ddi (Resendis-Antonio *et al.*, 2011).

El gen *cheY2* codifica para una proteína del sistema de dos componentes para la quimiotaxis. En *E. coli*, CheY regula la quimiotaxis transmitiendo información entre los quimiorreceptores y el interruptor del motor flagelar (Jenal y Galperin, 2009). En *S. meliloti* se caracterizó un regulador del sistema de dos componentes llamado CheY-like, el cual mediante represión catabólica regula el uso de succinato, mientras suprime la utilización de algunas fuentes secundarias de carbono, tales como α y β -galactosidasas. La mutante en *cheY*-like sobreproduce polihidroxibutirato (PHB) (García *et al.*, 2010).

El gen *clpS* codifica para una proteasa dependiente de ATP. En *S. meliloti* utilizando microarreglos para evaluar la expresión de la bacteria ante un pH bajo (5.75), observaron la expresión del gen *clpS* en el perfil que los autores llaman B. Las características es una sobreexpresión gradual que se mantiene a lo largo de 20 minutos después del cambio de pH. La diferencia entre este grupo y el perfil A, es la intensidad de la señal de expresión (de Lucena *et al.*, 2010). Este tipo de proteasas permite a la célula hacer frente al mal plegado o a las proteínas desnaturalizadas, condición que se incrementa en la presencia de una condición de estrés; por ejemplo, durante un estrés térmico previene la agregación de proteínas lo que permite el reciclaje de aminoácidos. En *R. etli* *clpS* es dependiente de la alarmona (p)ppGpp (Vercruyssen *et al.*, 2011).

El gen *ctpA* codifica para una proteína de procesamiento C-terminal peptidasa. La envoltura celular de las bacterias es de importancia crítica para el funcionamiento y sobrevivencia de la célula, ya que actúa por un lado como barrera para sustancias tóxicas y por el otro permite el flujo de nutrientes dentro de la célula. Durante la simbiosis también funciona para el contacto físico entre la bacteria y su hospedero. En *R. leguminosarum* utilizando una mutante *ctpA* se determinó que CtpA es indispensable para la viabilidad de la bacteria cuando es cultivada en un medio complejo semisólido (TY), pero es dispensable cuando se utiliza un medio líquido, lo que achacan a la susceptibilidad a la desecación en el medio semi-sólido. La mutante incrementa su sensibilidad a

detergentes, lo que indica una permeabilidad de la envoltura celular, lo que lleva a los autores a sugerir un papel para CtpA en la resistencia a la desecación (Gilbert *et al.*, 2007).

El gen *dppA* codifica para una proteína transportadora de dipéptidos tipo ABC. Mutantes en *R. leguminosarum* fueron afectadas severamente en el importe del ácido delta-aminolevulínico (ALA), un precursor del grupo hemo. La toma de ALA es antagonizada en la presencia de dipéptidos, lo que sugiere que ambas moléculas utilizan el mismo sistema de transporte. Utilizando como modelo *V. faba-R. leguminosarum*, mutaciones en *dppABCDF* no presentan un fenotipo simbiótico, lo que indica que el ALA necesario para la síntesis de hemo no es suplementado por la planta, o bien existe otro sistema de transporte que funciona en los bacteroides. Por otro lado, los transcritos de *dppA* y *dppB* contienen secuencias regulatorias inversas repetidas que pudieran estabilizar el RNA mensajero, el promotor para el operón *dppABCDF* es de tipo sigma 70 (Carter *et al.*, 2002).

El gen *etfA* codifica para una flavoproteína para transferencia de electrones. En simbiosis se ha determinado que los productos de los genes *fixABCX* participan en la transferencia de electrones; también han sido descritos como indispensables para la fijación de nitrógeno. En el perfil 5 (18, 35 ddi y MM), detectamos la máxima expresión de *etf*, el producto de *etfA* fue identificado mediante proteómica en bacteroides a los 18 ddi (Resendis-Antonio *et al.*, 2011).

El gen *gap* codifica para una gliceraldehído 3-fosfato deshidrogenasa. En *R. etli* CFN42 la proteína Gap también fue detectada mediante proteómica a los 18 ddi (Resendis-Antonio *et al.*, 2011). En *S. meliloti* SU47 se obtuvieron mutantes incapaces de crecer en succinato como fuente de carbono; la mutante *gap* fue una de éstas (Finan *et al.*, 1988). En este perfil también se localiza el gen *pckA*. Tanto la proteína Gap como la PckA participan en la gluconeogénesis.

Perfil 10: 18 ddi y VL.

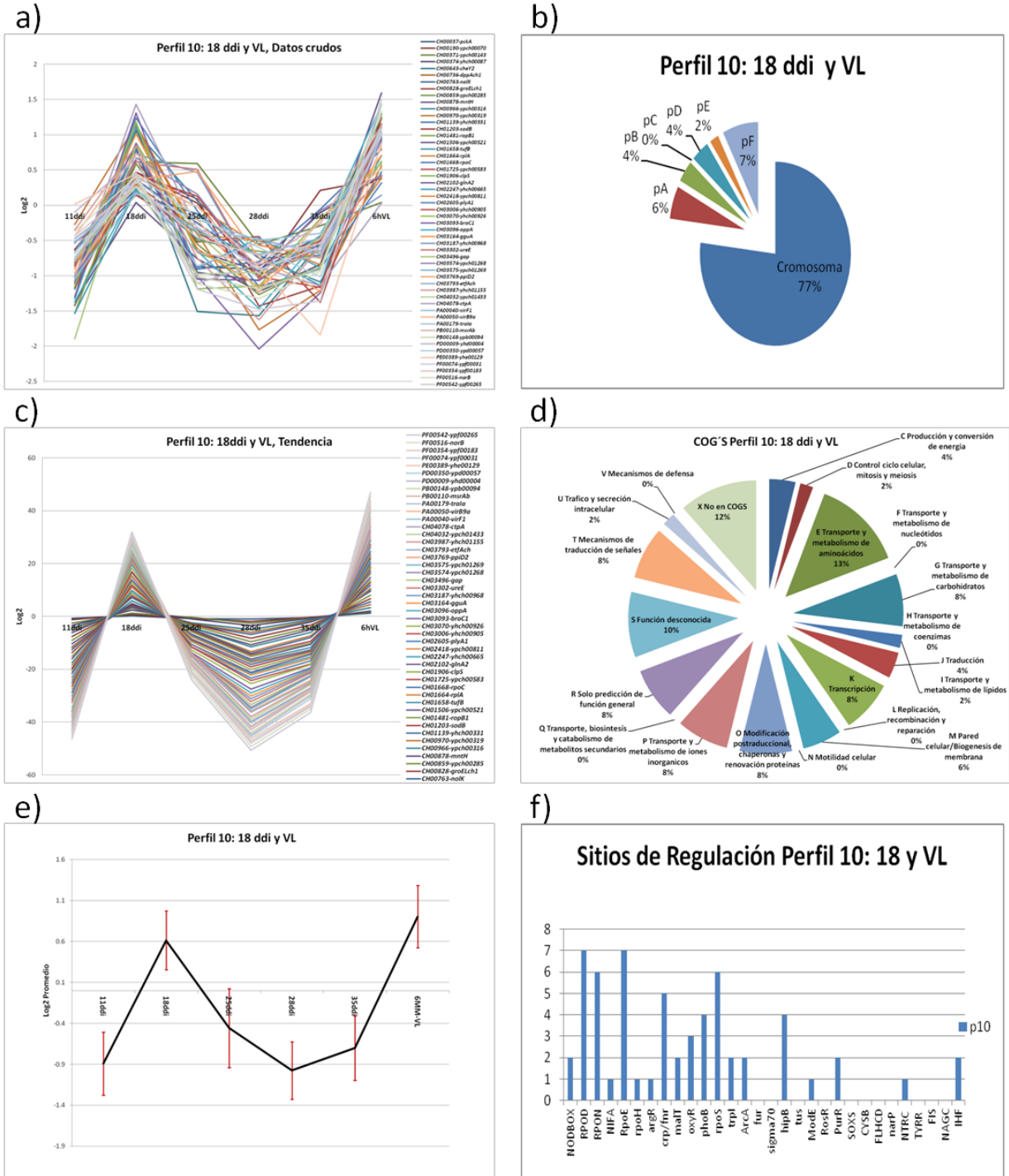


Figura 23. Perfil de expresión de los genes propios o expresados preferencialmente en vida libre (VL) y 18 ddi, a) datos crudos, b) distribución por replicón, c) tendencia global, d) distribución de COG'S, e) perfil con desviación estándar y f) sitios de regulación localizados.

El gen *gguA* codifica para un transportador de xilosa tipo ABC, de unión a ATP. Inicialmente se caracterizó en *Agrobacterium* por su vecindad con el gen *chvE* (perfil 3, MM), cuyo producto tiene un papel indispensable en la virulencia de la bacteria. Los genes *gguABC* fueron propuestos para el transporte de glucosa y galactosa. Mutantes en estos genes no presentan un efecto en la virulencia de *Agrobacterium* sobre *K. diagraphomontiana*; ni en el crecimiento en 1 mM de galactosa, glucosa, xilosa, ribosa, arabinosa, fucosa, o sacarosa; ni en la quimiotaxis en glucosa, galactosa, xilosa, o arabinosa (Kemner *et al.*, 1997).

El gen *glnA* codifica para una glutamino sintetasa. En *R. leguminosarum* la concentración de GS-I se incrementa de dos a cuatro veces cuando se cultiva en una fuente pobre de nitrógeno, comparado con un medio rico en nitrógeno. Cuando *R. leguminosarum* se cambia de un medio rico en nitrógeno a uno pobre, la actividad de GS-I es primordial para la desrepresión de GS-II. Lo que sugiere que bajo estas condiciones, GS-I se requiere como señal para la síntesis de glutamina y así esta se pueda utilizar en la síntesis de proteínas. En simbiosis se ha reportado una baja actividad de GS y mediante la manipulación genética han observado que ninguna de las isoformas de GS son necesarias para una simbiosis efectiva (Patriarca *et al.*, 2002). En *S. meliloti* una doble mutante *glnA-glnII* induce una invasión normal de nódulos (de Bruijin *et al.*, 1989) y en un reporte posterior detectan la actividad de una tercera GS-III (Shatters *et al.*, 1993). Lo contrario ocurre en *B. japonicum*, el cual no cuenta con una GS-III, donde la doble mutante *glnA-glnII* afecta severamente la nodulación (Carlson *et al.*, 1987). Por lo anterior, la necesidad de una glutamino sintetasa activa durante la fijación de nitrógeno es contradictoria, aunque Patriarca y col., lo clarifican indicando que la simbiosis es un proceso gradual, que la invasión de la bacteria puede depender de la capacidad para sintetizar glutamina, y por lo tanto la actividad de la GS parece ser dispensable para la función de los bacteroides (Patriarca *et al.*, 2002).

El gen *groELch1* codifica para una chaperona. Los genes *groEL* y *groES* se distribuyen en 3 diferentes perfiles: perfil 3 (MM), perfil 7 (25 ddi y MM) y perfil 10 (18 ddi y MM), pero en

todos la señal máxima de expresión se presenta en MM. Se hizo la descripción de estos en los perfiles 3 y 7.

El gen *mntH* codifica para una proteína transportadora de manganeso (Mn). En *B. japonicum*, MntH es esencial para el transporte de Mn y para mantener la homeóstasis celular. Este gen se induce bajo condiciones deficientes de Mn y no le afecta la presencia de hierro. El control de la expresión del gen *mntH* se pierde en una mutante *fur*, además ya ha sido reportado el sitio en el promotor *mntH* en donde se une Fur (Hohle y O'Brian, 2009). Los transportadores de Mn son esenciales para el crecimiento sin estrés en algunas bacterias, pero en otras es requerido para inducir las respuestas primarias ante el estrés. En *B. japonicum* en una mutante *mntH* crecida bajo limitaciones de manganeso, las actividades de la superóxido dismutasa (SodM) y de la piruvato cinasa (PykM) se ven afectadas. Una mutante *mntH* crece más lentamente en piruvato en condiciones de limitación de Mn que una doble mutante en *pykM* y *sodM*, lo que indica la implicación de otros procesos dependientes de Mn (Hohle y O'Brian, 2012).

El gen *mrsA* codifica para una sulfoxido metionina A reductasa. Los intermediarios reactivos de oxígeno y nitrógeno pueden causar daño a una gran cantidad de componentes celulares. Para contender con esto, las células tienen una variedad de mecanismos que destruyen estas moléculas reactivas o reparan el daño si este ocurre. En las proteínas, uno de los aminoácidos más fácil de ser oxidado es la metionina, la cual es convertida a sulfoxido metionina. La enzima MsrA cataliza la reducción de sulfoxido metionina para que las proteínas tengan disponibilidad de metionina. Por lo tanto, MsrA juega un importante papel en la protección de las células contra el daño oxidativo (Weissbach *et al.*, 2002).

El gen *nolK* codifica para una GDP-L-fucosa sintasa. El producto de este gen de nodulación participa en la biosíntesis y modificación de fucosa y manosa. En *A. caulinodans*, NolK está involucrado en la síntesis de GDP-fucosa, la cual es utilizada por NodZ como el donador de fucosil en la fucosilación de los factores nod (Mergaert *et al.*, 1997). Las bacterias *B. elkanii*, *B. japonicum*, *R. fredii*, y *Sinorhizobium* sp. NGR234 secretan factores nod

fucosilados, por lo que a NolK se le ha relacionado con la capacidad de nodulación. Mutantes *nodZ* de *Sinorhizobium* sp. NGR234 producen factores nod no fucosilados y la bacteria es incapaz de nodular *Pachyrhizus tuberosus* pero si nodulan soya (*G. max*). En tanto, en *B. japonicum* es requerido para nodular *M. atropurpureum* pero no para *G. max*. Incluso cepas transconjugantes de *R. leguminosarum* que contienen el gen *nodZ* de *B. japonicum* adquieren la capacidad de nodular *Macroptilium* pero no fijan nitrógeno (Perret *et al.*, 2000).

El gen *norB* codifica para una óxido nitrito reductasa. En *R. etli* los genes *nap* o *nar* están ausentes, lo que inhabilita a la bacteria para usar nitrato para respirar. Los genes *nirK* y *norCB* son regulados por NnrR en respuesta a óxido nítrico en microaerobiosis. El nivel del complejo nitrosil leghemoglobina en plantas de frijol expuestas a nitrato se incrementa en una mutante *norC*, pero decrece en una mutante *nirK*; lo que indica que la nitrito reductasa es un elemento fundamental en nódulos de frijol en respuesta a nitrato (Gómez-Hernández *et al.*, 2011).

El gen *oppA* codifica para un transportador de oligopéptidos tipo ABC, unión al sustrato. En el perfil 4 (35ddi) se expresa el gen *oppC*. Estos transportadores son requeridos para la toma de tetrapéptidos y de ciertos tripéptidos como 3Ala y bialafos. En *S. meliloti*, las mutantes en este tipo de transportadores no tienen un efecto en el establecimiento de la interacción (simbiosis) con la alfalfa. En vida libre, a diferencia de los genes *dpp* que presentan una expresión constitutiva, la máxima expresión para los genes *opp* se reporta en fase estacionaria (Nogales *et al.*, 2009). La proteína OppA fue detectada mediante proteómica a los 18 ddi (Resendis-Antonio *et al.*, 2011).

El gen *pckA* codifica para una PEP carboxilasa, una enzima clave en la gluconeogénesis. Esta enzima cataliza el primer paso en la conversión de los ácidos tricarbónicos intermediarios a azúcares de hexosa, conduciendo también a la síntesis de glicógeno. El fenotipo simbiótico de una mutante en *pckA* en *R. etli* CE3 es una baja producción de nódulos dentro de los cuales el hilo de infección parece no penetrar (Tatè *et al.*, 2004). En *R. etli* CFN42 fueron detectadas múltiples isoformas de PckA mediante proteómica y en

una mutación *in silico* sugieren que PckA es esencial para la fijación de nitrógeno (Resendis-Antonio *et al.*, 2011).

El gen *plyA2* codifica para una polisacaridasa, en el perfil 3 (MM) se ubica *plyA1*. Las proteínas PlyA, al igual que PlyB, degradan EPS y caboximetil celulosa. En *R. leguminosarum* *plyA* se expresa a muy bajo niveles en cultivos de vida libre; en simbiosis, una doble mutante *plyA-plyB* no afecta la fijación de nitrógeno y tiene un efecto insignificante en la nodulación (Finnie *et al.*, 1998).

El gen *ppiD2* codifica para una peptidil prolil cis-trans D isomerasa, péptido señal. En *Campylobacter jejuni* un patógeno gastrointestinal de humano, utilizando proteómica como herramienta de análisis en un ensayo de inducción por temperatura, se identificaron dos peptidil prolil cis-trans isomerasas PpiD2 y PpiB. El papel asignado es en el plegamiento de proteínas; no obstante, los autores señalan que la sobrerregulación de éstas tendría un papel relevante en el proceso de fijación de nitrógeno, debido a que en un reporte previo, también en *C. jejuni*, una PpiD está involucrada en la motilidad, invasión y colonización (Rathbun *et al.*, 2009).

El gen *ropB1* codifica para una proteína de membrana externa. En el perfil 7 (25ddi y MM) se localizan los genes *ropAch1* y *ropAch2*. A diferencia de *R. etli* CFN42 que tiene reportados dos genes *ropB*, *R. leguminosarum*, contiene los genes *ropB*, *ropB2* y *ropB3*, cuyas mutantes individuales no tienen fenotipo simbiótico, y su expresión en la simbiosis fue insignificante. La función de *ropB2* y 3 permanece desconocida, en tanto que una mutante *ropB* en vida libre es más sensible a los detergentes, antibióticos hidrofóbicos y ácidos débiles, por lo que los autores sugieren que la función de *ropB* es para la estabilidad de la membrana externa (Foreman *et al.*, 2010).

El gen *sodB*, codifica para una superóxido dismutasa. En el nódulo las bacterias están expuestas a especies reactivas de oxígeno. En *R. etli* CFN42 se detectó en simbiosis la proteína SodB que participa en la destoxificación de las especies reactivas de oxígeno (Resendis-Antonio *et al.*, 2011). En *S. meliloti* la mutante *sodA* presenta fenotipo simbiótico ya que la capacidad de fijar nitrógeno se deteriora. Incluso, la mayoría de las

bacterias deficientes en SOD no alcanzan el estado de diferenciación a bacteroides y entran rápidamente en senescencia (Santos *et al.*, 2000). Las actividades de SodA y SodB están íntimamente relacionadas con la presencia de manganeso como cofactor. En este perfil se localiza también el gen *mntH* (transportador de manganeso).

El gen *ureE* codifica para una ureasa. En nódulos de *M. truncatula* inducidos por *S. meliloti* y sometidos a estrés hídrico, una de las proteínas que se sobreexpresa en el bacteroide es UreE, junto con otras enzimas que participan en la asimilación de nitrógeno, lo que conduce a sugerir un intercambio activo de aminoácidos entre la bacteria y la planta (Larrainzar *et al.*, 2007).

El gen *virB9* codifica para una proteína de transporte del sistema de secreción tipo IV. En *Agrobacterium* es parte del complejo VirB7-VirB9. Estas proteínas de membrana externa participan en la formación del poro de transporte, siendo VirB9 la que forma el más grande oligómero. En esta bacteria VirB8 y VirB9 son indispensables para la transferencia del DNA (Das y Xie, 2000).

El gen *virF* codifica para una proteína de virulencia. Su probable función en *A. tumefaciens* es la de cortar o degradar las proteínas que cubren el T-DNA una vez que éste ha sido integrado a su hospedero. Se ha reportado que VirF puede ser exportado a las células de la planta (Tzfira *et al.*, 2004).

En este perfil destacan genes que codifican para enzimas que participarían en el manejo de la interacción planta-bacteria, desde el control de la asimilación de aminoácidos, nitrógeno, producción de energía, protección y otros. Los genes que se destacan son los relacionados para contender contra diversos tipos de estrés, por ejemplo el oxidativo (*sodB*). En este perfil el COG E-Transporte y metabolismo de aminoácidos, tiene su segundo nivel más alto entre los perfiles con el 13.5%, el porcentaje promedio de los 10 perfiles es del 10%. El COG T-Mecanismo de traducción de señales presenta un 7.7%, siendo el porcentaje promedio de los 10 perfiles de 3%. Además, los COG's P-Transporte y metabolismo de iones inorgánicos y O-Modificación postraducciona, chaperonas y renovación de proteínas incrementan del 4% en promedio de los 10 perfiles a

aproximadamente el 8% en este perfil. Por el contrario, los COG's F-Transporte y metabolismo de nucleótidos, H-Transporte y metabolismo de coenzimas, N-Motilidad celular, L-Replicación, recombinación y reparación, y finalmente el COG Q-Transporte, biosíntesis y catabolismo de metabolitos secundarios no tienen representantes en este perfil (Fig. 23, inciso d). Los sitios de regulación predominantes para este perfil son: RpoE con 7 blancos, 6 sitios para blancos de RpoS al igual que 6 sitios para RpoN. A diferencia de otros perfiles, donde el consenso para RpoN está sobrerrepresentado (perfiles 7 y 9), en este perfil no es acompañado por un enriquecimiento de los sitios para NifA. Si bien en *R. etli* no existe un RpoS identificado, el número de sitios que reconoce este regulador en este perfil alcanza su máximo con 6 elementos. Nuevamente, como en los diferentes perfiles que involucran el tiempo 18 ddi, en este perfil se identificaron 5 sitios para FNR. En este perfil, se ubican un probable regulador transcripcional y un gen cuyo producto pertenece a la familia de reguladores MarR. Los reguladores de esta familia son positivos y negativos, presentando un amplio espectro de blancos de regulación que se distribuyen en diversos mecanismos principalmente en virulencia; adicionalmente, compiten con otras proteínas para la unión a la RNA polimerasa y pueden activar la transcripción (Ellison y Miller, 2006). En el perfil 8 (18ddi) se ubican dos genes que codifican para proteínas de esta familia.

10 perfiles: Simbiosis y VL.

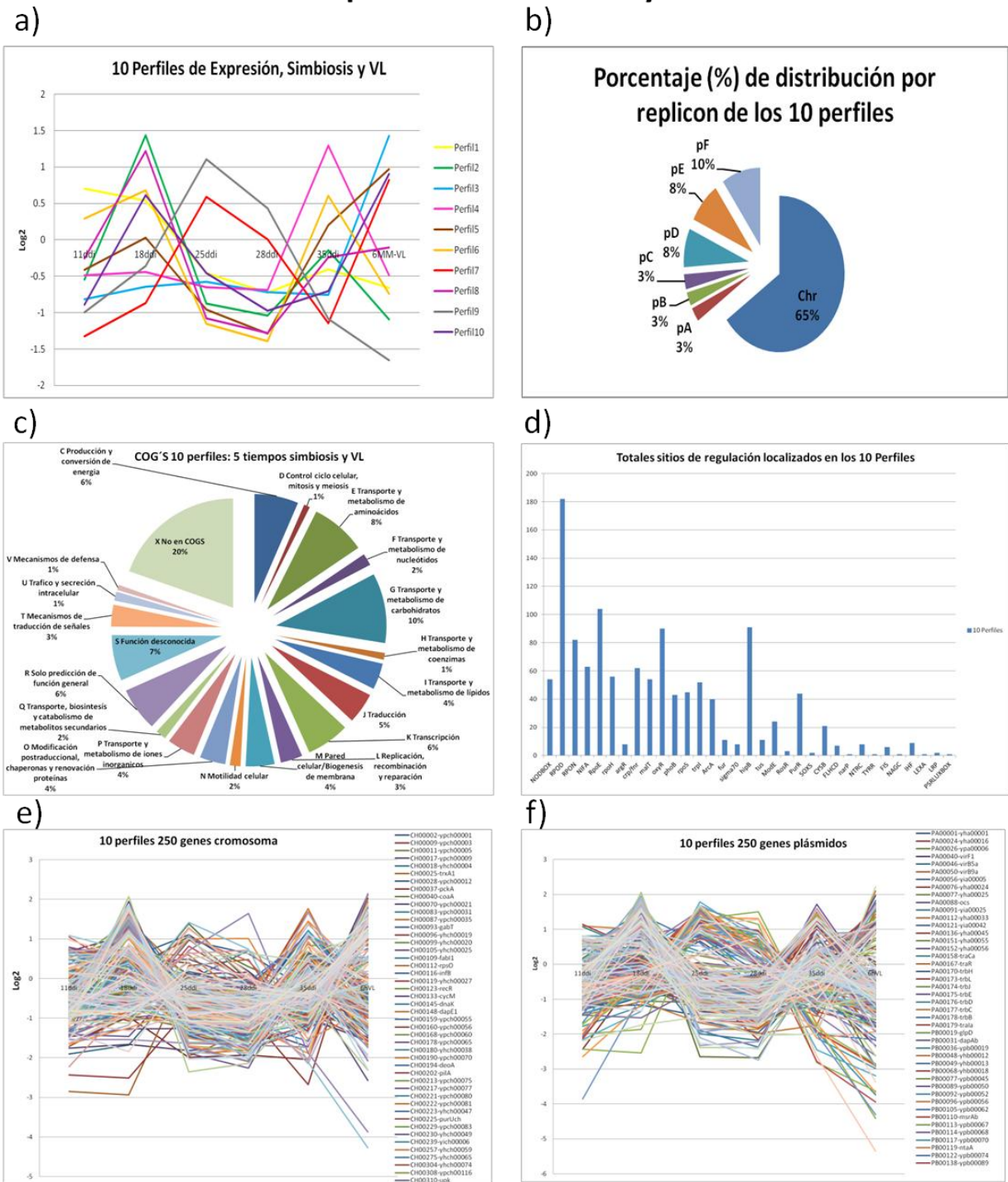


Figura 24. Perfiles de expresión del genoma de *R. etli* CFN42, presentes en los 5 puntos de la simbiosis y un tiempo de vida libre, a) 10 perfiles, b) distribución por replicón de los 898 genes, c) distribución de COG's, d) sitios de regulación localizados, e) tendencia global de 250 del genes cromosoma y f) tendencia global de 250 genes de plásmidos.

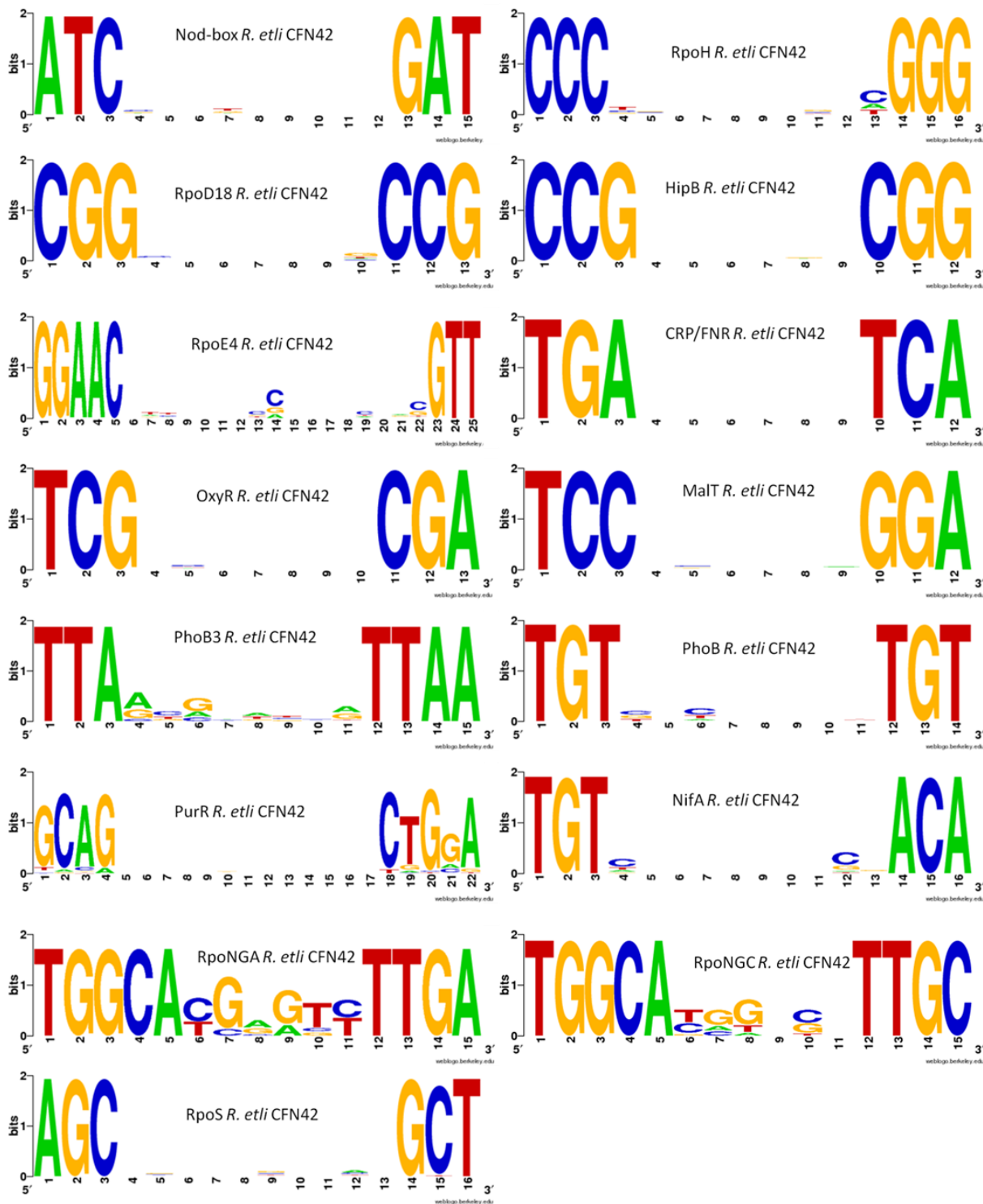


Figura 25. Consensos de los probables sitios de pegado para los diferentes reguladores y factores sigma localizados en los genes que conforman los 10 perfiles. En esta figura se presentan los sitios localizados con mayor frecuencia. La búsqueda de sitios se realizó con el programa RSA-Tools ver sección de material y métodos. Para obtener la imagen de los consensos se utilizó el programa WebLogo (Crooks *et al.*, 2004).

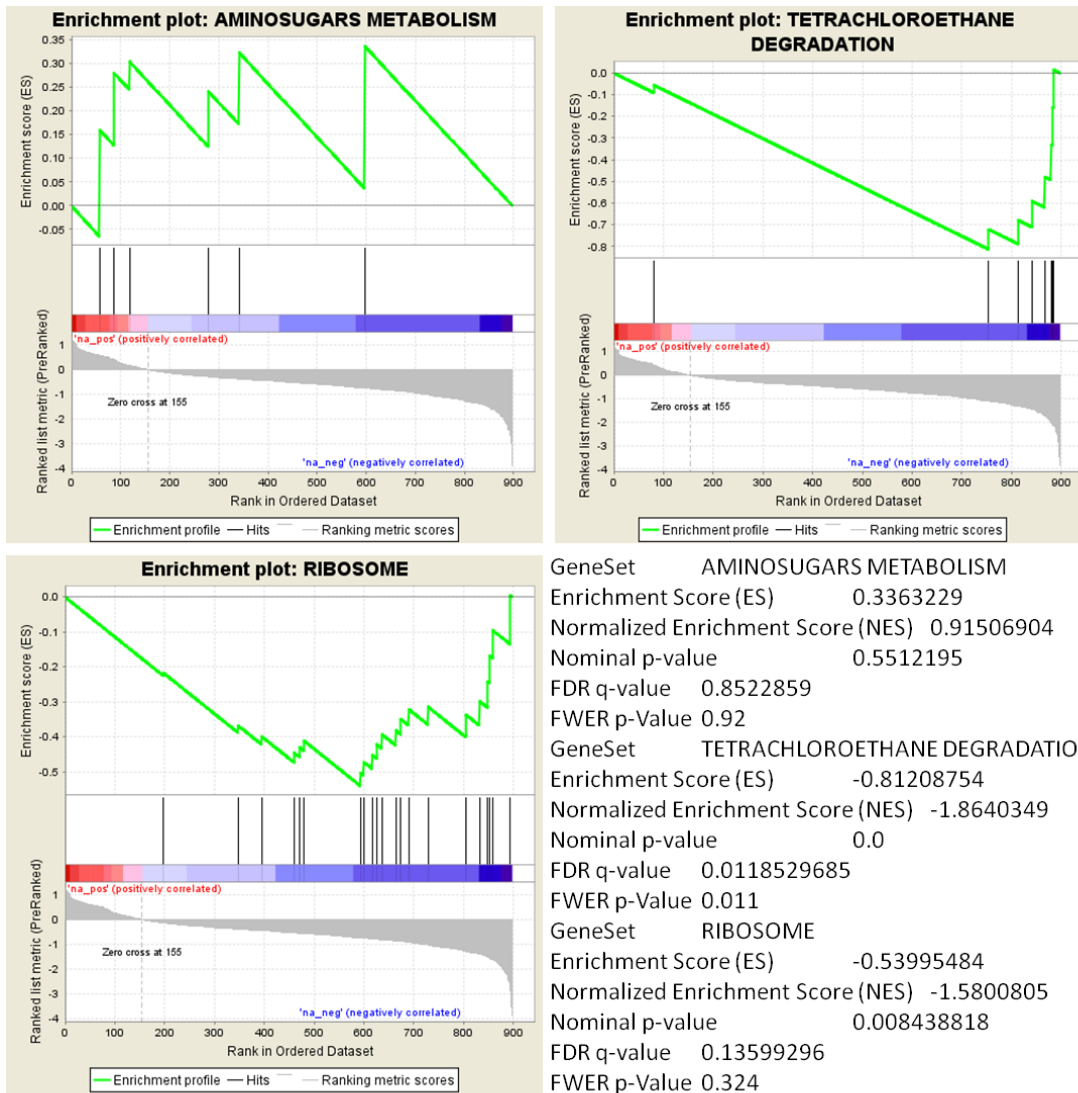


Figura 26. Análisis de los genes del tiempo 11 ddi, utilizando el programa GSEA, enriquecidos positivamente (metabolismo de amino-azúcares) y negativamente (degradación de tetracloroetano y ribosoma).

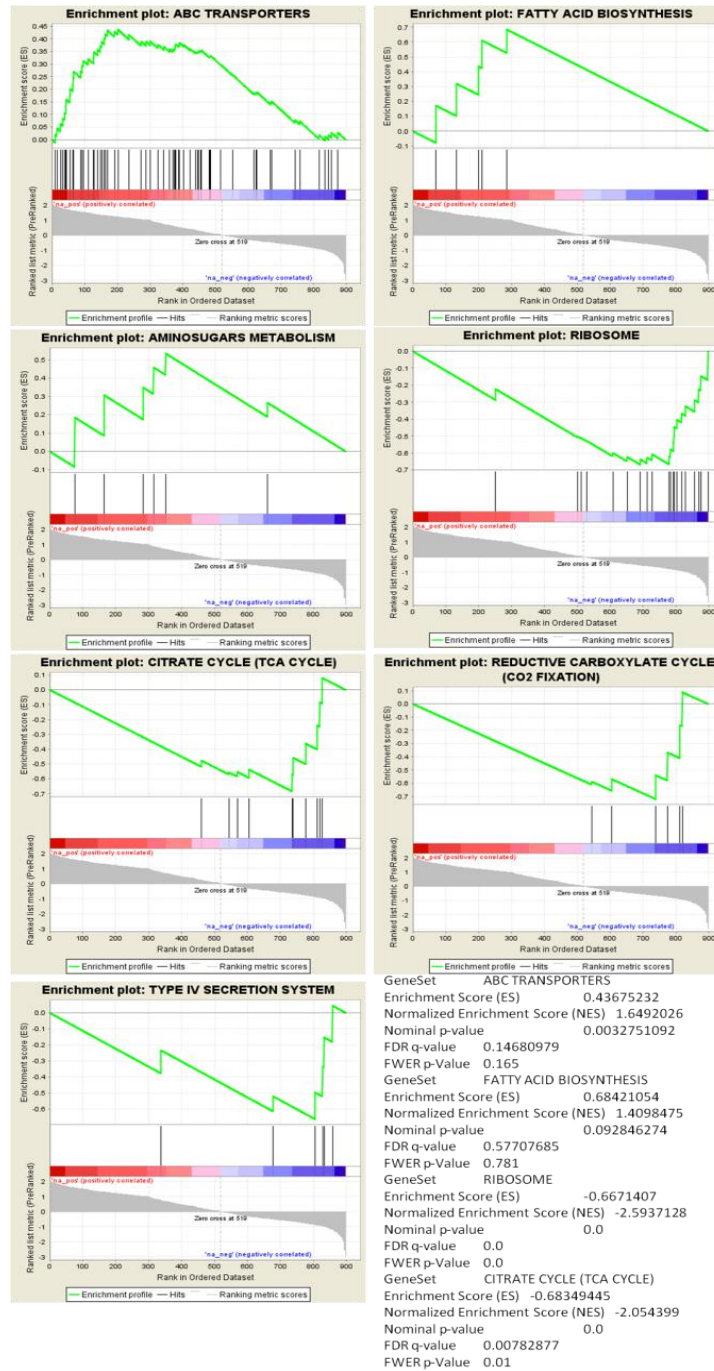


Figura 27. Análisis de los genes del tiempo 18 ddi, utilizando el programa GSEA; enriquecidos positivamente (transportadores tipo ABC, biosíntesis de ácidos grasos, metabolismo de amino-azúcares) y negativamente (ribosoma, ciclo TCA, fijación de CO₂, sistema de secreción tipo IV).

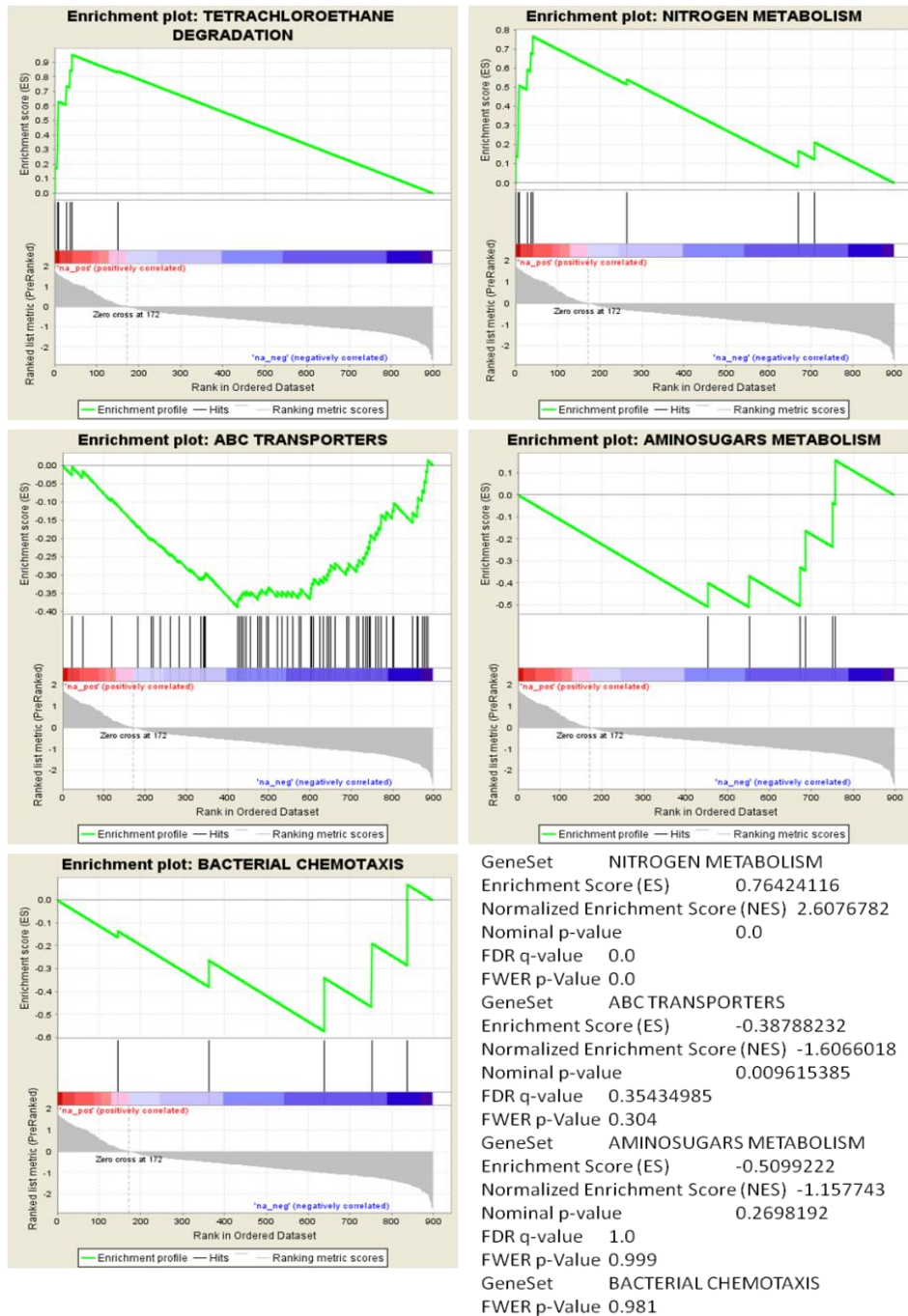


Figura 28. Análisis de los genes del tiempo 25 ddi, utilizando el programa GSEA; enriquecidos positivamente (metabolismo de nitrógeno, degradación de tetracloroetano) y negativamente (metabolismo de amino-azúcares, transportadores tipo ABC, quimiotaxis).

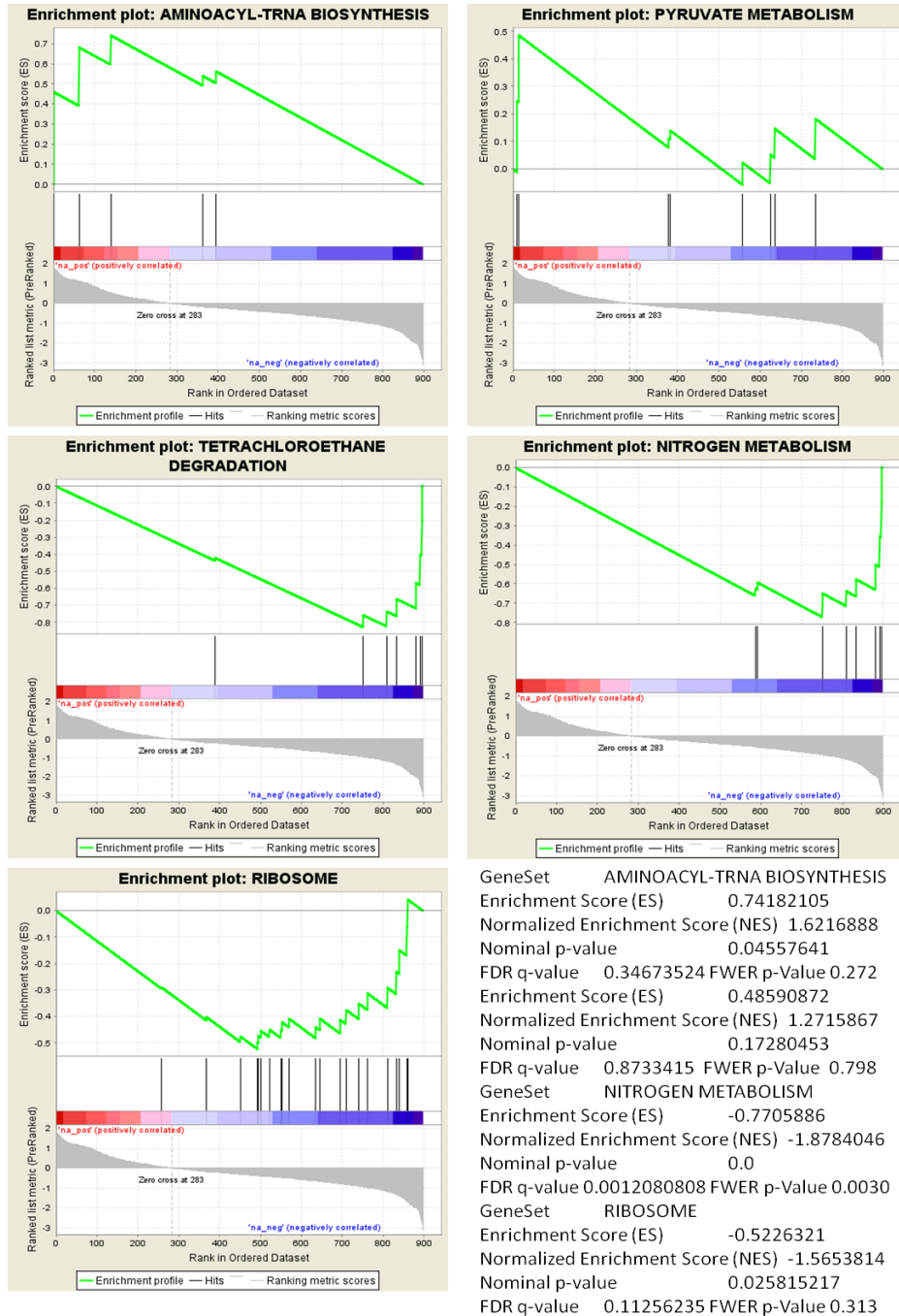


Figura 29. Análisis de los genes del tiempo 35 ddi, utilizando el programa GSEA; enriquecidos positivamente (biosíntesis de amino-acil-tRNAs, metabolismo de piruvato) y negativamente (metabolismo de nitrógeno, degradación de tetracloroetano, ribosoma).

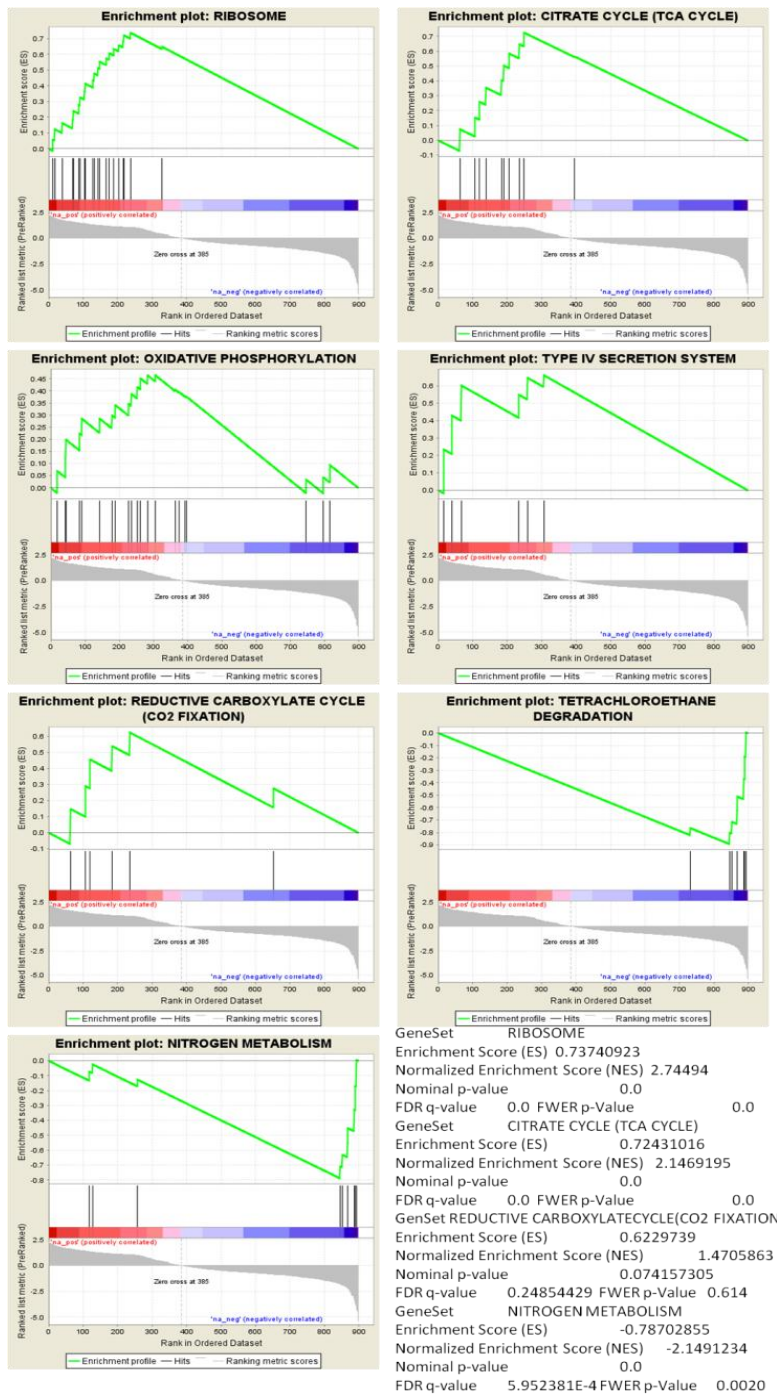


Figura 30. Análisis de los genes de vida libre (VL), utilizando el programa GSEA; enriquecidos positivamente (ribosoma, ciclo TCA, fosforilación oxidativa, sistema de secreción tipo IV, fijación de CO₂) y negativamente (metabolismo de nitrógeno, degradación de tetracloroetano, ribosoma).

12. DISCUSIÓN Y CONCLUSIONES

Los microorganismos en la mayoría de los nichos ecológicos están constantemente expuestos a las variaciones de múltiples factores ambientales entre éstos temperatura, oxígeno, disponibilidad de nutrientes, agua, presencia de compuestos tóxicos y la interacción con otros organismos incluyendo la simbiosis; los cambios en los patrones de expresión es la mayor respuesta adaptativa a éstas y otras variaciones. La expresión transcripcional es controlada por una variedad de reguladores transcripcionales entre los que destacan los factores sigma. Las bacterias diazótrofes, fijadoras de nitrógeno, presentan un genoma complejo, cuya regulación ha sido poco caracterizada. Nuestro modelo de estudio *R. etli* CFN42 no es la excepción. Su genoma posee un cromosoma constituido por alrededor del 67% de los genes, en tanto los plásmidos, que van de un tamaño de 184 Kb (pB) a 642 kb (pF), contienen el 33% restante. El número de genes codificados por el genoma dedicados a la transcripción y su regulación es de 479, de éstos 318 se localizan en el cromosoma y 161 en plásmidos (González *et al.*, 2006). La información sobre los mecanismos de regulación y análisis del comportamiento de la expresión de sus componentes y la interacción que existe entre ellos ha sido escasa, en donde destacan los resultados del primer regulón RpoE en Rhizobiaceas, trabajo donde participamos en colaboración (Martínez-Salazar *et al.*, 2009); posteriormente, en enero del 2010 se presentó un “Tiling array” de *R. etli* CFN42 cuyo objetivo principal consistió en identificar ncRNAs (RNAs no codificantes) expresados en la bacteria (Vercruyse *et al.*, 2010). Uno de los principales objetivos en los análisis de transcriptómicos es identificar los genes cuyos productos interactúan entre sí y que presentan un perfil transcripcional similar, aunque en ocasiones esto no siempre sucede, por lo que es necesario modificar y buscar nuevas estrategias para obtener el máximo rendimiento de éstas metodologías. El microarreglo de oligonucleótidos diseñado y construido para *R. etli* CFN42 y presentado en este trabajo, es una excelente herramienta para definir estimulones y conocer los mecanismos de regulación que sufre la bacteria, así como los niveles de expresión de los componentes del genoma durante la vida libre y simbiosis. Uno de los principales aportes de este trabajo fue la definición del regulón RpoN-NifA en vida libre (microaerobiosis) y

simbiosis, utilizando para ello diversas herramientas que van desde el análisis *in silico*, el análisis de los transcritos y la evaluación mediante proteómica. Previamente se había definido el regulón RpoN-NifA en *B. japonicum* pero sólo bajo condiciones de vida libre y utilizando un análisis *in silico* y la metodología de microarreglos (Hauser *et al.*, 2007).

Como ya se ha mencionado, NifA es un regulador central e importante en la fijación biológica de nitrógeno, el cual es un activador transcripcional que funciona a baja concentración de oxígeno y en conjunto con el sigma alternativo RpoN. La concentración de oxígeno en las rizobias se sensa a través de FixL, el cual contiene el dominio funcional PAS; este tipo de dominios se localiza en un amplio número de sensores que responden a oxígeno, potencial redox, luz y voltaje. FixL se auto fosforila y transfiere el grupo fosforil al dominio correspondiente del regulador FixJ, que a su vez activa la transcripción de los genes que participan en la fijación biológica de nitrógeno. Entre los genes regulados por FixJ, en *S. meliloti*, se encuentra *nifA*. El modo en que se regula es diferente a lo largo de las diferentes cepas y especies. Por ejemplo en *A. caulinodans*, *nifA* es regulado por FixK, un intermediario regulado a su vez por FixJ (Kaminski y Elmerich, 1998). En *R. etli* CNPAF512 la expresión de *nifA* es independiente de FixL, no obstante en microaerobiosis el gen *nifH* es dependiente del sistema FixLJ (D'Hooghe *et al.*, 1995). En *R. etli* CFN42, mutantes en *fixL*, *fixKf*, *fixKd*, *fnrNch*, *fnrNd* no tienen un efecto significativo en la fijación simbiótica de nitrógeno, aunque en la triple mutante *fnrNch*, *fnrNd* y *fixL* se reduce drásticamente (López *et al.*, 2001). En *R. etli*, aún no se han identificado elementos genéticos en la regulación de *nifA*, aunque mediante fusiones transcripcionales *nifA-gusA* se observó que la expresión no es exclusiva de la microaerobiosis (Michiels *et al.*, 1994). Además se ha reportado que la expresión de *nifA* aumenta alrededor de 3 veces en condiciones simbióticas (Benhassine *et al.*, 2007), lo que hace suponer que existe un mecanismo regulatorio que se activa en la simbiosis. Así, el análisis realizado en el presente trabajo acerca del regulón RpoN-NifA en simbiosis, nos aportó una mayor cantidad de elementos respecto a la vida libre en microaerobiosis, la mayoría localizados en el pSim lo que nos hace proponer que para identificar el o los reguladores de *nifA* en *R.*

etli CFN42 tenemos que buscar en las proteínas y metabolitos expresados en el bacteroide o en el simbiosoma.

El regulón RpoN-NifA de *R. etli* CFN42, que definimos en este trabajo (Salazar *et al.*, 2010), está constituido de 79 elementos y se complementó mediante el análisis de los perfiles obtenidos durante las diferentes etapas de la simbiosis, como ejemplo mencionaremos el gen *ypd00010* con ID PD00035. En el análisis del regulón NifA-RpoN (Salazar *et al.*, 2010) no se detectó su expresión en condiciones microaeróbicas, ni en simbiosis a los 11 ddi. Inclusive en el análisis entre vida libre y simbiosis, no se observó expresión a los 18 ddi (Resendis-Antonio *et al.*, 2011), pero sí observamos su expresión a los 35 ddi. En ensayos de expresión en nódulos de alfalfa observan la expresión de *nifA* y *fixK* en tiempos tardíos (incluye los 35 ddi), y no observan la expresión de *nifH* (Timmers *et al.*, 2000). En nuestros resultados observamos que a los 35 ddi la expresión de *nifH* es la más baja de la simbiosis. Por lo cual, proponemos que RpoN y NifA en este tiempo aún son funcionales pero ya no son los principales reguladores debido en parte al gradiente de O₂. El análisis del regulón NifA-RpoN lo realizamos a los 11 ddi, puesto que en la mutante *nifA* los nódulos y por ende los bacteroides no son viables a tiempos más lejanos. En nódulos de alfalfa inoculados con mutantes de *S. meliloti* que presentan fenotipo Fix-, la senescencia inicia mucho antes (entre los 14 y 21 ddi), que en nódulos inoculados con la cepa silvestre, cuya senescencia inicia entre los 28 y 35 ddi (Vasse *et al.*, 1990). Por otro lado, en el artículo de Salazar y col., discutimos sobre la expresión del gen *yhd00040* con ID PD00079 que fue detectado en microaerobiosis; sin embargo en simbiosis utilizando el microarreglo como herramienta, no observamos un cambio de expresión de este gen al comparar la cepa silvestre y la mutante *nifA* a los 11 ddi. El estudio que presentamos sobre los perfiles nos ofrece una explicación al respecto, al igual que en el estudio de expresión del regulón RpoN-NifA en simbiosis, el gen *yhd00040* a los 11 ddi presenta la más baja señal de los 5 puntos de simbiosis, incluso está cerca del nivel de expresión de la condición de VL (MM). Al igual que la mayoría de los genes del regulón RpoN-NifA, la señal máxima del gen *yhd00040* la observamos a los 25 ddi (perfil 9). Con la finalidad de buscar una explicación de este comportamiento en la expresión, observamos que en el análisis de búsqueda de

sitios de regulación (además de los sitios para RpoN y NifA), se identificaron sitios de unión para ModE y RpoD. El sitio para este último regulador es poco representativo del consenso, puesto que de las 2 regiones inverso repetidas que reconoce sólo se usó una como anzuelo para la búsqueda (Li *et al.*, 2002). Otro ejemplo de este tipo, es el gen *flgKch* con ID CH00680, el cual, al igual que para el gen *yhd00040*, no detectamos expresión diferencial a los 11 ddi, pero en el análisis de los perfiles observamos que el segundo punto más bajo de expresión de este gen se presenta a los 11 ddi (-1.45), sólo por debajo del de VL (-1.71); en cambio, su punto máximo se detecta a los 18 ddi (perfil 2), en el mismo perfil se expresa el gen *ypf00267* con el ID PF00546, detectado en el regulón NifA-RpoN de vida libre en microaerobiosis, sólo que a diferencia de *flgKch*, este alcanza su punto más bajo a los 28 ddi (-1.33), en VL (-1.31) y 25 ddi (-1.19). Los datos anteriores nos permiten sugerir a éstos cuatro genes como blancos para mutagenesis o para sobreexpresarlos y de esta manera conocer su papel en la simbiosis; de particular interés resulta el gen *yhd00040* que si bien en su descripción sólo se indica que es una proteína hipotética, es parte de la clase funcional I-Transporte y metabolismo de lípidos. Este tipo de proteínas pudieran ser relevantes al proveer precursores para membrana, así como en la señalización y traducción de señales (Resendis-Antonio *et al.*, 2011). En nuestro análisis en diferentes tiempos de la simbiosis, los otros 3 genes que son parte del regulón RpoN-NifA definidos en microaerobiosis: *thrB*, *ype00167* e *yhch00663*, en nuestro análisis en diferentes tiempos de la simbiosis no se ubican en ninguno de los 10 perfiles, por lo que asumimos que son exclusivos del regulón RpoN-NifA en vida libre (microaerobiosis) o bien su expresión en simbiosis es muy baja, y no se detecta debido al corte que establecimos.

Algunos reportes previos señalan que la baja concentración de oxígeno, si bien mimetiza una parte de lo que sucede en el nódulo, no es el único elemento que participa en la expresión de los genes simbióticos. Por ejemplo, en *M. loti* sólo un grupo de 50 genes se comparten entre la simbiosis y la microaerobiosis (Uchiumi *et al.*, 2004). En *B. japonicum*, en microaerobiosis, no se detectaron varios de los genes reportados como dependientes de RpoN-NifA (Bobik *et al.*, 2006). Por otro lado, es relevante la cantidad de genes que son parte del regulón NifA-RpoN, tanto en *B. japonicum* como en *R. etli*, y que no están

vinculados al menos directamente con la fijación de nitrógeno (Hauser *et al.*, 2007; Salazar *et al.*, 2010). En particular en los experimentos en microaerobiosis de la mutante *nifA*, comparando su expresión con la cepa silvestre, observamos 39 transcritos que alteran su expresión positivamente en la mutante, y de los 41 elementos que bajan su expresión, 16 de ellos no están vinculados con la fijación de nitrógeno, ni se les localizaron sitios de unión para NifA-RpoN. Existe un trabajo previo, en *S. meliloti*, donde la interrupción del gen *nifA* tiene un efecto sobre varios procesos celulares (Gong *et al.*, 2007). En este artículo indican que NifA tiene un papel regulatorio fuera de la simbiosis. Lo que podemos concluir es que en el caso de *R. etli* NifA tiene un papel tanto en la vida libre como en la simbiosis, más amplio e importante de lo esperado al inicio de este trabajo. Un dato que sustenta nuestra conclusión, lo encontramos en el trabajo de Hernández-Gómez y colaboradores 2010, en donde indican que en la mutante *nifA* el gen *nirK* es sobreexpresado. La carencia de un regulador transcripcional con diversos blancos puede tener un efecto más general al esperado, ya que si este regulador afecta otro regulador y así sucesivamente, se presenta lo que se conoce como una regulación en cascada, dificultando la posibilidad de identificar blancos específicos. En el regulón RpoN-NifA sólo identificamos un probable sensor de la familia Box/GGDEF; en el análisis de los perfiles este gen con ID PD00105 se localiza en el perfil 9, que corresponde a los genes expresados a los 25 ddi y en menor medida a los 28 ddi. En este perfil se encuentran los genes que participan en la fijación de nitrógeno y los previamente reportados como parte del regulón RpoN-NifA (Salazar *et al.*, 2010).

Los genes expresados por perfil nos indican una congruencia; por ejemplo, en todos los perfiles los tiempos 25 y 28 ddi se ubican prácticamente en la misma posición (ver inciso E de las figuras 14 a la 23 y los dendrogramas del agrupamiento jerárquico en las figuras 31, 32 y 33), la distancia que existe entre estos tiempos de la fijación de nitrógeno son 72 hrs, a diferencia de las 168 hrs que existe entre el resto de los tiempos. Con ello concluimos que el utilizar estos tiempos cercanos fue relevante ya que le dio certeza a los resultados presentados y discutidos en esta tesis. En los perfiles en los cuales están involucrados genes que se expresan en vida libre, se tienen los mayores porcentajes de elementos que

pertenecen al cromosoma, como en los perfiles: 7 (82%), 10 (77%) y el perfil 3 (75%). Revisando los grupos funcionales, tenemos que también los genes que pertenecen a los COG'S J-Traducción, y O-Modificación postraducciona, chaperonas y renovación de proteínas, están sobrerrepresentados en los perfiles donde se ubican los genes con una expresión alta en VL, como son el perfil 3, perfil 5 y perfil 7 (Figuras 16, 18, 20, inciso d y 34 inciso A). Otro ejemplo se observa en la figura 30, en donde los genes que codifican para proteínas ribosomales se enriquecen significativamente en la vida libre. Estos resultados nos permiten concluir que los datos analizados se acercan en gran medida a lo sucedido en la regulación de la expresión del genoma de *R. etli* CFN42, en los 5 tiempos de la simbiosis y en la vida libre. Datos que en el resto del análisis se fortalecen con la información recabada de los genes analizados individualmente y por el perfil donde se ubican, que incluye además los sitios de regulación localizados en éstos.

Consideramos que el análisis realizado mediante la comparación entre diferentes tiempos de la fijación de nitrógeno, es relevante porque no sólo se identificaron los genes que modifican su expresión en la vida libre o en la simbiosis, sino que además esta estrategia permitió localizar genes que se comparten entre la vida libre y algún punto de la simbiosis; así como genes que se comparten entre los diferentes tiempos de la simbiosis. Logramos también determinar algunos factores involucrados en la temporalidad de la expresión y logramos acercarnos a la identificación de probables sitios de pegado para diferentes reguladores transcripcionales, varios de éstos pobremente caracterizados en *R. etli* (Fig. 25). Con relación a la regulación observada en los 898 genes que constituyen los 10 perfiles, podemos proponer la participación de 47 reguladores o probables reguladores transcripcionales que pertenecen a diversas familias. De manera similar a nuestros resultados, en *B. japonicum* fueron identificados 39 reguladores transcripcionales inducidos en la simbiosis (21 ddi) comparando con la vida libre (Pessi *et al.*, 2007). En *S. meliloti* en simbiosis se identificaron sólo 15 reguladores transcripcionales, aunque el total de genes que modificaron su expresión fue de solamente 109 y debemos considerar que los nódulos inducidos por *S. meliloti* son indeterminados (Capela *et al.*, 2006). De los 47 reguladores que identificamos en este análisis, 43 se expresan en algún punto de la

simbiosis y 4 son exclusivos de la vida libre. Los perfiles que sobresalen son los representados por genes expresados a los 18 ddi, tanto el perfil 2 como el perfil 8. En éstos dos perfiles se encuentran 25 de los 43 factores de transcripción expresados en algún punto de la simbiosis. Los dos perfiles totalizan 280 genes, 201 pertenecen al perfil 2 y 79 genes se ubican en el perfil 8. Ambos perfiles tienen su máxima expresión a los 18 ddi. La diferencia en expresión entre los genes de éstos perfiles se presenta en vida libre, en donde en el perfil 2 se observa el punto más bajo (represión), a diferencia del perfil 8 en donde en esa condición se observa el segundo punto más alto (ver figuras 15 y 21 inciso e); adicionalmente, es en el perfil 2 (18 ddi) en donde se localizan 18 de los 25 factores transcripcionales, así mismo, en donde se agrupan más genes (como referencia el perfil 3 contiene 175 genes). Un elemento a destacar son los 7 reguladores del perfil 2, que pertenecen al plásmido E, entre estos, dos de la familia GntR involucrados en la utilización de carbohidratos, aminoácidos y ácidos grasos, así como en el metabolismo de la biotina. En el análisis de expresión entre los 18 ddi y VL, el plásmido E presenta un comportamiento similar al del plásmido simbiótico (pD); en simbiosis, ambos replicones expresan un mayor número de genes en comparación con el número de elementos expresados en vida libre (Fig. 4). Para el plásmido E, éstos 7 reguladores podrían ser la razón del incremento en la expresión en simbiosis.

Por otro lado, en *S. meliloti* se identificó un sistema tipo CRP, en donde están involucradas tres adenilato ciclasas (Tian *et al.*, 2012). En el perfil 2 (18 ddi), se localiza el transcrito de *cyaFch2*, éste codifica para una adenilato ciclasa que cataliza la síntesis de 3´5´-AMP cíclico (AMPC) a partir de ATP; además, en este mismo perfil el gen hipotético con ID CH02560 codifica también para una adenilato ciclasa. Incluso en el perfil 6 (18, 35 y 11 ddi), pero también con una mayor expresión a los 18 ddi, se observó expresado un gen (ID PE00177) el cual codifica para otra adenilato ciclasa. En simbiosis, el papel sugerido para las adenilato ciclasas es en respuesta a las señales de la planta. Las bacterias, a través de las adenilato ciclasas, tienen un papel activo en el control de la infección, e incluso es probable que el rizobio module la susceptibilidad de la planta a la infección (Tian *et al.*, 2012).

Los datos de la expresión en las diferentes etapas de la fijación, se relacionan directamente con la fisiología del nódulo de cada tiempo analizado. A los 11 ddi, se observa que la bacteria disminuye la expresión de la mayoría de sus elementos en comparación con otros tiempos, ya que sólo en el perfil 1 (48 genes) y en el perfil 6 (70 genes) se ubican los genes expresados en este tiempo; en el resto de los perfiles en el tiempo de 11 ddi se observa una represión (780 genes) en comparación con la expresión a los 18, 25, 28 y 35 ddi además de la condición de vida libre. También se observa una baja en la síntesis de compuestos que participan en la formación de membrana, además no se expresan genes que pudieran contender contra una respuesta de defensa por parte de la planta: en los perfiles 1 y 6, no hay elementos que representen el COG V-Mecanismos de defensa. Además alrededor de los 11 días inicia la expresión de la nitrogenasa (Puppo *et al.*, 2005; Valderrama *et al.*, 1996) y en consecuencia la fijación de nitrógeno, evitando que la planta sancione y termine con la simbiosis de manera temprana; es decir, suponemos que la bacteria utiliza este tipo de señales (envío de amonio y disminución de exopolisacáridos) para que la planta no active sus mecanismos de defensa y emita compuestos tóxicos a la bacteria. El perfil 1, como mencionamos anteriormente incluye los genes expresados a los 11 ddi, está influenciado aún por la regulación del activador transcripcional de la familia LysR, la proteína NodD, que tiene un papel fundamental en la regulación de los genes *nod*. Aún en ausencia de flavonoides, la proteína NodD se une a la región promotora aunque la transcripción se activa ante la presencia de flavonoides compatibles (Fisher y Long, 1993). En este perfil, se ubica un gen con ID CH03353 que codifica para un regulador tipo LysR, el cual además en la región promotora presenta una caja Nod. En este perfil se ubican los genes *exoR* y *exsB*, los cuales tienen un efecto negativo en la síntesis de exopolisacáridos (EPS) (Reeve *et al.*, 1997; Becker *et al.*, 1995). En particular, el producto de *exoR* presenta un claro fenotipo simbiótico, tanto *R. leguminosarum* como *S. meliloti* inician la infección pero esta no progresa (Reeve *et al.*, 1997; Ozga *et al.*, 1994) por lo que es un buen blanco a mutagenizar en *R. etli* CFN42. Finalmente, ambos genes presentan en sus regiones promotoras probables cajas Nod. En el perfil 1, el segundo tipo de región promotora en importancia, por el número de sitios

localizados, es para RpoH, un factor sigma alternativo de la familia Sigma 70 que se ha caracterizado para contender contra el estrés por calor o choque térmico, pero también tiene un papel durante los estreses por pH y oxidativo. En *R. etli* CFN42 se han identificado dos genes *rpoH*, ambos complementan la mutante *rpoH* de *E. coli*. La doble mutante y la mutante en *rpoH1* pueden inducir nódulos pero inician prematuramente la senescencia (a los 18 ddi), la actividad de la nitrogenasa se reduce, así como la viabilidad de las células (Martínez-Salazar *et al.*, 2009). Si bien varios de los blancos típicos como *groEL*, o genes que codifican para proteasas no están incluidos en estos 48 genes, la influencia de la regulación por RpoH puede observarse en este perfil con 5 probables blancos para este factor sigma, en donde 2 de ellos codifican para transportadores tipo ABC. La propuesta en este punto es que aún existe influencia de regulación, a través de NodD o algún miembro de la familia LysR, además de reguladores negativos para EPS y se presenta una regulación por medio de RpoH (Esquema del tiempo 11 ddi ver figura 34 inciso B).

A los 18 ddi ocurren la mayor cantidad de cambios en la expresión del genoma en simbiosis. Los principales cambios en expresión recaen en genes que codifican para productos que incrementaran la cantidad de compuestos de carbono en el bacteroide como transportadores tipo ABC, además de algunos genes clave para la transformación de bacteria a bacteroide. En este punto probablemente exista una población de bacterias que no se diferencia y por esta razón detectamos la expresión de genes involucrados en la síntesis de flagelo (Tabla 5 y Figuras 11, 27). Se ha reportado en nódulos de diferentes especies de leguminosas la presencia de bacterias vegetativas (no diferenciadas), y éstas se localizan en los espacios intercelulares (Timmers *et al.*, 2000). Se han reportado ejemplos en alfalfa (Paau *et al.*, 1980; Thornton 1930; Vance *et al.*, 1980), haba (Kijne 1975), trébol (Mosse 1964) y *Lotus corniculatus* (Vance *et al.*, 1982). A los 18 ddi, resalta el COG D-Motilidad celular, además mediante GSEA el grupo Ensamble de Flagelo se enriquece positivamente en la condición de simbiosis. En ambos análisis (comparación VL-Simbiosis 18 ddi y análisis de los 5 tiempos de simbiosis y un punto de VL) un antecedente se presenta en *R. etli* donde se reportó en nódulos de 21 ddi la expresión del gen *flaF* que codifica para un regulador de la biosíntesis de flagelo (Vercruyssen *et al.*, 2011). La

hipótesis sugerida es que en este punto (18 ddi), los bacteroides están a punto de alcanzar su maduración pero la parte de bacterias que abortó o bien nunca inició la diferenciación se mantienen en el nódulo o alrededor de este y es por eso que identificamos este tipo de genes en este punto particular. La figura 35 muestra un corte de nódulos a los 18 ddi y utilizando el método “Live/Dead BacLight (BL) Viability kit”, en donde se pueden observar algunas bacterias alrededor del nódulo. El método de extracción de RNA que utilizamos no considera una purificación previa de bacteroides, por lo que el RNA puede contener mensajeros de estas bacterias, y es probable que los RNA mensajeros sean específicos de este tiempo, puesto que en el resto de los tiempos analizados en simbiosis no detectamos señal para este tipo de genes. Hay una relación inversa entre la concentración de EPS y la transcripción de genes de la síntesis de flagelo, en donde a una alta concentración de EPS la expresión de elementos de flagelo disminuyen y cuando la concentración de EPS disminuye se activa la expresión de genes de la síntesis de flagelo. Cabe señalar que en el perfil 1 (11 y 18 ddi) se observa la expresión de 2 reguladores negativos para la síntesis de exopolisacáridos, lo que indicaría que la concentración de EPS es baja, incluso como se menciona párrafos arriba la mutante de uno de ellos, *exoR* en *S. meliloti*, resulta en la pérdida de flagelo y por lo tanto de motilidad. Los autores proponen que ExoR tiene un efecto directo en la regulación de la biosíntesis de flagelo adicional a la regulación de la biosíntesis de succinoglicanos (Yao *et al.*, 2004).

Adicionalmente en este tiempo de la fijación de nitrógeno observamos la máxima concentración de transportadores tipo ABC y la mayor cantidad de familias de reguladores expresados, por lo cual proponemos que en este punto es en donde maduran los bacteroides y se generan las condiciones para incrementar la fijación de nitrógeno. Otro dato a destacar en el tiempo 18 ddi, son los elementos involucrados en regulación así como los sitios de unión identificados, donde ambos aumentan ostensiblemente. Los reguladores en este punto son diversos. La familia más abundante es la GntR cuyos integrantes participan como represores de blancos involucrados en una diversidad de procesos biológicos. En total en el perfil 2 (18 ddi) son 5 reguladores de la familia GntR detectados, que incluyen a HutC. En *S. meliloti*, mutantes en los genes *gtrA* y *gtrB*

miembros de la familia GntR, tienen una baja motilidad celular y son deficientes en la simbiosis, con nódulos pequeños, redondos, ligeramente rosas y se localizan en las raíces laterales. Además, en las tres a cuatro semanas posteriores a la inoculación, las plantas inoculadas con las mutantes lucen atrofiadas y con una disminución en la actividad de la nitrogenasa, pero con un remarcado incremento en la cantidad de nódulos respecto a la cepa silvestre. Por último, las mutantes reducen su capacidad de competencia para la nodulación (Wang *et al.*, 2008). Otros reguladores presentes en este tiempo pertenecen a la familia TetR, sus blancos participan en la resistencia a antibióticos, y en el metabolismo de la pirimidinas, ácidos grasos y biotina. La familia LuxR, cuyos blancos están involucrados en la competencia, el “quorum sensing”, el metabolismo de los óxidos de nitrógeno y en el “switch” anaeróbico. Ambas familias presentan 2 elementos en este perfil. El factor sigma que presenta un número importante de sitios o blancos es RpoE, con 19 sitios localizados. Entre los probables blancos de RpoE destacan el gen *cyaFch2* el cuál codifica para una adenilato ciclasa que cataliza la síntesis de 3´5´-AMP cíclico (AMPC) de ATP; el gen *aceC* que codifica para una alfa-cellobiosil-difosfoprenil D-Manosa transferasa y junto con *aceB* participan catalizando el segundo y tercer paso de la vía biosintética del exopolisacárido acetano en *X. camprestis* (Griffin *et al.*, 1996); el gen *dapAb*, codifica para una dihidrodipicolinato sintetasa, en *S. meliloti* se expresa constitutivamente y está relacionado con la biosíntesis de lisina (García-Rodríguez *et al.*, 2000); el gen *flgE* codifica para una proteína del gancho flagelar. En el mismo sentido el gen hipotético con ID CH02245 que codifica para una proteína de varilla de cuerpo basal del flagelo también presenta un sitio para RpoE, y el gen *nuoB* que codifica para la subunidad B de una NADH deshidrogenasa. Otro sitio sobrerrepresentado en este perfil es para HipB, el cual forma un complejo con HipA para regular genes involucrados en la síntesis de péptidoglicanos y en la síntesis de DNA (Black *et al.*, 1994). En *E. coli* HipB-HipA participan en persistencia celular y en la resistencia a antibióticos (Schumacher *et al.*, 2009), aunque *R. etli* CFN42 no tiene anotados los genes *hipB* ni *hipA*. En el pC se localiza un gen hipotético con ID PC00094 que codifica, al igual que *hipA*, para una treonina/serina cinasa, el ID contiguo a PC00094 codifica para una proteína NodW (un regulador de repuesta de la nodulación). A

los 18 días localizamos sitios para RpoN (5) y NifA (13), pero ningún blanco comparte ambos, aunque RpoN podría funcionar con otra EBP como NtrC, cuyo transcrito se ubicó en el perfil 5 (18 y 35 ddi). Además identificamos sitios para un regulador Fis y uno TyrR que comparte el blanco con RpoN. La familia Fis abarca los reguladores TyrR, NifA, NtrC y otros. Este perfil es el más diverso en reguladores transcripcionales, y prueba de ello se observa al comparar el COG K-Transcripción, donde el porcentaje general de los 10 perfiles es de 6%, en contraste en el perfil 2 llega al 10%, un porcentaje similar al perfil 8 (18 ddi). El perfil 2 se complementa con el perfil 8, ambos aumentan el número de elementos involucrados en la transcripción. En ambos perfiles el sitio de unión a FNR aumenta, con 10 sitios en el perfil 8 y 9 sitios en el perfil 2. En *R. etli*, una triple mutante *fixL-fnrNd-fnrNchr* disminuye considerablemente la fijación de nitrógeno en todos los tiempos colectados, y la doble mutante *fnrNd-fnrNchr* presenta una reducción de poco más de la mitad respecto a la cepa silvestre a partir del tercer tiempo de colecta (López *et al.*, 2001). El fenotipo similar entre la doble mutante *fnrNd-fnrNchr* y la cepa silvestre en los nódulos jóvenes, es congruente con el aumento de los sitios FNR identificados en este trabajo a partir del segundo punto de colecta. Nosotros proponemos que es cuando se presentan las condiciones necesarias para la máxima expresión de los genes dependientes de FNR. Una de estas condiciones sería la concentración de oxígeno. A los 18 días, los sitios para HipB están entre los más representados, pero en cuanto a sitios para RpoE disminuyen de 19 sitios en el perfil 2 a sólo 4 en el perfil 8; probablemente a RpoE se deba que los genes de los perfiles 2 y 8 se distribuyan en diferentes grupos a pesar de presentar su máxima expresión a los 18 ddi, incluso en el perfil 2 observamos la expresión de un gen que es una de las cuatro copias de *rpoE* (PF00052) en el genoma de *R. etli*, en particular este se ubica en el pF. La comparación del consenso (“weblogo”) generada para los 19 sitios de RpoE del perfil 2 y la generada por los 4 sitios de RpoE del perfil 8, difieren considerablemente en las bases de las posiciones 2 y 3, donde la G del perfil 2 esta disminuida y la A de la posición 3 en el perfil 8 no existe (Fig. 34 inciso C). Como ya mencionamos, el segundo punto más alto en el perfil 2 es 35 ddi, a diferencia del perfil 8 donde el punto que representa MM es el segundo lugar en expresión. Al igual que en el

perfil 2, el perfil 8 presenta un número elevado de reguladores, entre los que destacan 2 de la familia MarR (ID's CH02389 y CH02727). Los reguladores de esta familia son positivos y negativos, presentan un amplio espectro de blancos de regulación que se distribuyen en diversos mecanismos principalmente en virulencia, resistencia a antibióticos, toma de zinc y utilización de compuestos aromáticos; adicionalmente, compite con otras proteínas para la unión a la RNA polimerasa y puede activar la transcripción (Ellison y Miller, 2006). De manera relevante, en el perfil 8 observamos la expresión de otro factor sigma ECF (PF00422) no obstante no se ha descrito su probable papel en *R. etli* ni otra bacteria relacionada, podría ser parte de la respuesta en simbiosis sobre genes cuyos productos estén involucrados en la envoltura celular.

La propuesta que tenemos como mecanismo de regulación en este tiempo es principalmente la concentración de oxígeno (la mayor cantidad de sitios FNR se localizan en este tiempo). La expresión también depende del factor sigma RpoE4, éste se induce por estrés osmótico, salino y oxidativo, sus blancos son diversos, y en total 38 genes dependen de la sobreexpresión de RpoE4, incluyendo un grupo de 16 genes que codifican para proteínas de envoltura celular (Martínez-Salazar *et al.*, 2009); otro factor transcripcional que influye a nivel de membrana es HipB, el complejo HipA-HipB está relacionado con la persistencia de una parte de la población. Finalmente, debido a que en este tiempo se concentran la mayoría de los reguladores transcripcionales, proponemos que a los 18 ddi es donde se presenta la mayor transformación de las bacterias (Esquema del tiempo 18 ddi ver figura 34 inciso C).

El tercer punto analizado (25 ddi), involucra un gran número de genes dependientes de RpoN-NifA. En nuestro análisis es donde se presenta una mayor expresión de la nitrogenasa y por ende de la fijación de nitrógeno, además se relaciona también con la máxima expresión en la simbiosis del gen *dctA*. En este punto proponemos que el bacteroide dedica sus componentes casi por completo a la fijación de nitrógeno. El tiempo o punto 4, genes de 28 ddi, es prácticamente una copia del tiempo 25 ddi, ya que se ubica, como mencionamos anteriormente, muy cercanos entre sí, aunque en el tiempo de 28 ddi observamos el inicio de la disminución en la expresión de los genes de la fijación de

nitrógeno y sólo algunos genes hipotéticos presentan una expresión mayor en este punto con respecto al punto de 25 ddi. Los productos de estos genes podrían ser señales claves del inicio de la decadencia de la fijación (ver el resumen del perfil 9). La fijación biológica de nitrógeno es un proceso altamente demandante de energía, y por lo tanto la expresión de los elementos que participan en este evento deben estar finamente controlados y no sólo en dependencia de la concentración de oxígeno en el medio. Basados en nuestros datos acerca del regulón RpoN-NifA en microaerobiosis y del regulón NifA en *B. japonicum* (Hauser *et al.*, 2006), así como en ensayos de expresión en *S. meliloti* 1021 en donde la expresión de los genes *nifHDK* y *fixABCX* fue observada en bacteroides a los 22 ddi y no en condiciones microaeróbicas (Becker *et al.*, 2004), concluimos que en nuestras condiciones en vida libre no fue posible obtener una señal significativa de varios de los genes previamente reportados dependientes de NifA, lo cual puede deberse a una baja expresión de *nifA* o bien a la regulación de algún otro elemento presente en el medio. Por ejemplo, en *S. meliloti* 1021 se reportó que en presencia de NH_4Cl el operón *nifHDK* no es expresado en condiciones microaeróbicas (Noonan *et al.*, 1992). El medio mínimo que nosotros utilizamos también contiene NH_4Cl que podría al igual que en *S. meliloti* tener un efecto adverso en la expresión de los genes de fijación de nitrógeno, en particular el operón *nifHDK*. En nuestros resultados, en el perfil 9 (perfil en donde se localizan la mayoría de los genes dependientes de RpoN-NifA), observamos que las regiones promotoras, además de contener los sitios de unión para RpoN y NifA, hay más sitios de unión para otras proteínas reguladoras. Podríamos concluir que los genes que pertenecen al regulón RpoN-NifA son altamente regulados transcripcionalmente por factores diferentes a RpoN y NifA. Los sitios detectados incluyen 9 cajas NOD, 52 sitios RpoD, 22 RpoE, 14 RpoH, 8 CRP/FNR, 11 MalT, 37 OxyR, 13 RpoS, 12 TrpI, 17 HipB, 3 sigma 70, 13 PurR, 9 CysB y 3 NtrC; en total se localizaron 301 sitios. El perfil 3 (VL) es el que sigue en número de sitios de unión detectados con 211, a pesar de que el perfil 3 dobla en número de genes al perfil 9, (178 y 86, respectivamente) (Figuras 16 y 22 inciso f). Basados en esta información, podemos concluir que existe una sobre regulación de los genes expresados en el punto donde se agrupan los genes del regulón RpoN-NifA. Por ejemplo, en *A.*

caulinodans, la regulación de los genes para la fijación de nitrógeno se modulan por la concentración de oxígeno y amonio (Michel-Reydellet y Kaminski., 1999), incluso una mutante en *phaC* (síntesis de PHB) tiene un efecto adverso en la expresión de la nitrogenasa (Mandon *et al.*, 1998).

A los 25 ddi, además se localizan varios de los genes reportados en *R. etli* como propios de la simbiosis (Vercruysse *et al.*, 2011). Entre éstos podemos mencionar a *azlC* (enzima de resistencia a azeleucina, transportador de cadenas ramificadas de aminoácidos), *casA* (unión a calcio) previamente su expresión era delimitada a bacteroides no maduros, y la mutante tiene un fenotipo en el desarrollo y la fijación de nitrógeno (Xi *et al.*, 2000), además de los genes *aptd*, *bacS*, el operón *fixABCX* y por supuesto *nifHDK*, entre otros. Nosotros observamos genes que los mismos autores consideran exclusivos de la simbiosis, compartidos con vida libre (perfil 7, 25 ddi y VL). Por ejemplo *dctA*, *ropA* y *nolRch*; sin embargo, es importante mencionar que en su análisis utilizaron sólo nódulos de 21 ddi (Vercruysse *et al.*, 2011). A los 25 ddi, como ya hemos señalado, observamos la máxima expresión de los genes relacionados directamente con la fijación biológica de nitrógeno, como los genes *fix* y *nif*. La expresión en este punto es controlada por RpoN-NifA, en donde los principales blancos son *nifH* copias 1, 2 y 3, que codifican para la nitrogenasa reductasa y los genes *nifD* y *nifK* codificantes de la nitrogenasa que forman parte de un operón con *nifH*. Los genes *cpxA* y *cpxP2* presentan ambos sitios en sus regiones promotoras, codifican para citocromos P450 monooxigenasas y a pesar de que en *B. japonicum* no presentan fenotipo simbiótico (Tully y Keister, 1993), es necesario evaluar su participación en la interacción *R. etli*-*P. vulgaris*. Para la apertura del complejo de transcripción, RpoN requiere de una proteína activadora ó EBP, y si bien localizamos blancos con ambos sitios también identificamos genes que en sus regiones promotoras está presente sólo el sitio para alguno de estos reguladores, ya sea para NifA o RpoN. Entre los blancos con sitios de unión a RpoN destacan: *osmC*, el cual en *E. coli* se induce por la presencia de sal, estudios de cristalografía de proteínas y datos adicionales en su estructura señalan que pudiera funcionar como una peroxiredoxina (Shin *et al.*, 2004). Este gen también fue observado en el análisis de Vercruysse y col., 2011. El gen *ctrE* que

codifica para una farnesil transferasa, la farnesilación es requerida para la interacción proteína-proteína y para la actividad de proteínas regulatorias clave como cinasas y Ras. Otro regulador con influencia en este punto es OxyR; este factor transcripcional es relevante para contender con las especies reactivas de oxígeno. Entre sus blancos en *R. etli* se encuentra la catalasa *katG* (Vargas *et al.*, 2003). A los 25 ddi, los blancos identificados para OxyR son: *alkA* el cual codifica para una DNA-3-metiladenina glicosidasa II, en *E. coli* se induce por la presencia de agentes alquilantes (Fernandez de Henestrosa y Barbé, 1991), el gen *mviN* que codifica para un factor de virulencia, el cual, su mutante en *R. etli* CIAT899 no presenta fenotipo simbiótico (O'Connell *et al.*, 1998), y el gen *coaA* que codifica para una pantotenato cinasa, *Plasmodium* cuando parasita al humano utiliza la biosíntesis de pantotenato para obtener nutrientes del ambiente del hospedero (Xu *et al.*, 2010). Los genes expresados a los 25 ddi pertenecen al perfil 9 y se complementan con los genes expresados en el perfil 7; no obstante, la expresión en este último es compartida con MM-VL. Existen genes con importancia para la simbiosis como *dctA* (transportador de C4-dicarboxilatos) la mutante en *S. meliloti* no fija nitrógeno y presenta senescencia temprana (Engelke *et al.*, 1989), *fixNd* (codifica para una citocromo oxidasa tipo C) en *R. etli* la mutante reduce la fijación de nitrógeno (Lopez *et al.*, 2001), y el gen *groES* (codifica para una cochaperona) que se propone participa en el plegamiento de la nitrogenasa. A éstos tres genes les localizamos sitios de unión a RpoN, por lo que proponemos que su regulación en simbiosis es similar a la que presentan los genes de fijación de nitrógeno. Los reguladores presentes a los 25 ddi son un regulador transcripcional y un miembro de la familia AraC cuya regulación es en genes que participan en la respuesta a estreses, utilización de carbohidratos y en la toma de sideróforos de hierro, funciones indispensables durante la fijación biológica de nitrógeno. Entre los reguladores presentes en el perfil 7 (25 ddi y MM) se encuentra NolRch, el cual se ha descrito como un regulador negativo para los genes de nodulación, lo que en este punto en particular correlaciona con la ausencia de expresión de genes de nodulación (perfiles 7 y 9). El regulador RosR actúa positivamente sobre los genes cuyos productos participan en la síntesis de exopolisacáridos. En ambos perfiles sólo localizamos un blanco para RosR,

específicamente en la propia región promotora de *rosR*. Además, el gen *exoF1* que codifica para una proteína de la biosíntesis de exopolisacáridos, presenta su máxima expresión a los 28 ddi (perfil 9) lo que sugiere el inicio de una posible reestructuración metabólica y funcional de la población de bacterias del nódulo, ya que además como señalamos previamente en este punto se inicia la disminución de la expresión de los genes de fijación de nitrógeno.

En resumen, la propuesta de regulación en este punto es la dependencia de los reguladores RpoN y NifA, sus blancos son los principales elementos de la fijación biológica de nitrógeno, y para el caso de RpoN se amplía a blancos involucrados en el transporte de dicarboxilatos. El regulador de respuesta a estrés por peróxido OxyR presenta diferentes blancos en este perfil, y por otro lado aún queda un grupo de blancos dependientes de RpoE, particularmente en el perfil 7 (Esquema del tiempo 25 y 28 ddi ver figura 34 inciso D).

Por lo que respecta a *R. etli* CFN42 e incluso para otras rhizobias, el punto 5 (35 ddi), es el tiempo más alejado analizado mediante transcriptómica en simbiosis, sin embargo es conveniente mencionar que en *B. japonicum* existe un estudio proteómico en el cual incluyen un tiempo de 49 ddi (Nomura *et al.*, 2010). En presente trabajo, este tiempo es el más parecido a la vida libre. Este tiempo muestra un grupo de genes que en su conjunto proponemos marcan la preparación metabólica por parte la bacteria para una probable restitución de la población de la bacterias en el suelo, pues se observan la expresión en genes de quimiotaxis y de privación de nutrientes. La señal para la activación de éstos seguramente proviene de la planta al bloquear el envío de compuestos de carbono y limitar fuentes de nitrógeno. El grupo de los genes involucrados directamente en la fijación de nitrógeno alcanzan el nivel más bajo de los 5 puntos evaluados en la simbiosis. El análisis para la búsqueda de sitios de unión de reguladores reportó pocos blancos, siendo el principal el consenso para RpoE. Entre los genes dependientes de RpoE, destaca el gen *thyA* el cual codifica para una timidilato sintasa; el timidilato es esencial para la construcción de bloques de DNA. El gen *trmU* codifica para una metiltransferasa tRNA (5-metil amino metil-2-tiouridilato) que participa en la traducción de proteínas, y el

gen *glgX* que codifica para una glicosil hidrolasa cuyo blanco es el glicógeno. La región promotora de ambos genes en su también presentan sitio de unión a RpoE. Al igual que en el perfil 2, en donde la mayor influencia de regulación es RpoE, el siguiente regulador por número de sitios localizados en los genes expresados a los 35 ddi es HipB, los genes con sitios HipB que destacan son: *atpI* que codifica para una ATP sintasa, el gen contiguo a este es *atpB* el cual también se localiza en el perfil 4 (35 ddi); el gen *bgIS* que codifica para una glucanasa (las bacterias que presentan éstos genes son capaces de degradar la celulosa), el gen *guaB* codifica para una inositol-5-monofosfato deshidrogenasa cuyo fenotipo simbiótico en *R. tropici* es nod⁺ Fix⁻ (Riccillo *et al.*, 2000), y el gen *mcpA* codifica para una proteína de quimiotaxis metil-aceptor. En *Burkholderia* se ha demostrado que la bacteria se mueve a través del micelio extra radicular de las estructura de los hongos y esto ocurre en la raíz de la planta hospedera, probablemente en respuesta a un estímulo quimiotáctil para la búsqueda de nutrientes y fuentes de energía (Bianciotto *et al.*, 1996 y Minerdi *et al.*, 2002). Los sitios Nod se presentan en menor medida en la región promotora de los genes de este perfil y tiene como probables blancos principales al gen *ocs* que codifica para la octopina deshidrogenasa (varios tipos de opinas proveen fuentes de nitrógeno, fosforo y carbono) y el gen *gabT* que codifica para una aminobutirato aminotransferasa (GABA), cuya función es para la producción de energía y la descarboxilación en el ciclo de TCA (Prell *et al.*, 2002). El otro sitio que destaca es FNR. Entre los blancos destaca el gen *ntaA* el cual codifica para una nitriloacetato monooxigenasa, la cual es la primera enzima para la biodegradación de NTA que es utilizado como fuente de carbono y nitrógeno (Xu *et al.*, 1997). El mayor grupo de genes con este sitio se observa en el tiempo de 18 ddi. En el perfil 4, identificamos 3 reguladores transcripcionales: uno la familia AraC sus blancos participan en la respuesta a estrés, utilización de carbohidratos y en la toma de sideróforos de hierro, y un miembro de la familia AsnC involucrado en el metabolismo de los aminoácidos.

La propuesta de regulación en este tiempo es complicada puesto que no hay un regulador que sobresalga de los demás, hay influencia de RpoE, HipB, Nod y FNR, pero algunos de

representativos son el J-Traducción con el 11.2%, el U- Tráfico y secreción intracelular con el 4.5%, el C-Producción de energía con el 10.7% y el O-Modificación postraducciona, chaperonas y renovación de proteínas con el 7.9%. Los genes más representativos son *rps*'s, *rpl*'s (ribosomales), *atpABC*, *vbs*'s y los genes *tra*'s.

En vida libre destaca la expresión de los genes que codifican para proteínas ribosomales y factores de traducción y otros “housekeeping genes”, además de genes cuyos productos participan en la generación de energía, división celular y los *trb* para la transferencia conjugativa. Los grupos funcionales enriquecidos con el programa GSEA corresponden a ribosomas, ciclo del TCA, fosforilación oxidativa, el sistema de transporte tipo IV y el ciclo reductivo del carboxilato (fijación de CO₂) (Fig. 30). El COG J-Traducción es casi exclusivo de la vida libre, y al igual que el COG O-Modificación postraducciona, chaperonas y renovación de proteína se presenta en los perfiles 3, 5, 7 y 10 que son los que contienen genes mayormente expresados en vida libre. Por último, el COG U-Tráfico y secreción intracelular (genes *trb*) es el más contrastante, ya que los genes que lo constituyen son en su mayoría exclusivos del perfil 3 (expresión en vida libre y represión en los 5 tiempos de la simbiosis). Entre los COG's que disminuyen en porcentaje de elementos en vida libre, destaca el COG F-Transporte y metabolismo de nucleótidos el cual presenta un bajo porcentaje a lo largo de los perfiles que contienen genes de vida libre (Figuras 16 y 24).

Los reguladores de la familia AraC que son activadores transcripcionales son los que predominan con 2 elementos de 4 reguladores exclusivos de vida libre. En cuanto a sitios de unión a DNA identificados, RpoE es el más abundante con 43 elementos seguido por HipB con 31 y RpoN con 26, esto tomando en cuenta los perfiles donde se expresan genes de vida libre (perfiles 3, 5, 7 y 10). Como señalamos anteriormente, en vida libre hay genes que indican que la bacteria se encuentra en una etapa de división constante y con un metabolismo general activo. Por ejemplo, los genes *ftsH*, *ftsZ2*, *mraZ* y *ftsW* que codifican para proteínas de división celular se expresan en vida libre, aunque a excepción de *mraZ* en donde sí identificamos un sitio de unión a RpoN, FNR, Tus y ArcA, para los 3 restantes no fue posible identificar ningún sitio de unión conocido. Otro grupo de genes relevantes de vida libre son los que pertenecen al COG C-Producción de energía, entre éstos los genes *atpC*, *atpD*, *atpA*, *atpE* y *atpF*, los cuales codifican para ATP sintetasas F0F1

subunidades épsilon, beta, alfa, C y B' respectivamente. De éstos previamente se había reportado su baja expresión en bacteroides, lo que respalda la baja actividad y división de este tipo celular. En *R. etli*, en un reporte reciente se estableció que la expresión de este grupo de genes también disminuye drásticamente cuando la bacteria entra en fase estacionaria (Vercruyse *et al.*, 2011). Un grupo de genes no reportado en los transcriptomas en donde se compara la expresión con la vida libre con respecto a la simbiosis, son los genes cuyos productos participan en la transferencia conjugativa, entre éstos *traCa*, *trbH*, *trbJ*, *trbE*, *trbD*, *trbC* y *trbB*. En *R. leguminosarum* los genes *trb* están involucrados en la transferencia del pSim (pRL1J1), una mutación en el gen *trbL* (ubicado en el perfil 4, 35 ddi) bloquea la transferencia del plásmido simbiótico (Wilkinson *et al.*, 2002). En *R. leguminosarum*, los genes *bisR* y *traR* se localizan adyacentes a la región *tral-trbBCDEJKLFGHI*. BisR induce la expresión de *traR*, en tanto TraR induce la expresión del operón *tral-trb*. La expresión es dependiente de "quorum sensing" a través de la concentración de homoserinas lactonas (Danino *et al.*, 2003). La disminución en la expresión durante la simbiosis de varios de los genes importantes para la bacteria podría tener sus bases en la coevolución. En un trabajo reciente, en donde analizan el transcriptoma de la interacción leguminosa-rhizobium observan que el efecto de la variación genética en la planta y en la bacteria afecta la expresión genética de la bacteria, pero la sola variación genética de la bacteria tiene un bajo efecto sobre la expresión genética de la planta. Los cambios en expresión por lo general repercuten en un beneficio en el acondicionamiento de la planta (Heath *et al.*, 2012). En el mismo sentido, se ha propuesto nombrar a los rhizobios fijadores de nitrógeno como aminoplastos cuya actividad principal es la de proveer nitrógeno a la planta (Terpolilli *et al.*, 2012). Lo observado en esta condición (VL) respalda en parte estas observaciones, la expresión de diferentes grupos de expresión benéficos para la bacteria son "reprimidos", en tanto que los grupos de genes cuyos productos aportaran un beneficio a la planta son expresados. Entre los genes con una drástica disminución en la expresión genética se encuentran los involucrados directamente en el ciclo de TCA, del que trataremos a detalle en los siguientes párrafos.

Perfil 4: 35 ddi

Perfil 5: VL, 35 y 18 ddi

Perfil 6: 11, 18 y 35 ddi

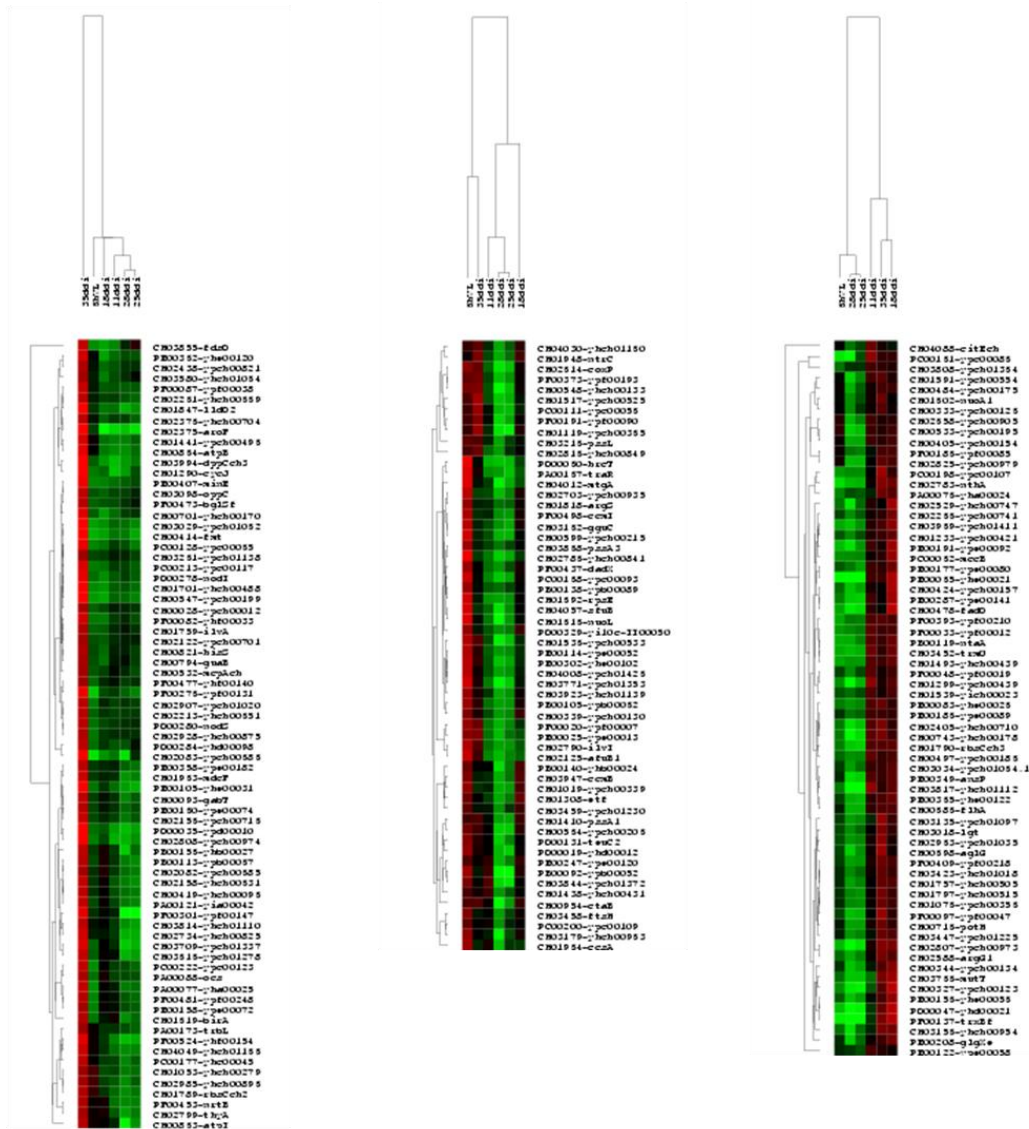


Figura 32. Cluster o agrupamiento jerárquico de los perfiles 4 (35 ddi), 5 (VL y 35 ddi) y 6 (11 y 18 ddi). El perfil 4 está representado por el 62% de genes del cromosoma y 12% del pF, sus principales reguladores son RpoD, HipB, RpoE, Nod y NifA. Los COG's más representativos son el G-Transporte y metabolismo de carbohidratos con el 13.3%, el E-Transporte y metabolismo de aminoácidos con 11% y el 4% de sus genes pertenecen al COG Q- Transporte y biosíntesis de metabolitos secundarios. Los principales genes de este perfil son *mcpA*, *birA* y *gabT*. El perfil 5 posee un 63% de genes del cromosoma, el 13% corresponden al pE y el 7% al pB. Sus principales reguladores son RpoD, RpoE, HipB y RpoH. Los COG's más representativos son el O-Modificación postraduccionnal, chaperonas y renovación de proteínas con el 10%, el COG P-Transporte y metabolismo de iones inorgánicos con el 8.3% y el COG M-Pared celular/Biogénesis de membrana con el 6.6% de los genes en este perfil. Los genes más representativos son *mtgA*, *ftsH*, *pssA* y *ntrC*. El perfil 6 está representado por un 66% de genes de cromosoma y un 16% de genes

pertenecen al pE. Sus principales reguladores son RpoD, RpoE, HipB y FNR. Los COG's más representativos son el COG G-Transporte y metabolismo de carbohidratos con el 24.3%, el E-Transporte y metabolismo de aminoácidos con un 15.7% y el COG I-Transporte y metabolismo de lípidos con un 8.6%. En este perfil destacan los genes *citEch*, *fadD*, *flhA* y *mccB*.

Además de la regulación, un factor que ofrece un nivel mayor de complejidad a este análisis, son los datos obtenidos y presentados en el artículo donde se analiza la expresión de los genes (transcriptoma) y proteínas (proteoma) durante la simbiosis, en donde se observa una baja correlación entre lo expresado e identificado entre ambas metodologías (Resendis-Antonio *et al.*, 2011). Los datos del transcriptoma referentes al ciclo de TCA, coinciden a lo reportado en otros modelos en donde se observa una represión durante la simbiosis, esto cuando los datos provienen de una comparación entre vida libre y simbiosis. Entre los datos que respaldan la baja expresión de éstos genes, tenemos que la mutante *acnA* (aconitasa) de *B. japonicum*, a pesar de que en vida libre la actividad de aconitasa disminuye en un 70%, esta cepa nodula y fija nitrógeno de manera similar a la cepa silvestre (Thöny-Meyer y Künzler, 1996); además mutantes en los genes *sucAB*, al igual que el fenotipo anterior, fijan nitrógeno al mismo nivel que la cepa silvestre (Green *et al.*, 2000). No obstante, en *R. leguminosarum* se reportó que las mutantes *sucA* y *sucC* tienen un fenotipo simbiótico Fix- (Walshaw *et al.*, 1997); además, en un análisis transcriptómico en nódulos de 28 ddi, observaron una sobreexpresión de los genes que componen el ciclo de TCA, entre éstos los genes *suc*, *acnA*, *citE*, entre otros (Karunakaran *et al.*, 2009), así también la mutante *acnA* de *S. meliloti* 1021 presenta un fenotipo Fix- (Koziol *et al.*, 2009). En *R. etli*, al comparar 18 ddi contra VL, observamos una clara tendencia de los genes involucrados en el ciclo del TCA hacia la vida libre, lo que nos indica una nula o baja expresión en simbiosis (Figuras 7, 9, 11 y Fig. 13 inciso b, Tablas 4 y 5). Prácticamente obtenemos el mismo patrón de expresión de éstos genes en el análisis de los perfiles generados a partir de los 5 tiempos de simbiosis y un tiempo de vida libre donde observamos la expresión de los genes del TCA en el perfil 3 (VL) (Fig. 16 incisos a y c, fig. 30). A diferencia de ambos resultados del transcriptoma, prácticamente todas las proteínas de este ciclo fueron identificadas en simbiosis a los 18 ddi (Resendis-Antonio *et*

al., 2011), las proteínas no identificadas (*CitEch* y *Sdh*) podrían ser la clave para la activación del ciclo en momentos específicos. Por ejemplo, observamos que *sdhC* que codifica para una succinato deshidrogenasa y *pdhA2* el cual codifica para piruvato deshidrogenasa subunidad beta, se expresan a los 25 ddi y en MM (perfil 7); en tanto *pdhA1* alcanza su máxima expresión en el perfil 3 (MM). Así mismo, en *R. etli* *PdhA2* fue identificada en simbiosis mediante proteómica (Resendis-Antonio *et al.*, 2011). El complejo de la piruvato deshidrogenasa cataliza la descaboxilación oxidativa del piruvato para producir acetil-CoA, lo que une la glicólisis con el ciclo de Krebs. En *S. meliloti* los genes *pdhA1* y *pdhA2* son mayormente expresados en simbiosis con respecto a vida libre (alrededor de 3 veces más), no obstante esta expresión no depende de la cascada FixLJ y en vida libre, al agregar piruvato, éstos genes se observan inducidos (Cabanes *et al.*, 2000). En nuestros resultados del análisis *in silico* no identificamos sitio de regulación alguno en la región promotora de *pdhA1* ni *pdhA2*. En *S. meliloti* una mutante en el gen *ada* cuyo producto es una arilesterasa, se observa disminuida la actividad de la PDH, presenta un fenotipo simbiótico al reducir la fijación de nitrógeno, lo que lleva a los autores a concluir que la actividad de la PDH es importante para la simbiosis (Soto *et al.*, 2001). En *A. caulinodans* a partir de mutantes *fixB* termosensibles, se ha propuesto que FixABCX pudiera facilitar la transferencia de electrones del complejo de la piruvato deshidrogenasa a la nitrogenasa (Scott y Ludwig, 2004). En nuestras observaciones la expresión de *fixB* coincide con la expresión de *pdhA2* a los 25 ddi, perfiles 9 y 7 respectivamente.

Por lo tanto, en *R. etli* si bien el análisis de los transcriptomas sostiene la hipótesis de que sólo una parte del ciclo de Krebs está activa, reveladoramente los genes detectados en el análisis de transcriptoma (perfiles de expresión) codifican para las proteínas que no fueron identificadas en el trabajo de Resendis-Antonio y col. 2011, y con excepción de la isocitrato deshidrogenasa (*icd*), tenemos identificados todos los elementos que participan en el ciclo del TCA, es decir observamos la expresión en simbiosis de los genes *citEch* (perfil 6, 18, 35 y 11-ddi, Fig. 19), y *sdh* (VL y 25 ddi, perfil 7, Fig. 20), proponemos que éstos genes podrían ser el cuello de botella para la activación o no del ciclo durante la

simbiosis. En *S. meliloti* se demostró que con el 7% de actividad de la citrato sintasa (*gltA*) es suficiente para fijar nitrógeno a los mismos niveles que la cepa silvestre, y solo si la actividad de la GltA disminuye a 0.5% se observa un fenotipo Fix- (Grzemeski *et al.*, 2005). Por lo tanto proponemos que los productos de estos genes son los que marcan el momento en el cual el ciclo del TCA es requerido, probablemente mediante regulación transcripcional o postraducciona. Abriendo un paréntesis para tocar el punto de regulación postraducciona, observamos la expresión de *pckA* en el perfil 10, 18 ddi y VL (Fig. 23) cuya mutante en *R. etli* presenta un fenotipo Fix- (Tate *et al.*, 2004). Resultados no publicados de un análisis proteómico han determinado que PckA presenta más de una entidad electroforética lo cual sugiere una modificación postraducciona (posible fosforilación) durante la simbiosis (Encarnación *et al.*, en preparación). Adicionalmente, en los genes expresados en simbiosis identificamos a *phoA* (fosfatasa alcalina) y al gen hipotético con ID CH003064 (probable fosfatasa), ambos en el perfil 2 (Fig. 15, 18 ddi), cuyos productos podrían funcionar en la regulación postraducciona durante la simbiosis. El dato más concluyente para indicar que en *R. etli* CFN42 el ciclo de TCA es activo en simbiosis, es reportado en el metaboloma del bacteroide a los 18 ddi, en donde se identificaron cinco metabolitos que participan en este ciclo; entre éstos, ácido cítrico, ácido succínico, ácido 2-oxoglutarico y ácidos fumarico y málico (Resendis-Antonio *et al.*, 2012). En cuanto a la regulación transcripcional de este grupo de genes, en la región promotora del gen *citEch* no identificamos sitios de regulación; por otro lado, por lo que respecta a la región promotora del gen *sdhC* presenta sitios para PhoB3, PurR, Tus y RpoD. Finalmente, los datos de los grupos funcionales (COG's, KEGG y GSEA) respaldan la hipótesis de la presencia y funcionalidad en la simbiosis del ciclo de TCA; por ejemplo, hay una baja o nula expresión en simbiosis de los genes involucrados en la síntesis de proteínas (COG J-Traducción), lo que nos indica una baja en la producción *de novo*, así como también una ausencia en los perfiles simbióticos de genes que codifican para proteínas involucradas en la degradación, chaperonas y modificación de proteínas, lo que sugiere la estabilidad proteica en el ambiente simbiótico. Es necesario evaluar la expresión de este grupo de genes en simbiosis en condiciones diferentes a las realizadas

en este trabajo, por ejemplo, cuando se reduzca la fotosíntesis (ensayos en oscuridad) o cuando la planta sea senescente.

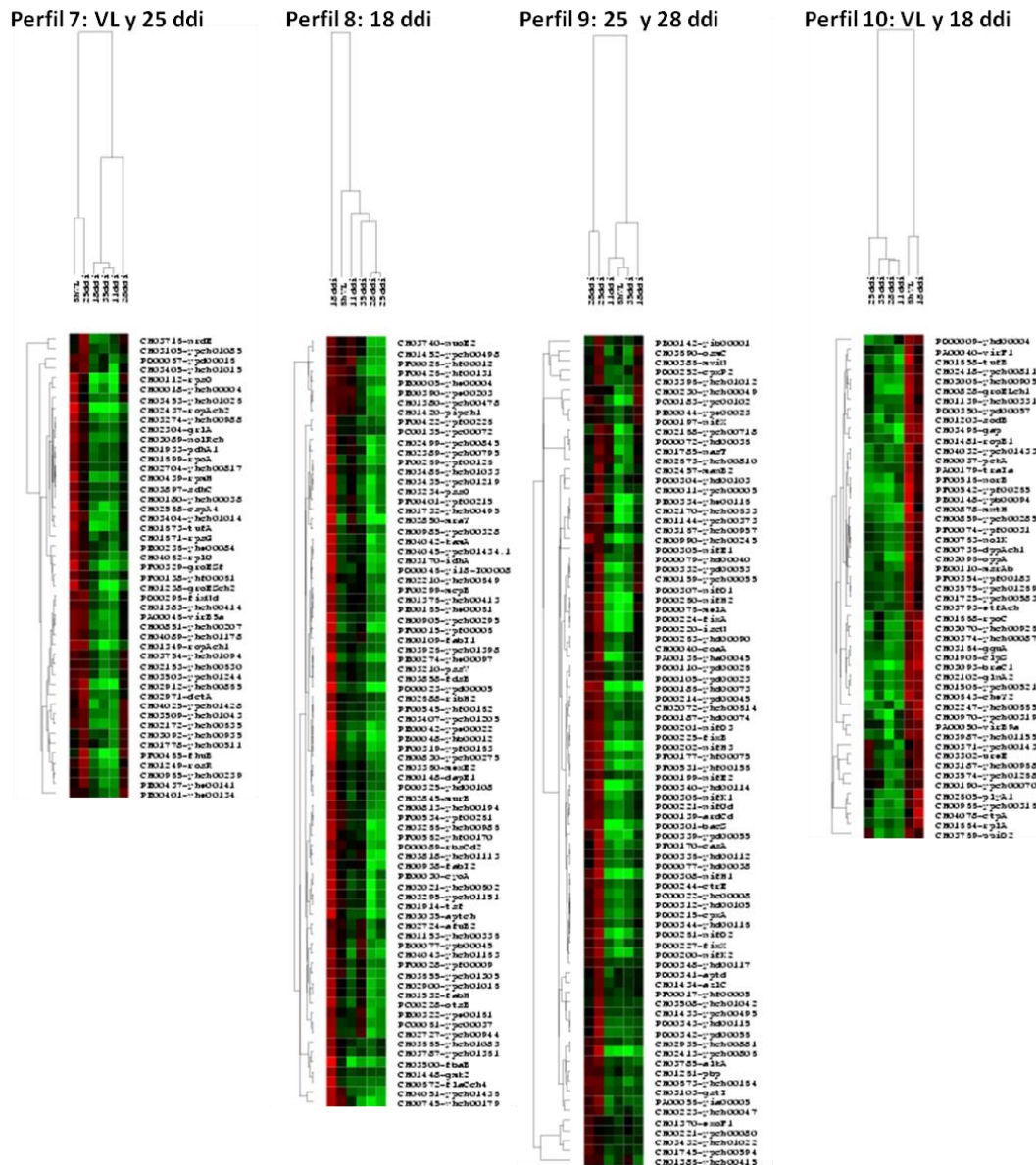


Figura 33. Cluster o agrupamiento jerárquico de los perfiles 7 (VL y 25 ddi), 8 (18 ddi), 9 (25 y 28 ddi) y 10 (VL y 18 ddi). El perfil 7 está representado por el 82 % de genes del cromosoma, sus principales reguladores son RpoE, RpoD, RpoN, NifA, PhoB y PurR. Los COG's más representativos son el K-Transcripción, el J-Traducción ambos con el 10.2% y el COG O-Modificación postraduccional, chaperonas y renovación de proteínas con el 8.2%. Los principales genes de este perfil son *pdhA2*, *dctA*, *sdhC* y *groES*. El perfil 8 posee un 63% de genes del cromosoma y el 16% al pF. Sus principales reguladores son RpoD, FNR, HipB, MalT y OxyR. Los COG's más representativos son el K-Transcripción con el 10.1%, el COG I-transporte y metabolismo de lípidos con un 7.6%, y el COG M-Pared celular/Biogénesis de membrana con el 6.3% de los genes en este perfil. Los genes representativos de este perfil

son *mraY*, *murB*, *aptch* e *idhA*. El perfil 9 está representado por un 49% de genes del pD y por un 37% de cromosoma. Los reguladores principales son RpoN, RpoE, NifA, RpoD y OxyR. Los COG's más representativos son el C-Producción y conversión de energía con el 12.64% y el COG H-Transporte y metabolismo de coenzimas con el 2.3%. Los genes representativos son *nifHDK*, *fixABCX*, *iscN*, *leuD* y *gstI*. El perfil 10 el 77% son de cromosoma, un 7% son del pF y un 6% del pA. Los principales reguladores son RpoD, RpoE, RpoN, RpoS y FNR. Los COG's representativos son el E-Transporte y metabolismo de aminoácidos con un 13.46%, el COG T-Mecanismo de traducción de señales y el COG P-Transporte y metabolismo de iones inorgánicos ambos presentan un 7.7%. Los genes más representativos son *sodB*, *mntH*, *clpS* y *mrsA*.

Un grupo de genes poco estudiados en simbiosis es el relacionado a la vía del metabolismo de las purinas. El antecedente directo en *R. etli* CFN42 es que observaron que mutantes en los genes *purF*, *purD*, *purQ*, *purY*, *purL* y *purE* aumentan la expresión de *fixNOQP* pero en la mutante *purH* (cuyo producto cataliza la conversión de ribonucleótido 5-aminoimidazole-4-carboxamida (AICAR) a ribonucleótido 5-form-aminoimidazole-4-carboxamida (FAICAR) e inosina no se modifica su expresión. Los autores proponen que AICAR o un metabolito cercano inhibe la expresión de *fixN* (Soberón *et al.*, 1997). En *S. meliloti* al utilizar un precursor de AICAR se observó la disminución de la expresión de *fixK*, *fixN*, *fixT* y *nifA*, esto en condiciones microaeróbicas (Soberón *et al.*, 2001). En nuestro análisis observamos la expresión de los genes *purC* (codifica para una defosforibosil aminoimidazol succino sintetasa) y *purU*, el cual codifica para una deformilasa de formil tetra hidrofolato que produce formato a partir de formil-tetrahidrofolato, siendo la fuente de formato para PurT en la biosíntesis "de novo" de purinas, tiene un papel en el metabolismo de carbono, se activa por metionina y se inhibe por glicina. Una mutante en *purU* de *E. coli* disminuye la infección sobre violetas africanas (Okinaka *et al.*, 2002). La máxima expresión de *purUe* se observa a los 11 ddi, en tanto que la expresión de *purUch* y *purC* la observamos a los 18 ddi. En este punto también se observa la expresión de los genes *cyaFch2*, CH02560 y PE00177 las tres codifican para adenilato ciclasas, además de los genes *gmk2* (guanilato cinasa o GMP cinasa) enzima esencial para la biosíntesis de GTP y dGTP) y *aptch* (una adenina fosforibosil transferasa). La copia del gen *aptd* se localiza en pSim y su máxima expresión se observa a los 25 ddi, incluso forma parte del regulón RpoN-NifA (Salazar *et al.*, 2010). Al igual que en nuestro análisis, en *R. etli* utilizando

nódulos de 21 ddi, fue detectada la expresión de un gen *aptD* exclusivo de simbiosis (Vecruyse *et al.*, 2011). La función de Apt es catalizar una reacción de salvamento que da por resultado la formación de AMP difosfato a partir de adenina + 5-fosfo-alfa-D-ribosa 1-difosfato, lo que resulta menos costoso energéticamente que la síntesis “*de novo*”. A los 25 ddi, compartido con vida libre, se observa la expresión de *rpoA* (la RNA polimerasa subunidad alfa) y de *nrdE* el cual codifica para una ribonucleótido-difosfato reductasa subunidad alfa. En *S. meliloti* la sobreexpresión de *nrdJ* (de *E. coli*) que codifica para una ribonucleótido reductasa dependiente de cobalamina (RNR) afecta la simbiosis. Mediante micrografías se observa que las bacterias de esta cepa pueden inducir nódulos pero que son incapaces de diferenciarse a bacteroides y en cierto momento son lisadas en el citoplasma de la planta. Este defecto puede deberse a la inactivación de la RNR por especies reactivas de oxígeno, o por una inadecuada habilidad de contender con el oxígeno en el nódulo, o bien una combinación de ambas (Taga y Walker, 2010). A los 35 ddi, se expresa expreado el gen *guaB* el cual codifica para una inositol-5-monofosfato deshidrogenasa. En *R. tropici* la mutante *guaB* presenta sensibilidad al calor, y un fenotipo simbiótico al formar nódulos carentes de rhizobios (Ricciolo *et al.*, 2000). Utilizando los promotores para *nodA*, *bacA* y *nifH* fusionados al gen *guaB*, sólo el promotor para *nodA* recobró el fenotipo silvestre, lo que indicaría que GuaB se requiere en las etapas iniciales de la interacción. La mutante *guaB* en *S. meliloti* induce nódulos silvestres en alfalfa y en *M. truncatula*, y analizando la mutante *guaB* de *R. tropici* en *L. leucocephala* (nódulos indeterminados) ésta no presenta fenotipo simbiótico, por lo que sugieren que *guaB* es requerido para la simbiosis pero sólo en el caso de plantas que forman nódulos determinados (Collavino *et al.*, 2005). El metabolismo de las purinas podría ser un referente en el proceso de la fijación biología de nitrógeno, y a través de la concentración de metabolitos como AICAR (represión metabólica), pueden controlar la expresión de genes. No obstante, en varios de los ejemplos el impacto de genes relacionados a la purinas se presenta en tiempos tempranos de la interacción, también pudiera tener un efecto en las etapas tardías, incluso para el caso de AICAR en *R. etli* se observa el mayor impacto en dichos tiempos (Soberón *et al.*, 1997).

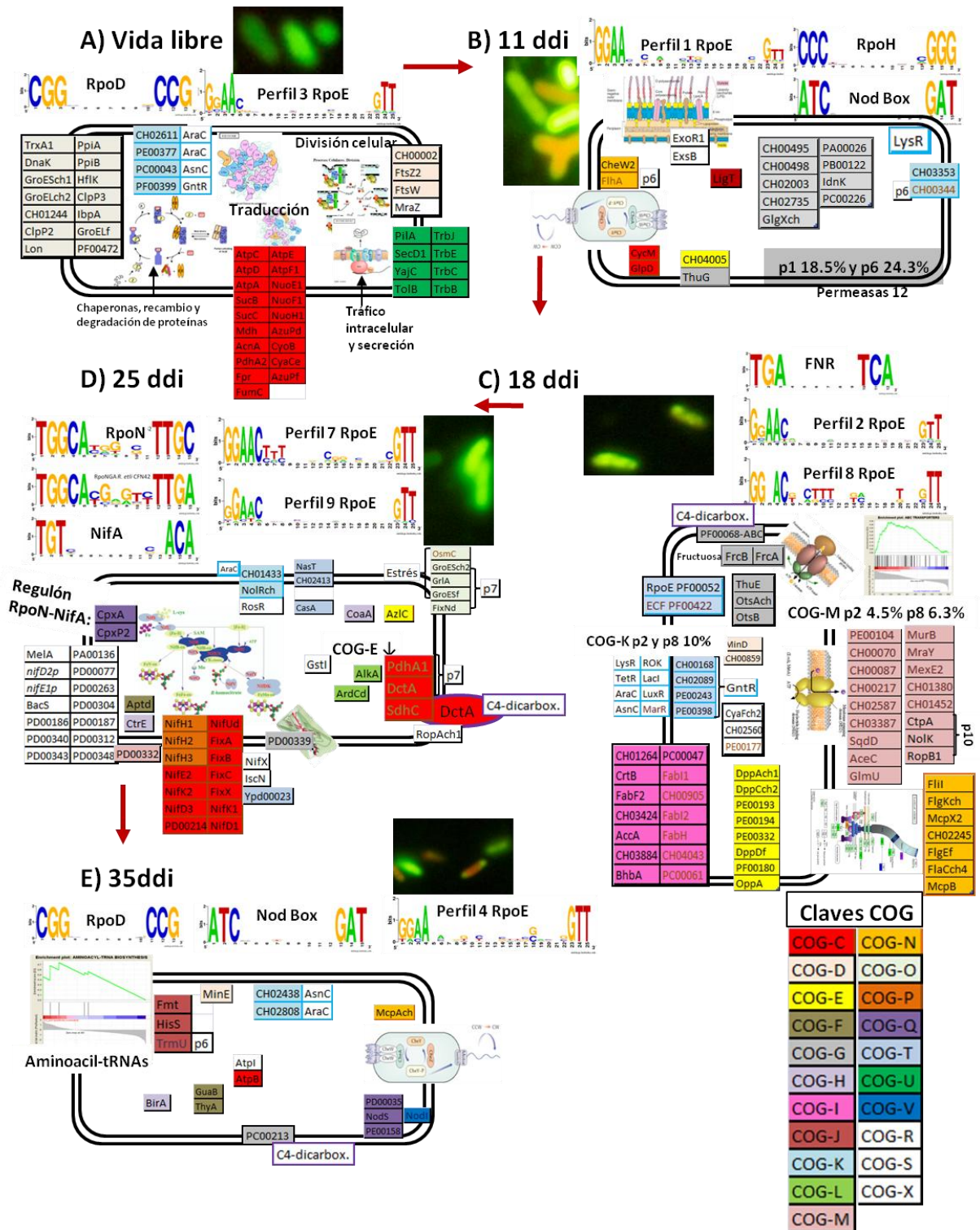


Figura 34. Representación de la temporalidad de expresión basada en cuatro puntos de la simbiosis y un punto de vida libre de acuerdo a sus niveles de expresión. La descripción de los mismos se presenta en la sección de resultados 11.3 y en la discusión. A) Vida libre, B) 11 ddi, C) 18 ddi, D) 25 y 28 ddi, E) 35 ddi. Los claves de los COG's son: C-Producción y conversión de energía, D-Control del ciclo celular, E-Transporte y metabolismo de

aminoácidos, F-Transporte y metabolismo de nucleótidos, G-Transporte y metabolismo de carbohidratos, H-Transporte y metabolismo de coenzimas, I-Transporte y metabolismo de lípidos, J-Traducción, K-Transcripción, L-Replicación, recombinación y reparación, M-Biogénesis de membrana y pared celular, N-Motilidad celular, O-Modificación postraduccional, degradación de proteínas y chaperonas, P-Transporte y metabolismo de iones inorgánicos, Q-Transporte, catabolismo y biosíntesis de metabolitos secundarios, T-Mecanismos de traducción de señales, U-Tráfico intracelular y secreción, V-Mecanismos de defensa, R-Solo predicción de función general, S-Función desconocida y X-No en COG's. Se incluyen los sitios de regulación representativos de cada tiempo, así como el consenso para RpoE. Las micrografías mostradas son representativas de una condición de vida libre y de bacteroides colectados en cada tiempo posteriormente etiquetados con Syto 9 (verde) y PI (rojas).

El metabolismo de las purinas se complementa con la expresión de los genes propios del metabolismo de las pirimidinas. Los genes *trxBf* (una tiorredoxina (NADPH) reductasa, *deoA* (fosforilasa de timidina) y *pyrE* (orato ribosiltransferasa) se expresan a los 18 ddi, en tanto que a los 35 observamos la expresión de los genes *thyA* (una timidilato sintasa) y nuevamente *trxBf* (perfil 6, 18 y 35 ddi). TrxB, al igual que KatA, AhpCF y Bcp, confieren resistencia al estrés oxidativo, en *S. aureus* y son reguladas por el mismo factor transcripcional PerR, además son necesarias para la virulencia (Horsburgh *et al.*, 2001). Las tiorredoxinas en plantas constituyen una familia de pequeñas proteínas. En *M. truncatula*, se localizaron dos isoformas específicas de leguminosas, su expresión se observa en simbiosis y están clasificadas como *trxs*'s simbióticas (Alkhalfioui *et al.*, 2008). El gen *deoA* es expresado a los 18 ddi y participa en la fosforilación reversible de la timidina, pero en simbiosis aún no ha sido caracterizada. En *S. meliloti* una mutante *pyrE* (una orato ribosiltransferasa, perfil 2) retrasa la nodulación (Pobigaylo *et al.*, 2008), además la mutante *pyrF/pyrE* produce nódulos pero no fija nitrógeno. Los nódulos producidos por la mutante *pyrF/pyrE* presentan características estructurales más avanzadas que los nódulos de la mutante *pyrC*, esto podría indicar que algunos intermediarios de la biosíntesis de las pirimidinas tienen un papel sobre la transformación del bacteroide y del desarrollo del nódulo (Vineetha *et al.*, 2001). Respecto a ThyA (una timidilato sintasa, perfil 4), el timidilato es esencial para la construcción de bloques de DNA, se sintetiza a partir de deoxiuridilato utilizando la enzima timidilato sintasa (TS) o bien de timidina (dT) utilizando

la timidina cinasa (TK). En *M. pneumonidae* la sobre expresión de la timidina cinasa complementa la pérdida de la timidilato sintasa, aunque los autores proponen que podría existir una TS alternativa en el genoma de *M. pneumonidae* (Wang *et al.*, 2010). Los datos de expresión y fenotipos de este grupo de genes nos indica la importancia de incorporarlos para evaluar su expresión y fenotipo simbiótico a lo largo de las diferentes etapas de la fijación biológica de nitrógeno.

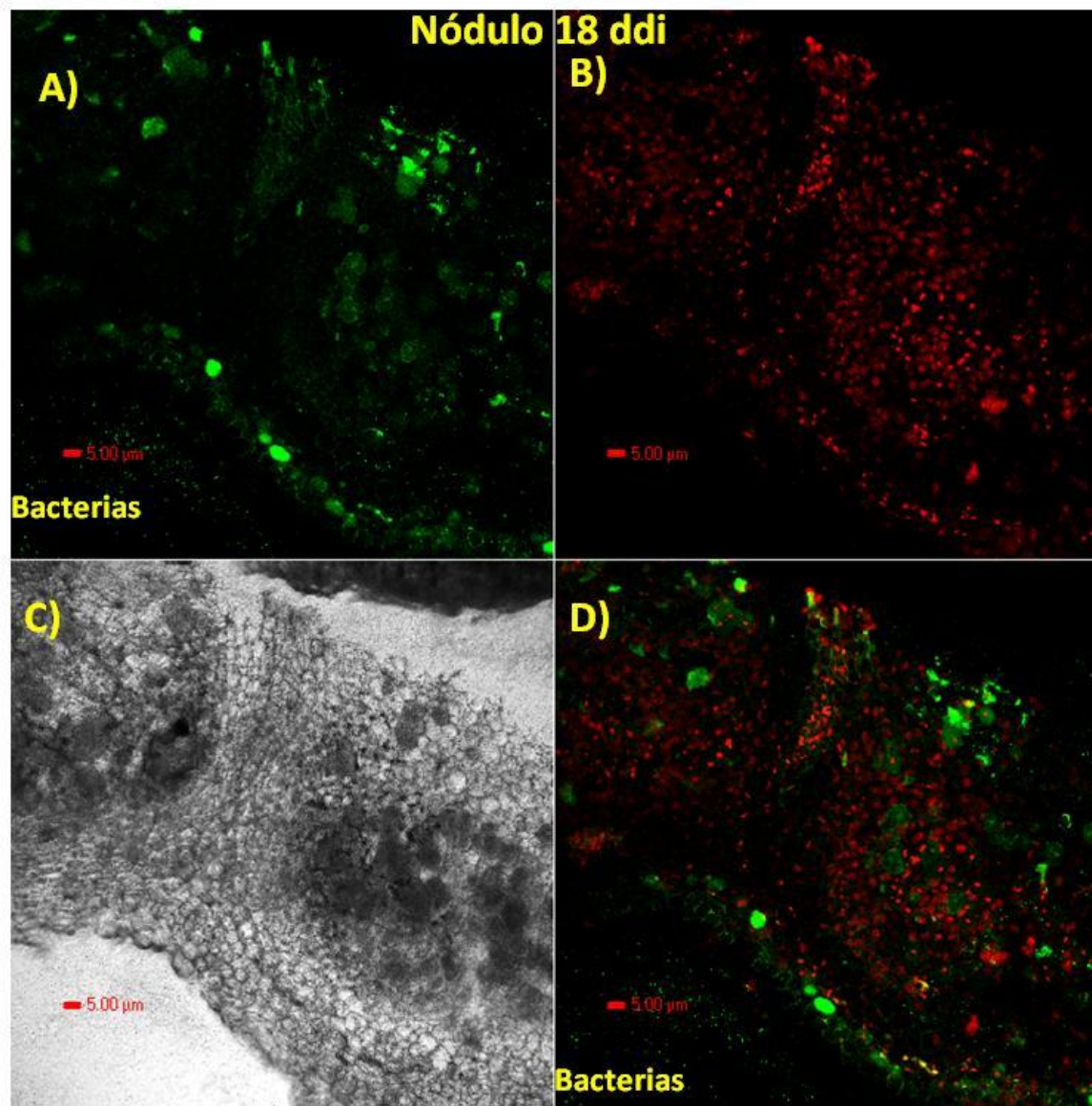


Figura 35. Corte de nódulo a los 18 ddi (*Phaseolus vulgaris*-*R. etli* CFN42), A) Fluorocromo SYTO 9, tinción verde (bacteria viable), B) Yoduro de propidio o PI, tinción roja (célula muerta o dañada), C) Sin tinción, y D) Sobrelape de imágenes SYTO 9 y Yoduro de propidio.

El análisis expuesto a continuación incluye sólo los componentes del bacteroide expresados en un nódulo fijador (la mayoría de genes corresponden a los tiempos 18 y 25 ddi). El modelo general de cómo funcionan los bacteroides en un nódulo determinado lo presentamos en la figura final (Fig. 36), la cual fue tomada de una revisión reciente acerca de la nodulación y fijación de nitrógeno por rhizobios (Oldroyd *et al.*, 2011) y modificada con nuestros datos. Al igual que en el resto de las Rhizobiaceas, el malato, succinato y fumarato son los principales dicarboxilatos provistos por la planta (Westenberg y Guerinot, 1999). En *R. etli* CFN42, detectamos la expresión de los genes involucrados en el transporte de C4 dicarboxilatos, entre ellos el principal, *dctA* (Yurgel y Kahn, 2004), localizado en el perfil 7 que corresponde a genes expresados principalmente en vida libre y en un segundo plano a los 25 ddi. La regulación de *dctA* es llevada a cabo por sigma 54 y por el complejo DctD-DctB de cuyos transcritos no detectamos expresión utilizando nuestros parámetros. Cabe destacar que con los mismos parámetros tampoco detectamos expresión del gen *nifA*. En la región promotora del gen *dctA* localizamos el sitio de unión a RpoN, por lo que proponemos que al igual que los genes de fijación de nitrógeno, *dctA* es regulado por RpoN2 en simbiosis. A los 18 ddi (perfil 2), se observa la expresión de un gen hipotético con ID PF00068, que codifica para un transportador tipo ABC de C4-dicarboxilatos, en la región promotora del gen con el que posiblemente este acoplado (PF00069) identificamos 2 sitios, uno para RpoN y otro para RpoD. En el mismo perfil donde se localiza *dctA* se encuentra el gen *sdhC* cuyo producto cataliza la oxidación de succinato a fumarato, si bien las principales fuentes de carbono son las que señalamos previamente, tenemos un grupo diverso involucrado en el transporte, asimilación y uso de diversas fuentes de carbono. Un ejemplo que destaca son los transportadores tipo ABC enriquecidos positivamente a los 18 ddi y que son enriquecidos negativamente a los 25 y 28 ddi. Por ejemplo, los productos del operón *frcAB* que participan en el transporte de fructosa y se localizan en el perfil 2 (18 ddi). Al igual que nosotros Vercruyse y colaboradores (2011) indican que el operón constituido por los genes hipotéticos con ID's PB00113, PB00114 y PB00115 codifican para transportadores de azúcar tipo ABC se expresan a los 21 ddi (Vercruyse *et al.*, 2011), en nuestro análisis nosotros detectamos

los genes con ID's PB00113 y PB00114 en los perfiles 4 y 9, respectivamente. Para el gen del perfil 4, no identificamos sitio de regulación alguno, pero en la región promotora del PB00114, perfil 9, localizamos 1 sitio para NifA, 1 para RpoE, 1 para Sox y uno para RpoD. Una vez que se han establecido las condiciones óptimas en los nódulos se incrementa la fijación de nitrógeno, por lo que los bacteroides disminuyen significativamente la asimilación de amonio en los amino ácidos y secretan en mayor medida el amonio a la planta, este patrón lo tenemos representado en los genes que se reprimen a los 25 ddi y en menor medida a los 28 ddi, donde *braC* (perfil 10) y *braF* (perfil 2) presentan el nivel más alto de represión respecto a los otros tiempos de simbiosis e incluso a la vida libre. En contraste a esto, los genes involucrados directamente en la fijación presentan su máxima expresión a los 25 ddi (perfil 9). En este perfil también se localiza el gen *gstI*; en *R. leguminosarum* la presencia de GstI reduce la capacidad de asimilación de NH_4^+ (Tatè *et al.*, 2001). Un evento relacionado con la disponibilidad de aminoácidos se presenta cuando el bacteroide está en su punto máximo de fijación, a diferencia del resto de los perfiles, en el perfil 7 (25 ddi y MM) observamos que no hay genes pertenecientes al COG E-Transporte y metabolismo de aminoácidos.

Una serie de sistemas de transporte para diferentes elementos son requeridos para el óptimo funcionamiento del bacteroide, de los cuales incluso sumando los identificados en diversos modelos fijadores de nitrógeno en simbiosis con plantas leguminosas aún quedan huecos en algunas funciones. En este modelo, se presentan los identificados en *R. etli* CFN42 utilizando principalmente los datos del transcriptoma generados en este trabajo. El transportador identificado para Mg^{2+} en *R. leguminosarum* es MgtE, una mutación en el gen *mtgE* produce un fenotipo Fix- (Karunakaran *et al.*, 2010). En *R. etli* CFN42 en el perfil 10 observamos al gen *mntH*, que codifica para un transportador de manganeso (Mn^{2+}) y que podría también transportar Mg^{2+} , en *B. japonicum* el gen es regulado por Fur, y una mutante *mntH* cultivada en limitaciones de Mn^{2+} tiene una deficiencia en el crecimiento, utilizando al piruvato como fuente de carbono (Hohle y O'Brian, 2009). En la región promotora de este gen no detectamos de sitios de regulación conocidos. En cuanto al hierro, en la revisión publicada por Oldroyd y colaboradores, (2011), se analiza este

sistema de transporte, quedando sin definir un probable gen candidato para codificar a la permeasa. No obstante, en el trabajo de Vercruyssen y colaboradores señalan que *AfuA3* transporta el hierro al interior del bacteroide. El transcrito para *afuA3* fue localizado en nódulos de 21 ddi (Vercruyssen *et al.*, 2011), en nuestros datos la expresión de este gen se ubica en el perfil 2 (18 ddi), además el gen *afuB2* se observa expresado a los 18 ddi (perfil 8). Adicionalmente en la región promotora de *afuA3* identificamos un sitio para PhoB y uno para RpoD, en la región promotora del gen *afuB2* no identificamos sitio alguno pero los ID's son contiguos CH02724 y CH02725 respectivamente, lo que indicaría que probablemente forman un operón. Para el caso de fosfato, en los perfiles 2 y 9 identificamos los genes hipotéticos con ID's CH00159 y CH00160 respectivamente, ambos pertenecen a la familia de transportadores para fosfonato tipo ABC y se han propuesto para el transporte de fosfato en simbiosis. Recientemente en nódulos de 21 ddi se detectó la expresión de CH00158 que codifica también para un probable transportador de fosfonato tipo ABC (Vercruyssen *et al.*, 2011) el cual sugerimos está en operón con CH00159 y CH00160. En la región promotora de CH00159 localizamos los sitios para HipB, OxyR, RpoD y una caja Nod; en cuanto a CH00160 localizamos un sitio para NifA y uno para PhoB. Además del gen hipotético con ID CH02045 el cual codifica para un transportador tipo ABC para nitrato, sulfonato y carbonato, su máxima expresión se presenta a los 11 ddi (perfil 1), y en su región promotora no identificamos ningún sitio de regulación conocido.

La activación transcripcional de los promotores de bacterias en su medio no es una simple decisión de encendido o apagado. Los genes o grupos de genes aquí descritos, requieren una coordinada expresión entre los genes de la bacteria y los de la planta, la expresión se modifica en dependencia a los cambios continuos de las señales moleculares que se presentan a lo largo de la interacción. Así, nosotros presentamos algunos aspectos de la regulación de *R. etli*, pero enfatizando los elementos de regulación que gobiernan la transcripción durante la simbiosis. Para estudiar a detalle los diferentes promotores que regulan los genes expresados en simbiosis y vida libre utilizamos secuencias previamente reportadas y herramientas bioinformáticas para buscar los sitios de regulación presentes

en los genes que modificaron su expresión. Además, empleamos la información de los genes que codifican para reguladores transcripcionales presentes en cada perfil para tener un panorama de regulación más detallado. La expresión de los genes en el bacteroide presentados y analizados en esta tesis, valida información previa y genera nueva información que consideramos da respuesta del funcionamiento general del bacteroide permitiéndonos concluir que a través de diferentes vías y sistemas de regulación, la serie de genes básicos para el funcionamiento correcto del bacteroide es similar tanto en nódulos determinados como no determinados (Fig. 36). Aunque existen características propias al comparar bacterias que inducen nódulos indeterminados y bacterias con nódulos determinados, hay diferencias que incluso se presentan entre bacterias afines, por ejemplo la mutante *pckA* en *R. leguminosarum* 3841, fija nitrógeno normalmente (Mulley *et al.*, 2010), pero la mutante en *R. etli* CFN42 presenta un fenotipo Fix- (Tate *et al.*, 2004). Otra de las diferencias es que en nuestros resultados observamos la expresión de genes de nodulación principalmente a los 11 y 35 ddi, lo que contrasta con los transcriptomas en nódulos indeterminados, donde no observan la expresión de este tipo de genes. Un aspecto fundamental entre las bacterias que inducen nódulos indeterminados respecto a los determinados se observa en su viabilidad, así en nódulos indeterminados las bacterias sufren endoduplicación (división del cromosoma hasta alcanzar 18C en una misma célula) y son inviables (el yoduro de propidio, IP, ingresa lo que indica que la célula está muerta o dañada). Los bacteroides de los nódulos determinados no sufren endoduplicación y una gran parte de la población es viable (Mergaert *et al.*, 2006, Oono y Denison, 2010).

Los diferentes modelos de fijación biológica de nitrógeno coinciden en aspectos fundamentales, tal como la necesidad de una baja concentración de oxígeno, y en el intercambio de nutrientes donde la planta aporta carbono y la bacteria nitrógeno, pero en los trabajos reportados hasta el momento no se ha obtenido un mapa de expresión que refleje la temporalidad de esta y con ello se conozcan las bases moleculares presentes en la interacción necesarias para el funcionamiento armónico entre ambos organismos. La ventaja que otorga el análisis en nódulos determinados es la identificación de la expresión

de genes de acuerdo a su temporalidad. Para un análisis similar en nódulos indeterminados se requeriría utilizar microdisecciones con láser como se hizo para evaluar el transcriptoma tejido específico de nódulos de *Lotus japonicus* (Takanashi *et al.*, 2012). Una de los aspectos relevantes de este tipo de herramientas de análisis de expresión global es que permite delimitar redes de regulación o como en este caso regulones; no obstante, para extraer al máximo la información generada por este tipo de metodologías, en particular los microarreglos, es necesario apoyarse en programas que faciliten la extracción e interpretación de los mismos. Por ello, nosotros utilizamos programas como GSEA que nos permitió agrupar los genes de acuerdo a su clasificación funcional, "RsaTools" que a partir de secuencias previas aportó los sitios probables de regulación para cada uno de los genes, KEGG Color Pathway que nos ayudó a la ubicación en las diferentes vías en que participa cada gen de manera gráfica, el programa Cluster a partir de los valores de expresión de los genes en cada condición y tiempo aquí presentada, nos ayudó al agrupamiento de cada gen. El análisis de búsqueda de los sitios de regulación de acuerdo a la temporalidad de la expresión es el primero que se realiza en Rhizobios, incluso varios de los sitios identificados en este trabajo no se han reportado previamente en *R. etli* CFN42. Como se mencionó en varias partes del trabajo, la regulación del genoma de *R. etli* CFN42 es compleja, pero este trabajo clarifica una parte sustancial del funcionamiento esta bacteria de importancia agrícola y comercial durante la fijación biológica del nitrógeno.

13. PERSPECTIVAS

De los datos presentados en este trabajo, uno de los más complejos y complicados de explicar es la expresión de genes que codifican para proteínas de flagelo, principalmente en los perfiles 2 y 8, que corresponden a los 18 ddi; una de las hipótesis es que las bacterias no diferenciadas en este punto tienen un primer evento para la repoblación del suelo, por lo que es necesario evaluar si la población de bacterias en el exterior del nódulo en este punto es mayor a los otros tiempos analizados, también se podría hacer un seguimiento de las bacterias en los diferentes estadíos, incluso analizando los nódulos de 35 ddi o posteriores; para ello sería necesario utilizar además de la microscopía, anticuerpos dirigidos a proteínas de flagelo, además se requiere que las observaciones en condiciones propias del nódulo afecten lo menos posible el sistema.

Los sitios de unión detectados en este trabajo para RpoN tienen una ligera variación, principalmente en el sitio de -12, donde el consenso reportado es GC y en este análisis observamos que hay cambio por un GA (Fig. 25); por lo que habría que evaluar si RpoN1 de cromosoma puede reemplazar al RpoN2 del plásmido simbiótico. En *R. etli* CNPAF512, durante la simbiosis RpoN1 no es capaz de complementar a la mutante *rpoN2*, pero lo observado en nuestros datos indican que la estructura y regulación es diferente, por lo que en el futuro cercano deberán de analizarse las regiones promotoras de cada *rpoN* creando quimeras, por ejemplo, utilizar la región cadena arriba de *rpoN2* intentando expresar a *rpoN1* en microaerobiosis y evaluar su actividad en la simbiosis; además se requieren ensayos mediante geles de retardo con las diferentes proteínas RpoN y evaluar su capacidad de unión para cada consenso reportado en este trabajo.

La carencia de información sobre la regulación de *nifA* deja abierta la puerta para identificar el o los elementos que participan en ésta, los posibles candidatos pueden ser proteínas con dominios PAS, para lo cual proponemos realizar los ensayos principalmente en condiciones simbióticas.

Los sitios completos localizados para RpoD o sigma 70 fueron los menos a lo largo de los 10 perfiles, siendo este el sigma universal o “house keeping” es necesario generar un

consenso que nos permita identificar los blancos para este sigma fundamental, un acercamiento fue hecho por el grupo del Dr. Ramírez-Romero y col. en 2006, en el que señaló que a diferencia del consenso de *E. coli* donde el sitio -10 es primordial en *R. etli* tiene mayor importancia el sitio -35.

Varios de los genes reportados en este trabajo por la importancia sugerida son candidatos para obtenerse la mutante respectiva, para su posterior evaluación durante la simbiosis. En especial nos interesa evaluar a los diferentes reguladores transcripcionales localizados en cada perfil, como el posible regulador transcripcional *ypd00023* (ID PD00105) el cual hemos definido como miembro del regulón RpoN-NifA. En rhizobias es poca la información que se tiene sobre sitios de unión a DNA, con base en este trabajo consideramos necesario generar una base de datos que contenga los sitios de pegado verificados experimentalmente, sitios localizados *in silico*, que además incluyan mutantes en reguladores transcripcionales, fenotipos de éstos y de los genes miembros de los respectivos regulones. Esta información podría ayudar en un futuro cercano a modificar la capacidad de la bacteria en la fijación de nitrógeno, en la sobrevivencia y en la competencia.

Es necesario analizar con un mayor detalle cada uno de los componentes de la simbiosis descritos en esta tesis, en especial los genes hipotéticos o de función desconocida. De muchos de los genes previamente reportados aún falta evaluar su esencialidad durante la simbiosis, particularmente la mayoría de los genes del perfil 9 (regulón RpoN-NifA). Las características a revisar deben incluir además de las evaluaciones clásicas; nodulación y nivel de fijación de nitrógeno, la competencia y capacidad para repoblar el suelo, si bien de éstos 2 últimos ensayos se han llevado a cabo en diversos trabajos, consideramos primordial su realización en todos los estudios que utilicen el modelo rhizobium-leguminosa.

Es también necesaria una validación mediante RT-PCR de algunos de los genes expresados que son referencia del tiempo en el que se expresan, así como identificar sus sitios de inicio de transcripción. Esto último se puede realizar con ensayos de RNAseq.

Finalmente, dado que los resultados obtenidos utilizando el microarreglo de oligonucleótidos de *R. etli* CFN42 junto con los programas de análisis presentadas en este trabajo, han funcionado para proponer cuáles son los elementos necesarios en la simbiosis entre *R. etli-P. vulgaris*, este tipo de estudios puede ser utilizado para el análisis de mas condiciones a las aquí presentadas, asi como en mutantes de reguladores transcripcionales, para definir el grupo de genes dependientes de los 21 sigmas que aún faltan por evaluar de los 23 que se propone tiene *R. etli* CFN42, de acuerdo a la secuencia de su genoma.

14. REFERENCIAS

Aguilera L, Campos E, Giménez R, Badía J, Aguilar J, Baldoma L. 2008. Dual role of LldR in regulation of the *lldPRD* operon, involved in L-lactate metabolism in *Escherichia coli*. J Bacteriol. 190(8): 2997-3005.

Alifano P, Fani R, Liò P, Lazcano A, Bazzicalupo M, Carlomagno MS, Bruni CB. 1996. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol Rev. 60(1): 44-69.

Alkhalfioui F, Renard M, Frendo P, Keichinger C, Meyer Y, Gelhaye E, Hirasawa M, Knaff DB, Ritzenthaler C, Montrichard F. 2008. A novel type of thioredoxin dedicated to symbiosis in legumes. Plant Physiol. 148(1): 424-35.

Aly KA, Baron C. 2007. The VirB5 protein localizes to the T-pilus tips in *Agrobacterium tumefaciens*. Microbiology. 153(Pt 11): 3766-75.

Angerer A, Gaisser S, Braun V. 1990. Nucleotide sequences of the *sfuA*, *sfuB*, and *sfuC* genes of *Serratia marcescens* suggest a periplasmic-binding-protein-dependent iron transport mechanism. J Bacteriol. 172(2): 572-8.

Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, Gurjal M, Hong A, Huizar L, Hyman RW, Kahn D, Kahn ML, Kalman S, Keating DH, Palm C, Peck MC, Surzycki R, Wells DH, Yeh KC, Davis RW, Federspiel NA, Long SR. 2001. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc Natl Acad Sci U S A. 98(17): 9883-8.

Barrios H, Valderrama B, Morett E. 1999. Compilation and analysis of sigma (54)-dependent promoter sequences. Nucleic Acids Res. 27(22): 4305-13.

Becker A, Bergès H, Krol E, Bruand C, Rüberg S, Capela D, Lauber E, Meilhoc E, Ampe F, de Bruijn FJ, Fourment J, Francez-Charlot A, Kahn D, Küster H, Liebe C, Pühler A, Weidner S, Batut J. 2004. Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. Mol Plant Microbe Interact. 17(3): 292-303.

Becker A, Küster H, Niehaus K, Pühler A. 1995. Extension of the *Rhizobium meliloti* succinoglycan biosynthesis gene cluster: identification of the *exsA* gene encoding an ABC transporter protein, and the *exsB* gene which probably codes for a regulator of succinoglycan biosynthesis. Mol Gen Genet. 249(5): 487-97.

Beijerinckm, W. 1888. Die Bacterien der Papilionaceen-Kniillchen. Rot. Ztg., 46, 725, 757.

- Belitsky BR, Gustafsson MC, Sonenshein AL, Von Wachenfeldt C. 1997. An *Irp*-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. *J Bacteriol.* 179(17): 5448-57.
- Belitsky BR, Sonenshein AL. An enhancer element located downstream of the major glutamate dehydrogenase gene of *Bacillus subtilis*. 1999. *Proc Natl Acad Sci U S A.* 96(18): 10290-5.
- Benhassine T, Fauvart M, Vanderleyden J, Michiels J. 2007. Interaction of an IHF-like protein with the *Rhizobium etli nifA* promoter. *FEMS Microbiol Lett.* 271(1): 20-6.
- Bergès H, Lauber E, Liebe C, Batut J, Kahn D, de Bruijn FJ, Ampe F. 2003. Development of *Sinorhizobium meliloti* pilot macroarrays for transcriptome analysis. *Appl Environ Microbiol.* 69(2): 1214-9.
- Bernardelli CE, Luna MF, Galar ML, Boiardi JL. 2001. Periplasmic PQQ-dependent glucose oxidation in free-living and symbiotic rhizobia. *Curr Microbiol.* 42(5): 310-5.
- Berrios-Rivera SJ, Bennett GN, San KY. 2002. Metabolic engineering of *Escherichia coli*: increase of NADH availability by overexpressing an NAD(+)-dependent formate dehydrogenase. *Metab Eng.* 4(3): 217-29.
- Bevers LE, Hagedoorn PL, Santamaria-Araujo JA, Magalon A, Hagen WR, Schwarz G. 2008. Function of MoaB proteins in the biosynthesis of the molybdenum and tungsten cofactors. *Biochemistry.* 47(3): 949-56.
- Bianciotto V, Bandi C, Minerdi D, Sironi M, Tichy HV, Bonfante P. 1996. An obligately endosymbiotic mycorrhizal fungus itself harbors obligately intracellular bacteria. *Appl Environ Microbiol.* 62: 3005–3010.
- Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J, McClean P, Attene G, Papa R. 2012. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc Natl Acad Sci U S A.* 109(14): E788-96.
- Bittner AN, Foltz A, Oke V. 2007. Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. *J Bacteriol.* 189(5): 1884-9.
- Black DS, Irwin B, Moyed HS. 1994. Autoregulation of *hip*, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis. *J Bacteriol.* 176(13): 4081-91.
- Bobik C, Meilhoc E, Batut J. 2006. FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *J Bacteriol.* 188(13): 4890-902.

- Bongaerts J, Zoske S, Weidner U, Uden G. 1995. Transcriptional regulation of the proton translocating NADH dehydrogenase genes (*nuoA-N*) of *Escherichia coli* by electron acceptors, electron donors and gene regulators. *Mol Microbiol.* 16(3): 521-34.
- Bonomi HR, Marchesini MI, Klinke S, Ugalde JE, Zylberman V, Ugalde RA, Comerchi DJ, Goldbaum FA. 2010. An atypical riboflavin pathway is essential for *Brucella abortus* virulence. *PLoS One.* 5(2): e9435.
- Braunstein M, Brown AM, Kurtz S, Jacobs WR Jr. 2001. Two nonredundant SecA homologues function in mycobacteria. *J Bacteriol.* 183(24): 6979-90.
- Buck M. 1986. Deletion analysis of the *Klebsiella pneumoniae* nitrogenase promoter: importance of spacing between conserved sequences around positions -12 and -24 for activation by the *nifA* and *ntnC* (*glnG*) products. *J Bacteriol.* 166(2): 545-51.
- Buckel SD, Bell AW, Rao JK, Hermodson MA. 1986. An analysis of the structure of the product of the *rbsA* gene of *Escherichia coli* K12. *J Biol Chem.* 261(17): 7659-62.
- Cabanes D, Boistard P, Batut J. 2000. Symbiotic induction of pyruvate dehydrogenase genes from *Sinorhizobium meliloti*. *Mol Plant Microbe Interact.* 13(5): 483-93.
- Capela D, Filipe C, Bobik C, Batut J, Bruand C. 2006. *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. *Mol Plant Microbe Interact.* 19(4): 363-72.
- Carter RA, Worsley PS, Sawers G, Challis GL, Dilworth MJ, Carson KC, Lawrence JA, Wexler M, Johnston AW, Yeoman KH. 2002. The *vbs* genes that direct synthesis of the siderophore vibriobactin in *Rhizobium leguminosarum*: their expression in other genera requires ECF sigma factor Rpol. *Mol Microbiol.* 44(5): 1153-66.
- Carter RA, Yeoman KH, Klein A, Hosie AH, Sawers G, Poole PS, Johnston AW. 2002. *dpp* genes of *Rhizobium leguminosarum* specify uptake of delta-aminolevulinic acid. *Mol Plant Microbe Interact.* 15(1): 69-74.
- Charles TC, Cai GQ, Aneja P. 1997. Megaplasmid and chromosomal loci for the PHB degradation pathway in *Rhizobium (Sinorhizobium) meliloti*. *Genetics.* 146(4): 1211-20.
- Chen H, Gao K, Kondorosi E, Kondorosi A, Rolfe BG. 2005. Functional genomic analysis of global regulator NolR in *Sinorhizobium meliloti*. *Mol Plant Microbe Interact.* 18(12): 1340-52.

Cho MJ, Kim YH, Shin K, Kim YK, Kim YS, Kim TJ. 2010. Symbiotic adaptation of bacteria in the gut of *Reticulitermes speratus*: low endo-beta-1,4-glucanase activity. *Biochem Biophys Res Commun.* 395(3): 432-5.

Cianciotto NP, Cornelis P, Baysse C. 2005. Impact of the bacterial type I cytochrome c maturation system on different biological processes. *Mol Microbiol.* 56(6): 1408-15.

Cloutier J, Laberge S, Prévost D, Antoun H. 1996. Sequence and mutational analysis of the common *nodBCIJ* region of *Rhizobium* sp. (*Oxytropis arctobia*) strain N33, a nitrogen-fixing microsymbiont of both arctic and temperate legumes. *Mol Plant Microbe Interact.* 9(6): 523-31.

Collavino M, Riccillo PM, Grasso DH, Crespi M, Aguilar M. 2005. GuaB activity is required in *Rhizobium tropici* during the early stages of nodulation of determinate nodules but is dispensable for the *Sinorhizobium meliloti*-alfalfa symbiotic interaction. *Mol Plant Microbe Interact.* 18(7): 742-50.

Cooper CL, Hsu L, Jackowski S, Rock CO. 1989. 2-Acylglycerolphosphoethanolamine acyltransferase/acyl-acyl carrier protein synthetase is a membrane-associated acyl carrier protein binding protein. *J Biol Chem.* 264(13): 7384-9.

Cuadrado B, Rubio G, Santos W. 2009. Caracterización de cepas de *rhizobium* y *bradyrhizobium* (con habilidad de nodulación) seleccionados de los cultivos de fríjol caupi (*Vigna unguiculata*) como potenciales bioinóculos. *Revista Colombiana de Ciencias Químico Farmacéuticas*, 38(1).

Daniels R, De Vos DE, Desair J, Raedschelders G, Luyten E, Rosemeyer V, Verreth C, Schoeters E, Vanderleyden J, Michiels J. 2002. The cin quorum sensing locus of *Rhizobium etli* CNPAF512 affects growth and symbiotic nitrogen fixation. *J Biol Chem.* 277(1): 462-8.

Danino VE, Wilkinson A, Edwards A, Downie JA. 2003. Recipient-induced transfer of the symbiotic plasmid pRL1JI in *Rhizobium leguminosarum* bv. viciae is regulated by a quorum-sensing relay. *Mol Microbiol.* 50(2): 511-25.

Das A, Xie YH. 2000. The *Agrobacterium* T-DNA transport pore proteins VirB8, VirB9, and VirB10 interact with one another. *J Bacteriol.* 182(3): 758-63.

Davies BW, Walker GC. 2007. Identification of novel *Sinorhizobium meliloti* mutants compromised for oxidative stress protection and symbiosis. *J Bacteriol.* 189(5): 2110-3.

De Costa DM, Suzuki K, Satou M, Yoshida K. 2001. Genome analysis of *Agrobacterium tumefaciens*: linkage map and genetic features of the left region of the linear chromosome. *Genes Genet Syst.* 76(6): 363-71.

de Lucena DK, Pühler A, Weidner S. 2010. The role of sigma factor RpoH1 in the pH stress response of *Sinorhizobium meliloti*. *BMC Microbiol.* 10: 265.

D'hooghe I, Michiels J, Vlassak K, Verreth C, Waelkens F, Vanderleyden J. 1995. Structural and functional analysis of the fixLJ genes of *Rhizobium leguminosarum* biovar phaseoli CNPAF512. *Mol Gen Genet.* 249(1): 117-26.

Ding H, Yip CB, Geddes BA, Oresnik IJ, Hynes MF. 2012. Glycerol utilization by *Rhizobium leguminosarum* requires an ABC transporter and affects competition for nodulation. *Microbiology.* 158(Pt 5): 1369-78.

Dombrecht B, Heusdens C, Beullens S, Verreth C, Mulkers E, Proost P, Vanderleyden J, Michiels J. 2005. Defence of *Rhizobium etli* bacteroids against oxidative stress involves a complexly regulated atypical 2-Cys peroxiredoxin. *Mol Microbiol.* 55(4): 1207-21.

Dombrecht B, Marchal K, Vanderleyden J, Michiels J. 2002. Prediction and overview of the RpoN-regulon in closely related species of the Rhizobiales. *Genome Biol.* 3(12): RESEARCH0076.

Domínguez-Ferreras A, Pérez-Arnedo R, Becker A, Olivares J, Soto MJ, Sanjuán J. 2006. Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. *J Bacteriol.* 188(21): 7617-25.

Duong F, Wickner W. 1997. The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO J.* 16(16): 4871-9.

Egawa T, Ganesan K, Lin MT, Yu MA, Hosler JP, Yeh SR, Rousseau DL, Gennis RB. Differential effects of glutamate-286 mutations in the aa(3)-type cytochrome c oxidase from *Rhodobacter sphaeroides* and the cytochrome bo(3) ubiquinol oxidase from *Escherichia coli*. *Biochim Biophys Acta.* 2011 1807(10): 1342-8.

Ellison DW, Miller VL. 2006. Regulation of virulence by members of the MarR/SlyA family. *Curr Opin Microbiol.* 9(2): 153-9.

Encarnación S, del Carmen Vargas M, Dunn MF, Dávalos A, Mendoza G, Mora Y, Mora J. 2002. AniA regulates reserve polymer accumulation and global protein expression in *Rhizobium etli*. *J Bacteriol.* 184(8): 2287-95.

- Encarnación S, Guzmán Y, Dunn MF, Hernández M, del Carmen Vargas M, Mora J. 2003. Proteome analysis of aerobic and fermentative metabolism in *Rhizobium etli* CE3. *Proteomics*. 3(6): 1077-85.
- Engelke T, Jording D, Kapp D, Pühler A. 1989. Identification and sequence analysis of the *Rhizobium meliloti* *dctA* gene encoding the C4-dicarboxylate carrier. *J Bacteriol*. 171(10): 5551-60.
- Eskra L, Canavessi A, Carey M, Splitter G. 2001. *Brucella abortus* genes identified following constitutive growth and macrophage infection. *Infect Immun*. 69(12): 7736-42.
- Fales L, Kryszak L, Zeilstra-Ryalls J. 2001. Control of *hemA* expression in *Rhodobacter sphaeroides* 2.4.1: effect of a transposon insertion in the *hbdA* gene. *J Bacteriol*. 183(5): 1568-76.
- Fauvart M, Michiels J. 2008. Rhizobial secreted proteins as determinants of host specificity in the rhizobium-legume symbiosis. *FEMS Microbiol Lett*. 285(1): 1-9.
- Fay A, Dworkin J. 2009. *Bacillus subtilis* homologs of MviN (MurJ), the putative *Escherichia coli* lipid II flippase, are not essential for growth. *J Bacteriol*. 191(19): 6020-8.
- Ferluga S, Venturi V. 2009. OryR is a LuxR-family protein involved in interkingdom signaling between pathogenic *Xanthomonas oryzae* pv. *oryzae* and rice. *J Bacteriol*. 191(3): 890-7.
- Fernandez de Henestrosa AR, Barbé J. 1991. Induction of the *alkA* gene of *Escherichia coli* in gram-negative bacteria. *J Bacteriol*. 173(23): 7736-40.
- Finan TM, Oresnik I, Bottacin A. 1988. Mutants of *Rhizobium meliloti* defective in succinate metabolism. *J Bacteriol*. 170(8): 3396-403.
- Finnie C, Zorreguieta A, Hartley NM, Downie JA. 1998. Characterization of *Rhizobium leguminosarum* exopolysaccharide glycanases that are secreted via a type I exporter and have a novel heptapeptide repeat motif. *J Bacteriol*. 180(7): 1691-9.
- Fischer HM, Babst M, Kaspar T, Acuna G, Arigoni F, & Hennecke H. 1993. One member of a *gro-ESL*-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. *EMBO J*. 12(7): 2901-12.
- Fisher RF, Long SR. 1993. Interactions of NodD at the *nod* Box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. *J Mol Biol*. 233(3): 336-48.

Flemming D, Stolpe S, Schneider D, Hellwig P, Friedrich T. 2005. A possible role for iron-sulfur cluster N2 in proton translocation by the NADH: ubiquinone oxidoreductase (complex I). *J Mol Microbiol Biotechnol.* 10(2-4): 208-22.

Flores-Mireles AL, Eberhard A, Winans SC. 2012. *Agrobacterium tumefaciens* can obtain sulfur from an opine that is synthesized by octopine synthase using S-methylmethionine as a substrate. *Mol Microbiol.* 84(5): 845-56.

Foreman DL, Vanderlinde EM, Bay DC, Yost CK. 2010. Characterization of a gene family of outer membrane proteins (*ropB*) in *Rhizobium leguminosarum* bv. *viciae* VF39SM and the role of the sensor kinase ChvG in their regulation. *J Bacteriol.* 192(4): 975-83.

Fry J, Wood M, Poole PS. 2001. Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Mol Plant Microbe Interact.* 14(8): 1016-25.

Fuchsius, L. 1542. *De Historia Stirpium Commentarii Insignes*. Ed. Palma Ising. Basileae, Suiza. A4, Pag.6.

Fujisawa M, Misawa N. 2010. Enrichment of carotenoids in flaxseed by introducing a bacterial phytoene synthase gene. *Methods Mol Biol.* 643: 201-11.

Garcia PP, Bringhurst RM, Arango Pinedo C, Gage DJ. 2010. Characterization of a two-component regulatory system that regulates succinate-mediated catabolite repression in *Sinorhizobium meliloti*. *J Bacteriol.* 192(21): 5725-35.

García-Rodríguez FM, Zekri S, Toro N. 2000. Characterization of the *Sinorhizobium meliloti* genes encoding a functional dihydrodipicolinate synthase (*dapA*) and dihydrodipicolinate reductase (*dapB*). *Arch Microbiol.* 173(5-6): 438-44.

Gilbert KB, Vanderlinde EM, Yost CK. 2007. Mutagenesis of the carboxy terminal protease CtpA decreases desiccation tolerance in *Rhizobium leguminosarum*. *FEMS Microbiol Lett.* 272(1): 65-74.

Girard L, Brom S, Dávalos A, López O, Soberón M, Romero D. 2000. Differential regulation of *fixN*-reiterated genes in *Rhizobium etli* by a novel *fixL*-*fixK* cascade. *Mol Plant Microbe Interact.* 13(12): 1283-92.

Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Verméglie A, Médigue

C, Sadowsky M. 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science*. 316(5829): 1307-12.

Glucksmann MA, Reuber TL, Walker GC. 1993. Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. *J Bacteriol*. 175(21): 7045-55.

Gómez-Hernández N, Reyes-González A, Sánchez C, Mora Y, Delgado MJ, Girard L. 2011. Regulation and symbiotic role of *nirK* and *norC* expression in *Rhizobium etli*. *Mol Plant Microbe Interact*. 24(2): 233-45.

Gong Z, Zhu J, Yu G, Zou H. 2007. Disruption of *nifA* gene influences multiple cellular processes in *Sinorhizobium meliloti*. *J Genet Genomics*. 34(9): 783-9.

González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G. 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci U S A*. 103(10): 3834-9.

González V, Bustos P, Ramirez-Romero MA, Medrano-Soto A, Salgado H, Hernandez-Gonzalez I, Hernandez-Celis JC, Quintero V, Moreno-Hagelsieb G, Girard L, Rodriguez O, Flores M, Cevallos MA, Collado-Vides J, Romero D, Davila G. 2003. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol*. 4: R36.

Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MH Jr. 2004. Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J Bacteriol*. 186(11): 3516-24.

Granados-Baeza MJ, Gómez-Hernández N, Mora Y, Delgado MJ, Romero D, Girard L. 2007. Novel reiterated Fnr-type proteins control the production of the symbiotic terminal oxidase *cbb3* in *Rhizobium etli* CFN42. *Mol Plant Microbe Interact*. 20(10): 1241-9.

Green LS, Li Y, Emerich DW, Bergersen FJ, Day DA. 2000. Catabolism of alpha-ketoglutarate by a *sucA* mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. *J Bacteriol*. 182(10): 2838-44.

Griffin AM, Morris VJ, Gasson MJ. 1996. Genetic analysis of the acetan biosynthetic pathway in *Acetobacter xylinum*: nucleotide sequence analysis of the *aceB*, *aceC*, *aceD* and *aceE* genes. *DNA Seq*. 6(5): 275-84.

Grzemeski W, Akowski JP, Kahn ML. 2005. Probing the *Sinorhizobium meliloti*-alfalfa symbiosis using temperature-sensitive and impaired function citrate synthase mutants. *Mol. Plant-Microbe Interact.* 18: 134–141.

Güell M, Yus E, Lluch-Senar M, Serrano L. 2011. Bacterial transcriptomics: what is beyond the RNA hori-z-ome? *Nat Rev Microbiol.* 9(9): 658-69.

Guo W, Zou LF, Li YR, Cui YP, Ji ZY, Cai LL, Zou HS, Hutchins WC, Yang CH, Chen GY. 2012. Fructose-bisphosphate aldolase exhibits functional roles between carbon metabolism and the *hrp* system in rice pathogen *Xanthomonas oryzae* pv. *oryzicola*. *PLoS One.* 7(2): e31855.

Gutterson N, Ziegler JS, Warren GJ, Layton TJ. 1988. Genetic determinants for catabolite induction of antibiotic biosynthesis in *Pseudomonas fluorescens* HV37a. *J Bacteriol.* 170(1): 380-5.

Hankins JS, Denroche H, Mackie GA. 2010. Interactions of the RNA-binding protein Hfq with *cspA* mRNA, encoding the major cold shock protein. *J Bacteriol.* 192(10): 2482-90.

Haraszthy VI, Jordan SF, Zambon JJ. 2006. Identification of Fur-regulated genes in *Actinobacillus actinomycetemcomitans*. *Microbiology.* 152(Pt 3): 787-96.

Hassani R, Prasad CK, Vineetha KE, Vij N, Singh P, Sud R, Yadav S, Randhawa GS. 2002. Symbiotic characterization of isoleucine+valine and leucine auxotrophs of *Sinorhizobium meliloti*. *Indian J Exp Biol.* 40(10): 1110-20.

Hauberg-Lotte L, Klingenberg H, Scharf C, Böhm M, Plessl J, Friedrich F, Völker U, Becker A, Reinhold-Hurek B. 2012. Environmental factors affecting the expression of *pilAB* as well as the proteome and transcriptome of the grass endophyte *Azoarcus* sp. strain BH72. *PLoS One.* 7(1): e30421.

Hauser F, Lindemann A, Vuilleumier S, Patrignani A, Schlapbach R, Fischer HM, Hennecke H. 2006. Design and validation of a partial-genome microarray for transcriptional profiling of the *Bradyrhizobium japonicum* symbiotic gene region. *Mol Genet Genomics.* 275(1): 55-67.

Hauser F, Pessi G, Friberg M, Weber C, Rusca N, Lindemann A, Fischer HM, Hennecke H. 2007. Dissection of the *Bradyrhizobium japonicum* NifA+sigma54 regulon, and identification of a ferredoxin gene (*fdxN*) for symbiotic nitrogen fixation. *Mol Genet Genomics.* 278(3): 255-71.

- Hawkins FK, Johnston AW. 1988. Transcription of a *Rhizobium leguminosarum* biovar phaseoli gene needed for melanin synthesis is activated by *nifA* of *Rhizobium* and *Klebsiella pneumoniae*. *Mol Microbiol.* 2(3): 331-7.
- Heath KD, Burke PV, Stinchcombe JR. 2012. Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Mol Ecol.* 21(19): 4735-47.
- Hellriegel H. y Wilfarth H. 1888. Untersuchungen über die stickstoffnahrung der gramineen und leguminosen. Beilageheft zu der Ztschr. Ver. Rübenzucker-Industrie Deutschen Reichs.
- Herrgård MJ, Covert MW, Palsson BØ. 2004. Reconstruction of microbial transcriptional regulatory networks. *Curr Opin Biotechnol.* 15(1): 70-7.
- Herridge DF, Peoples MB, Boddey RM. 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil.* 311: 1–18
- Heungens K, Cowles CE, Goodrich-Blair H. 2002. Identification of *Xenorhabdus nematophila* genes required for mutualistic colonization of *Steinernema carpocapsae* nematodes. *Molecular Microbiology.* 45(5): 1337-53.
- Hinz A, Lee S, Jacoby K, Manoil C. 2011. Membrane proteases and aminoglycoside antibiotic resistance. *J Bacteriol.* 193(18): 4790-7.
- Hohle TH, O'Brian MR. The *mntH* gene encodes the major Mn(2+) transporter in *Bradyrhizobium japonicum* and is regulated by manganese via the Fur protein. 2009. *Mol Microbiol.* 72(2): 399-409.
- Horsburgh MJ, Clements MO, Crossley H, Ingham E, Foster SJ. 2001. PerR controls oxidative stress resistance and iron storage proteins and is required for virulence in *Staphylococcus aureus*. *Infect Immun.* 69(6): 3744-54.
- Hoshino T, Kose-Terai K, Sato K. 1992. Solubilization and reconstitution of the *Pseudomonas aeruginosa* high affinity branched-chain amino acid transport system. *J Biol Chem.* 267(30): 21313-8.
- Howden AJ, Preston GM. 2009. Nitrilase enzymes and their role in plant-microbe interactions. *Microb Biotechnol.* 2(4): 441-51.
- Humann JL, Ziemkiewicz HT, Yurgel SN, Kahn ML. 2009. Regulatory and DNA repair genes contribute to the desiccation resistance of *Sinorhizobium meliloti* Rm1021. *Appl Environ Microbiol.* 75(2): 446-53.

- Hunt TP, Magasanik B. 1985. Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. Proc Natl Acad Sci U S A. 82(24): 8453-7.
- Jabbouri S, Fellay R, Talmont F, Kamalaprija P, Burger U, Relić B, Promé JC, Broughton WJ. 1995. Involvement of *nodS* in N-methylation and *nodU* in 6-O-carbamoylation of *Rhizobium* sp. NGR234 nod factors. J Biol Chem. 270(39): 22968-73.
- Jahn OJ, Davila G, Romero D, Noel KD. 2003. BacS: an abundant bacteroid protein in *Rhizobium etli* whose expression ex planta requires *nifA*. Mol. Plant-Microbe Interact. 16: 65–73.
- Janczarek M, Kutkowska J, Piersiak T, Skorupska A. 2010. *Rhizobium leguminosarum* bv. trifolii *rosR* is required for interaction with clover, biofilm formation and adaptation to the environment. BMC Microbiol. 10: 284.
- Janczarek M, Skorupska A. 2011. Modulation of *rosR* Expression and Exopolysaccharide Production in *Rhizobium leguminosarum* bv. trifolii by Phosphate and Clover Root Exudates. Int J Mol Sci. 12(6): 4132-55.
- Jensen JB, Peters NK, Bhuvaneswari TV. 2002. Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti*. J Bacteriol. 184(11): 2978-86.
- Jiang G, Krishnan AH, Kim YW, Wacek TJ, Krishnan HB. 2001. A functional myo-inositol dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of *Sinorhizobium fredii* USDA191 to nodulate soybean (*Glycine max* [L.] Merr.). J Bacteriol. 183(8): 2595-604.
- Kaminski PA, Elmerich C. 1998. The control of *Azorhizobium caulinodans nifA* expression by oxygen, ammonia and by the HF-I-like protein, NrfA. Mol Microbiol. 28(3): 603-13.
- Karunakaran R, Haag AF, East AK, Ramachandran VK, Prell J, James EK, Scocchi M, Ferguson GP, Poole PS. 2010. BacA is essential for bacteroid development in nodules of galeoid, but not phaseoloid, legumes. J Bacteriol. 192(11): 2920-8.
- Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS. 2009. Transcriptomic analysis of *R. leguminosarum* biovar viciae in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. J Bacteriol. 191(12): 4002-14.
- Kemner JM, Liang X, Nester EW. 1997. The *Agrobacterium tumefaciens* virulence gene *chvE* is part of a putative ABC-type sugar transport operon. J Bacteriol. 179(7): 2452–2458.

Keseler IM, Kaiser D. 1997. sigma54, a vital protein for *Myxococcus xanthus*. Proc Natl Acad Sci U S A. 94(5): 1979-84.

Kiers ET, Rousseau RA, West SA, Denison RF. 2003. Host sanctions and the legume-rhizobium mutualism. Nature. 425(6953): 78-81.

Kijne, J. W. 1975. The fine structure of pea root nodules. 2. Senescence and disintegration of the bacteroid tissue. Physiological Plant Pathology 7: 17–21.

Kim H, Farrand SK. 1998. Opine catabolic loci from *Agrobacterium* plasmids confer chemotaxis to their cognate substrates. Mol Plant Microbe Interact. 11(2): 131-43.

Kim YS. 2002. Malonate metabolism: biochemistry, molecular biology, physiology, and industrial application. J Biochem Mol Biol. 35(5): 443-51.

Kobza K, Sarath G, Zemleni J. 2008. Prokaryotic BirA ligase biotinylates K4, K9, K18 and K23 in histone H3. BMB Rep. 41(4): 310-5.

Köhler T, Epp SF, Curty LK, Pechère JC. 1999. Characterization of MexT, the regulator of the MexE-MexF-OprN multidrug efflux system of *Pseudomonas aeruginosa*. J Bacteriol. 181(20): 6300-5.

Kok RG, D'Argenio DA, Ornston LN. 1998. Mutation analysis of PobR and PcaU, closely related transcriptional activators in acinetobacter. Bacteriol. 180(19): 5058-69.

Kozioł U, Hannibal L, Rodríguez MC, Fabiano E, Kahn ML, Noya F. 2009. Deletion of citrate synthase restores growth of *Sinorhizobium meliloti* 1021 aconitase mutants. J Bacteriol. 191(24): 7581-6.

Kroll TC, Wöfl S. 2002. Ranking: a closer look on globalisation methods for normalisation of gene expression arrays. Nucleic Acids Res. 30(11): e50.

Kumar A, Vij N, Randhawa GS. 2003. Isolation and symbiotic characterization of transposon Tn5-induced arginine auxotrophs of *Sinorhizobium meliloti*. Indian J Exp Biol. 41(10): 1198-204.

Kustu S, Santero E, Keener J, Popham D, Weiss D. 1989. Expression of sigma 54 (*ntrA*)-dependent genes is probably united by a common mechanism. Microbiol Rev. 53(3): 367-76.

Lambert A, Østerås M, Mandon K, Poggi MC, Le Rudulier D. 2001. Fructose uptake in *Sinorhizobium meliloti* is mediated by a high-affinity ATP-binding cassette transport system. J Bacteriol. 183(16): 4709-17.

- Landeta C, Dávalos A, Cevallos MÁ, Geiger O, Brom S, Romero D. 2011. Plasmids with a chromosome-like role in rhizobia. *J Bacteriol.* 193(6): 1317-26.
- Langklotz S, Schäkermann M, Narberhaus F. 2011. Control of Lipopolysaccharide Biosynthesis by FtsH-Mediated Proteolysis of LpxC Is Conserved in Enterobacteria but Not in All Gram-Negative Bacteria. *J Bacteriol.* 193(5): 1090–1097.
- Larrainzar E, Wienkoop S, Weckwerth W, Ladrera R, Arrese-Igor C, González EM. 2007. *Medicago truncatula* root nodule proteome analysis reveals differential plant and bacteroid responses to drought stress. *Plant Physiol.* 144(3): 1495-507.
- Lazzaroni JC, Germon P, Ray MC, Vianney A. 1999. The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. *FEMS Microbiol Lett.* 177(2): 191-7.
- Leang C, Krushkal J, Ueki T, Puljic M, Sun J, Juárez K, Núñez C, Reguera G, DiDonato R, Postier B, Adkins RM, Lovley DR. 2009. Genome-wide analysis of the RpoN regulon in *Geobacter sulfurreducens*. *BMC Genomics.* 10: 331.
- Levitan O, Kranz SA, Spungin D, Prásil O, Rost B, Berman-Frank I. 2010. Combined effects of CO₂ and light on the N₂-fixing cyanobacterium *Trichodesmium* IMS101: a mechanistic view. *Plant Physiol.* 154(1): 346-56.
- Li F, Hou B, Hong G. 2008. Symbiotic plasmid is required for NoIR to fully repress nodulation genes in *Rhizobium leguminosarum* A34. *Acta Biochim Biophys Sin (Shanghai).* 40(10): 901-7.
- Li H, Rhodius V, Gross C, Siggia ED. 2002. Identification of the binding sites of regulatory proteins in bacterial genomes. *Proc Natl Acad Sci U S A.* 99(18): 11772-7.
- Liu X, Taber HW. 1998. Catabolite regulation of the *Bacillus subtilis* *ctaBCDEF* gene cluster. *J Bacteriol.* 180(23): 6154-63.
- Lopez O, Morera C, Miranda-Rios J, Girard L, Romero D, Soberón M. 2001. Regulation of gene expression in response to oxygen in *Rhizobium etli*: role of FnrN in *fixNOQP* expression and in symbiotic nitrogen fixation. *J Bacteriol.* 183(24): 6999-7006.
- Lowry JE, Goodridge L, Vernati G, Fluegel AM, Edwards WH, Andrews GP. 2010. Identification of *Brucella abortus* genes in elk (*Cervus elaphus*) using in vivo-induced antigen technology (IVIAT) reveals novel markers of infection. *Vet Microbiology.* 142(3-4): 367-72.

Lundgren HK, Björk GR. 2006. Structural alterations of the cysteine desulfurase IscS of *Salmonella enterica* serovar Typhimurium reveal substrate specificity of IscS in tRNA thiolation. *J Bacteriol.* 188(8): 3052-62.

Madigan MT, Martinko JM, Parker J. 2000. Brock Biology of Microorganisms. Edi. Prentice Hall, Upper Saddle River. Novena edición. Cap. 12.9.

Malpighi, M. 1675. Anatomie plantarum: Cui subjungitur appendix, iteratas. Impensis Johannis Martyn. Pag. 103. Fig. 105B.

Mandon K, Michel-Reydellet N, Encarnación S, Kaminski PA, Leija A, Cevallos MA, Elmerich C, Mora J. 1998. Poly-beta-hydroxybutyrate turnover in *Azorhizobium caulinodans* is required for growth and affects *nifA* expression. *J Bacteriol.* 180(19): 5070-6.

Marais A, Mendz GL, Hazell SL, Mégraud F. 1999. Metabolism and genetics of *Helicobacter pylori*: the genome era. *Microbiol Mol Biol Rev.* 63(3): 642-74.

Marco DE, Carbajal JP, Cannas S, Pérez-Arnedo R, Hidalgo-Perea A, Olivares J, Ruiz-Sainz JE, Sanjuán J. 2009. An experimental and modelling exploration of the host-sanction hypothesis in legume-rhizobia mutualism. *J Theor Biol.* 259(3): 423-33.

Marczak M, Mazur A, Gruszecki WI, Skorupska A. 2008. PssO, a unique extracellular protein important for exopolysaccharide synthesis in *Rhizobium leguminosarum* bv. trifolii. *Biochimie.* 90(11-12): 1781-90.

Marroquí S, Zorreguieta A, Santamaría C, Temprano F, Soberón M, Megías M, Downie JA. 2001. Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants. *Journal of Bacteriology.* 183(3): 854-64.

Martínez-Salazar JM, Salazar E, Encarnación S, Ramírez-Romero MA, Rivera J. 2009. Role of the extracytoplasmic function sigma factor RpoE4 in oxidative and osmotic stress responses in *Rhizobium etli*. *J Bacteriol.* 191(13): 4122-32.

Martínez-Salazar JM, Sandoval-Calderón M, Guo X, Castillo-Ramírez S, Reyes A, Loza MG, Rivera J, Alvarado-Affantranger X, Sánchez F, González V, Dávila G, Ramírez-Romero MA. 2009. The *Rhizobium etli* RpoH1 and RpoH2 sigma factors are involved in different stress responses. *Microbiology.* 155(Pt 2): 386-97.

Martínez-Salazar JM, Zuñiga-Castillo J, Romero D. 2009. Differential roles of proteins involved in migration of Holliday junctions on recombination and tolerance to DNA damaging agents in *Rhizobium etli*. *Gene.* 432(1-2): 26-32.

- Martins-Pinheiro M, Marques RC, Menck CF. 2007. Genome analysis of DNA repair genes in the alpha proteobacterium *Caulobacter crescentus*. BMC Microbiol. 12: 7-17.
- McIntyre HJ, Davies H, Hore TA, Miller SH, Dufour JP, Ronson CW. 2007. Trehalose biosynthesis in *Rhizobium leguminosarum* bv. trifolii and its role in desiccation tolerance. Appl Environ Microbiol. 73(12): 3984-92.
- Mellegård H, Kovács AT, Lindbäck T, Christensen BE, Kuipers OP, Granum PE. 2011. Transcriptional Responses of *Bacillus cereus* towards Challenges with the Polysaccharide Chitosan. PLoS One. 6(9): e24304.
- Mercer KL, Weiss DS. 2002. The *Escherichia coli* cell division protein FtsW is required to recruit its cognate transpeptidase, FtsI (PBP3), to the division site. J Bacteriol. 184(4): 904-12.
- Mergaert P, D'Haese W, Fernández-López M, Geelen D, Goethals K, Promé JC, Van Montagu M, Holsters M. Fucosylation and arabinosylation of Nod factors in *Azorhizobium caulinodans*: involvement of *nolK*, *nodZ* as well as *noeC* and/or downstream genes. Mol Microbiol. 1996 21(2): 409-19.
- Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proc Natl Acad Sci U S A. 103(13): 5230-5.
- Michel-Reydellet N, Kaminski PA. 1999. *Azorhizobium caulinodans* PII and GlnK proteins control nitrogen fixation and ammonia assimilation. J Bacteriol. 181(8): 2655-8.
- Michiels J, D'hooghe I, Verreth C, Pelemans H, Vanderleyden J. 1994. Characterization of the *Rhizobium leguminosarum* biovar phaseoli *nifA* gene, a positive regulator of nif gene expression. Arch Microbiol. 161(5): 404-8.
- Michiels J, Moris M, Dombrecht B, Verreth C, Vanderleyden J. 1998. Differential regulation of *Rhizobium etli* *rpoN2* gene expression during symbiosis and free-living growth. J. Bacteriol. 180: 3620–3628.
- Michiels J, Van Soom T, D'Hooghe I, Dombrecht B, Benhassine T, de Wilde P, Vanderleyden J. 1998. The *Rhizobium etli* *rpoN* locus: DNA sequence analysis and phenotypical characterization of *rpoN*, *ptsN*, and *ptsA* mutants. J. Bacteriol. 180: 1729–1740.

- Miller WG, Parker CT, Rubenfield M, Mendz GL, Wösten MM, Ussery DW, Stolz JF, Binnewies TT, Hallin PF, Wang G, Malek JA, Rogosin A, Stanker LH, Mandrell RE. 2007. The complete genome sequence and analysis of the epsilonproteobacterium *Arcobacter butzleri*. PLoS One. 2(12): e1358.
- Minamino T, Shimada M, Okabe M, Saijo-Hamano Y, Imada K, Kihara M, Namba K. 2010. Role of the C-terminal cytoplasmic domain of FlhA in bacterial flagellar type III protein export. J Bacteriol. 192(7): 1929-36.
- Minerdi D, Fani R, Bonfante P. 2002. Identification and evolutionary analysis of putative cytoplasmic mcpA-like protein in a bacterial strain living in symbiosis with a mycorrhizal fungus. J Mol Evol. 54(6): 815-24.
- Molina-López JA, Santero E. An artificial enhancer with multiple response elements stimulates prokaryotic transcriptional activation mediated by various regulatory proteins. Mol Gen Genet. 1999 262(2): 291-301.
- Moreno S, Patriarca EJ, Chiurazzi M, Meza R, Defez R, Lamberti A, Riccio A, Iaccarino M, Espin G. 1992. Phenotype of a *Rhizobium leguminosarum ntrC* mutant. Res Microbiol. 143(2): 161-71.
- Morett E, Buck M. 1989. In vivo studies on the interaction of RNA polymerase-sigma 54 with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters. The role of NifA in the formation of an open promoter complex. J Mol Biol. 210(1): 65-77.
- Morett E, Segovia L. 1993. The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. J Bacteriol. 175(19): 6067-74.
- Moris M, Dombrecht B, Xi C, Vanderleyden J, Michiels J. 2004. Regulatory role of *Rhizobium etli* CNPAF512 *fnrN* during symbiosis. Appl Environ Microbiol. 70(3): 1287-96.
- Mosse B. 1964. Electron-microscope studies of nodule development in some clover species. J. Gen. Microbiol. 36: 49-66.
- Münchbach M, Nocker A, Narberhaus F. 1999. Multiple small heat shock proteins in rhizobia. J Bacteriol. 181(1): 83-90.
- Mulley G, Lopez-Gomez M, Zhang Y, Terpolilli J, Prell J, Finan T, Poole P. 2010. Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. J Bacteriol. 192(19): 4944-53.

- Mulley G, White JP, Karunakaran R, Prell J, Bourdes A, Bunnewell S, Hill L, Poole PS. 2011. Mutation of GOGAT prevents pea bacteroid formation and N₂ fixation by globally downregulating transport of organic nitrogen sources. *Mol Microbiol.* 80(1): 149-67.
- Naggert J, Narasimhan ML, DeVeaux L, Cho H, Randhawa ZI, Cronan JE Jr, Green BN, Smith S. 1991. Cloning, sequencing, and characterization of *Escherichia coli* thioesterase II. *J Biol Chem.* 266(17): 11044-50.
- Neef A, Latorre A, Peretó J, Silva FJ, Pignatelli M, Moya A. 2011. Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities. *Genome Biol Evol.* 3: 1437-48.
- Negrea A, Bjur E, Puiac S, Ygberg SE, Aslund F, Rhen M. 2009. Thioredoxin 1 participates in the activity of the *Salmonella enterica* serovar Typhimurium pathogenicity island 2 type III secretion system. *J Bacteriol.* 191(22): 6918-27.
- Newcomb W, Callahan D, Torrey JG, Peterson RL. 1979. Morphogenesis and fine structure of the actinomycetous endophyte of nitrogen-fixing root nodules of *Comptonia peregrina*. *Bot. Gaz.* 140 (Suppl.): S22-S34.
- Nickel M, Homuth G, Böhnisch C, Mäder U, Schweder T. 2004. Cold induction of the *Bacillus subtilis bkd* operon is mediated by increased mRNA stability. *Mol Genet Genomics.* 272(1): 98-107.
- Nogales J, Campos R, BenAbdelkhalek H, Olivares J, Lluch C, Sanjuan J. 2002. *Rhizobium tropici* genes involved in free-living salt tolerance are required for the establishment of efficient nitrogen-fixing symbiosis with *Phaseolus vulgaris*. *Mol Plant Microbe Interact.* 15(3): 225-32.
- Nogales J, Domínguez-Ferreras A, Amaya-Gómez CV, van Dillewijn P, Cuéllar V, Sanjuán J, Olivares J, Soto MJ. 2010. Transcriptome profiling of a *Sinorhizobium meliloti fadD* mutant reveals the role of rhizobactin 1021 biosynthesis and regulation genes in the control of swarming. *BMC Genomics.* 11: 157.
- Nogales J, Muñoz S, Olivares J, Sanjuán J. 2009. Genetic characterization of oligopeptide uptake systems in *Sinorhizobium meliloti*. *FEMS Microbiol Lett.* 293(2): 177-87.
- Nomura M, Arunothayanan H, Dao TV, Le HTP, Kaneko T, Sato S, Tabata S, Tajima S. 2010. Differential protein profiles of *Bradyrhizobium japonicum* USDA110 bacteroid during soybean nodule development. *Soil Sci. Plant Nutr.* 56, 579–590.

- Noonan B, Motherway M, O'Gara F. 1992. Ammonia regulation of the *Rhizobium meliloti* nitrogenase structural and regulatory genes under free-living conditions: involvement of the fixL gene product? *Mol Gen Genet.* 234(3): 423-8.
- O'Connell KP, Raffel SJ, Saville BJ, Handelsman J. 1998. Mutants of *Rhizobium tropici* strain CIAT899 that do not induce chlorosis in plants. *Microbiology.* 144: 2607-17.
- Okazaki S, Hattori Y, Saeki K. 2007 The *Mesorhizobium loti purB* gene is involved in infection thread formation and nodule development in *Lotus japonicus*. *J Bacteriol.* 189(22): 8347-52.
- Okinaka Y, Yang CH, Perna NT, Keen NT. 2002. Microarray profiling of *Erwinia chrysanthemi* 3937 genes that are regulated during plant infection. *Mol Plant Microbe Interact.* 15(7): 619-29.
- Okugawa S, Moayeri M, Pomerantsev AP, Sastalla I, Crown D, Gupta PK, Leppla SH. 2012. Lipoprotein biosynthesis by prolipoprotein diacylglyceryl transferase is required for efficient spore germination and full virulence of *Bacillus anthracis*. *Mol Microbiol.* 83(1): 96-109.
- Oldroyd GE, Murray JD, Poole PS, Downie JA. 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet.* 45: 119-44.
- Oliveira LR, Marcelino FC, Barcellos FG, Rodrigues EP, Megías M, Hungria M. 2010. The *nodC*, *nodG*, and *glgX* genes of *Rhizobium tropici* strain PRF 81. *Funct Integr Genomics.* 10(3): 425-31.
- Oono R, Denison RF. 2010. Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* 154(3): 1541-8.
- Ortuño-Olea L, Durán-Vargas S. 2000. The L-asparagine operon of *Rhizobium etli* contains a gene encoding an atypical asparaginase. *FEMS Microbiol Lett.* 189(2): 177-82.
- Ozga DA, Lara JC, Leig JA. 1994. The regulation of exopolysaccharide production is important at two levels of nodule development in *Rhizobium meliloti*. *Mol Plant Microbe Interact.* 7(6): 758-65.
- Paaü AS, Bloch CB, Brill WJ. 1980. Developmental fate of *Rhizobium meliloti* bacteroids in alfalfa nodules. *J Bacteriol.* 143(3): 1480-90.
- Paget MS, Helmann JD. 2003. The sigma70 family of sigma factors. *Genome Biol.* 4(1): 203.
- Pallen M. 1999. RpoN-dependent transcription of *rpoH*? *Mol Microbiol.* 31(1): 393.

- Park SJ, Gunsalus RP. 1995. Oxygen, iron, carbon, and superoxide control of the fumarase *fumA* and *fumC* genes of *Escherichia coli*: role of the *arcA*, *fnr*, and *soxR* gene products. *J Bacteriol.* 177(21): 6255-62.
- Perret X, Staehelin C, Broughton WJ. 2000. Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev.* 64(1): 180-201.
- Pessi G, Ahrens CH, Rehrauer H, Lindemann A, Hauser F, Fischer HM, Hennecke H. 2007. Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Mol Plant Microbe Interact.* 20(11): 1353-63.
- Pino C, Olmo-Mira F, Cabello P, Martínez-Luque M, Castillo F, Roldán MD, Moreno-Vivián C. 2006. The assimilatory nitrate reduction system of the phototrophic bacterium *Rhodobacter capsulatus* E1F1. *Biochem Soc Trans.* 34(Pt 1): 127-9.
- Pinto UM, Winans SC. 2009. Dimerization of the quorum-sensing transcription factor TraR enhances resistance to cytoplasmic proteolysis. *Mol Microbiol.* 73(1): 32-42.
- Planamente S, Vigouroux A, Mondy S, Nicaise M, Faure D, Moréra S. 2010. A conserved mechanism of GABA binding and antagonism is revealed by structure-function analysis of the periplasmic binding protein Atu2422 in *Agrobacterium tumefaciens*. *J Biol Chem.* 285(39): 30294-303.
- Pobigaylo N, Szymczak S, Nattkemper TW, Becker A. 2008. Identification of genes relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants. *Mol Plant Microbe Interact.* 21(2): 219-31.
- Polesky AH, Ross JT, Falkow S, Tompkins LS. 2001. Identification of *Legionella pneumophila* genes important for infection of amoebas by signature-tagged mutagenesis. *Infect Immun* 69: 977-987.
- Popp C, Ott T. 2011. Regulation of signal transduction and bacterial infection during root nodule symbiosis. *Curr Opin Plant Biol.* 14(4):458-67.
- Prell J, Boesten B, Poole P, Priefer UB. 2002. The *Rhizobium leguminosarum* bv. viciae VF39 gamma-aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids. *Microbiology.* 148(Pt 2): 615-23.
- Prell J, Bourdès A, Karunakaran R, Lopez-Gomez M, Poole P. 2009. Pathway of gamma-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. *J Bacteriol.* 191(7): 2177-86.

- Prell J, Bourdès A, Kumar S, Lodwig E, Hosie A, Kinghorn S, White J, Poole P. 2010. Role of symbiotic auxotrophy in the *Rhizobium*-legume symbioses. PLoS One. 5(11): e13933.
- Prell J, Mulley G, Haufe F, White JP, Williams A, Karunakaran R, Downie JA, Poole PS. 2012. The PTS(Ntr) system globally regulates ATP-dependent transporters in *Rhizobium leguminosarum*. Mol Microbiol. 84(1): 117-29.
- Prell J, Poole P. 2006. Metabolic changes of rhizobia in legume nodules. Trends Microbiol. 14(4): 161-8.
- Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS. 2009. Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. Proc Natl Acad Sci U S A. 106(30): 12477-82.
- Puppo A, Groten K, Bastian F, Carzaniga R, Soussi M, Lucas MM, de Felipe MR, Harrison J, Vanacker H, Foyer CH. 2005. Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. New Phytol. 165(3): 683-701.
- Qian D, Zhou D, Ju R, Cramer CL, Yang Z. 1996. Protein farnesyltransferase in plants: molecular characterization and involvement in cell cycle control. Plant Cell. 8(12): 2381-2394.
- Ramírez-Romero MA, Masulis I, Cevallos MA, González V, Dávila G. The *Rhizobium etli* sigma70 (SigA) factor recognizes a lax consensus promoter. 2006. Nucleic Acids Res. 34(5): 1470-80.
- Ramírez-Trujillo JA, Encarnación S, Salazar E, de los Santos AG, Dunn MF, Emerich DW, Calva E, Hernández-Lucas I. 2007. Functional characterization of the *Sinorhizobium meliloti* acetate metabolism genes *aceA*, SMC00767, and *glcB*. J Bacteriol. 189(16): 5875-84.
- Ramseier TM, Winteler HV, Hennecke H. 1991. Discovery and sequence analysis of bacterial genes involved in the biogenesis of c-type cytochromes. J Biol Chem. 266(12): 7793-803.
- Rashid MH, Kornberg A. 2000. Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A. 97(9): 4885-90.
- Rathbun KM, Hall JE, Thompson SA. 2009. Cj0596 is a periplasmic peptidyl prolyl cis-trans isomerase involved in *Campylobacter jejuni* motility, invasion, and colonization. BMC Microbiol. 9: 160.

Real G, Henriques AO. 2006. Localization of the *Bacillus subtilis murB* gene within the *dcw* cluster is important for growth and sporulation. *J Bacteriol.* 188(5): 1721-32.

Reeve WG, Dilworth MJ, Tiwari RP, Glenn AR. 1997. Regulation of exopolysaccharide production in *Rhizobium leguminosarum* biovar *viciae* WSM710 involves *exoR*. *Microbiology.* 143 (6): 1951-8.

Reeve WG, Tiwari RP, Guerreiro N, Stubbs J, Dilworth MJ, Glenn AR, Rolfe BG, Djordjevic MA, Howieson JG. 2004. Probing for pH-regulated proteins in *Sinorhizobium medicae* using proteomic analysis. *J Mol Microbiol Biotechnol.* 7(3): 140-7.

Reitzer LJ, Magasanik B. 1986. Transcription of *glnA* in *Escherichia coli* is stimulated by activator bound to sites far from the promoter. *Cell.* 45(6): 785-92.

Resendis-Antonio O, Hernández M, Mora Y, Encarnación S. 2012. Functional modules, structural topology, and optimal activity in metabolic networks. *PLoS Comput Biol.* 8(10): e1002720.

Riccillo PM, Collavino MM, Grasso DH, England R, de Bruijn FJ, Aguilar OM. 2000. A *guaB* mutant strain of *Rhizobium tropici* CIAT899 pleiotropically defective in thermal tolerance and symbiosis. *Mol Plant Microbe Interact.* 13(11): 1228-36.

Ritz D, Thöny-Meyer L, Hennecke H. 1995. The *cycHJKL* gene cluster plays an essential role in the biogenesis of c-type cytochromes in *Bradyrhizobium japonicum*. *Mol Gen Genet.* 247(1): 27-38.

Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE, Felbeck H. 2008. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ Microbiol.* 10(3): 727-37.

Rodionov DA. 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev.* 107(8): 3467-97.

Rolfe BG, Gresshoff PM. 1988. Genetic analysis of legume nodule initiation. *Annu. Rev. Plant Physiol Plant Mol. Biol.* 39: 297-319.

Roy AM, Coleman J. 1994. Mutations in *firA*, encoding the second acyltransferase in lipopolysaccharide biosynthesis, affect multiple steps in lipopolysaccharide biosynthesis. *J Bacteriol.* 176(6): 1639-46.

Rudnick PA, Arcondéguy T, Kennedy CK, Kahn D. 2001. *glnD* and *mviN* are genes of an essential operon in *Sinorhizobium meliloti*. *J Bacteriol.* 183(8): 2682-5.

- Sandmann G, Misawa N. 1992. New functional assignment of the carotenogenic genes *crtB* and *crtE* with constructs of these genes from *Erwinia* species. FEMS Microbiol Lett. 69(3): 253-7.
- Santos R, Hérouart D, Puppo A, Touati D. 2000. Critical protective role of bacterial superoxide dismutase in rhizobium-legume symbiosis. Mol Microbiol. 38(4): 750-9.
- Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 270(5235): 467-70.
- Schinko T, Berger H, Lee W, Gallmetzer A, Pirker K, Pachlinger R, Buchner I, Reichenauer T, Güldener U, Strauss J. 2010. Transcriptome analysis of nitrate assimilation in *Aspergillus nidulans* reveals connections to nitric oxide metabolism. Mol Microbiol. 78(3): 720-38.
- Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quéré A, Wollherr A, Heinemeyer I, Morgenstern B, Pommerening-Röser A, Flores M, Palacios R, Brenner S, Gottschalk G, Schmitz RA, Broughton WJ, Perret X, Strittmatter AW, Streit WR. 2009. *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. Appl Environ Microbiol. 75(12): 4035-45.
- Schumacher MA, Piro KM, Xu W, Hansen S, Lewis K, Brennan RG. 2009. Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. Science. 323(5912): 396-401.
- Scott DB, Wilson R, Shaw GJ, Petit A, and Tempe J. 1987. Biosynthesis and degradation of nodule-specific *Rhizobium loti* compounds in *Lotus* nodules. J Bacteriol. 169(1): 278–282.
- Scott JD, Ludwig RA. 2004. *Azorhizobium caulinodans* electron-transferring flavoprotein N electrochemically couples pyruvate dehydrogenase complex activity to N₂ fixation. Microbiology. 150(Pt 1): 117-26.
- Shin DH, Choi IG, Busso D, Jancarik J, Yokota H, Kim R, Kim SH. 2004. Structure of OsmC from *Escherichia coli*: a salt-shock-induced protein. Acta Crystallogr D Biol Crystallogr. 60(Pt 5): 903-11.
- Sieira R, Arocena GM, Bukata L, Comerci DJ, Ugalde RA. 2010. Metabolic control of virulence genes in *Brucella abortus*: HutC coordinates *virB* expression and the histidine utilization pathway by direct binding to both promoters. J Bacteriol. 192(1): 217-24.
- Skorupska A, Janczarek M, Marczak M, Mazur A, Król J. 2006. Rhizobial exopolysaccharides: genetic control and symbiotic functions. Microb Cell Fact. 5: 7.

Smith AH, Blevins JS, Bachlani GN, Yang XF, Norgard MV. 2007. Evidence that RpoS (σ^S) in *Borrelia burgdorferi* is controlled directly by RpoN (σ^{54}/σ^N). J Bacteriol. 189(5): 2139-44.

Soberón M, Lopez O, Miranda J, Tabche ML, Morera C. 1997. Genetic evidence for 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) as a negative effector of cytochrome terminal oxidase *cbb3* production in *Rhizobium etli*. Mol Gen Genet. 254(6): 665-73.

Soberón M, Morera C, Kondorosi A, Lopez O, Miranda J. 2001. A purine-related metabolite negatively regulates *fixNOQP* expression in *Sinorhizobium meliloti* by modulation of *fixK* expression. Mol Plant Microbe Interact. 14(4): 572-6.

Song J, Wurtele ES, Nikolau BJ. 1994. Molecular cloning and characterization of the cDNA coding for the biotin-containing subunit of 3-methylcrotonoyl-CoA carboxylase: identification of the biotin carboxylase and biotin-carrier domains. Proc Natl Acad Sci U S A. 91(13): 5779-83.

Soto MJ, Fernández-Pascual M, Sanjuan J, Olivares J. 2002. A *fadD* mutant of *Sinorhizobium meliloti* shows multicellular swarming migration and is impaired in nodulation efficiency on alfalfa roots. Mol Microbiol. 43(2): 371-82.

Soto MJ, Sanjuan J, Olivares J. 2001. The disruption of a gene encoding a putative arylesterase impairs pyruvate dehydrogenase complex activity and nitrogen fixation in *Sinorhizobium meliloti*. Mol Plant Microbe Interact. 14(6): 811-5.

Stahelin C, Forsberg LS, D'Haeze W, Gao MY, Carlson RW, Xie ZP, Pellock BJ, Jones KM, Walker GC, Streit WR, Broughton WJ. 2006. Exo-oligosaccharides of *Rhizobium* sp. strain NGR234 are required for symbiosis with various legumes. J Bacteriol. 188(17): 6168-78.

Stevens JB, Carter RA, Hussain H, Carson KC, Dilworth MJ, Johnston AW. 1999. The *fhu* genes of *Rhizobium leguminosarum*, specifying siderophore uptake proteins: *fhuDCB* are adjacent to a pseudogene version of *fhuA*. Microbiology. 145 (Pt 3): 593-601.

Suárez R, Wong A, Ramírez M, Barraza A, Orozco Mdel C, Cevallos MA, Lara M, Hernández G, Iturriaga G. 2008. Improvement of drought tolerance and grain yield in common bean by overexpressingtrehalose-6-phosphate synthase in rhizobia. Mol Plant Microbe Interact. 21(7): 958-66.

Subramanian A, et al. 2005. Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545-15550.

- Sugawara M, Cytryn EJ, Sadowsky MJ. 2010. Functional role of *Bradyrhizobium japonicum* trehalose biosynthesis and metabolism genes during physiological stress and nodulation. *Appl Environ Microbiol.* 76(4): 1071-81.
- Suzuki S, Aono T, Lee KB, Suzuki T, Liu CT, Miwa H, Wakao S, Iki T, Oyaizu H. 2007. Rhizobial factors required for stem nodule maturation and maintenance in *Sesbania rostrata*-*Azorhizobium caulinodans* ORS571 symbiosis. *Appl Environ Microbiol.* 73(20): 6650-9.
- Szeto WW, Nixon BT, Ronson CW, Ausubel FM. 1987. Identification and characterization of the *Rhizobium meliloti ntrC* gene: *R. meliloti* has separate regulatory pathways for activation of nitrogen genes in free-living and symbiotic cells. *J Bacteriol.* 169: 1423-1432.
- Tabche ML, García EG, Miranda J, Escamilla JE, Soberón M. 1998. *Rhizobium etli cycHJKL* gene locus involved in c-type cytochrome biogenesis: sequence analysis and characterization of two *cycH* mutants. *Gene.* 208(2): 215-9.
- Taga ME, Walker GC. 2010. *Sinorhizobium meliloti* requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. *Mol Plant Microbe Interact.* 23(12): 1643-54.
- Takanashi K, Takahashi H, Sakurai N, Sugiyama A, Suzuki H, Shibata D, Nakazono M, Yazaki K. 2012. Tissue-specific transcriptome analysis in nodules of *Lotus japonicus*. *Mol Plant Microbe Interact.* 25(7): 869-76.
- Tambalo DD, Bustard DE, Del Bel KL, Koval SF, Khan MF, Hynes MF. 2010. Characterization and functional analysis of seven flagellin genes in *Rhizobium leguminosarum* bv. *viciae*. Characterization of *R. leguminosarum* flagellins. *BMC Microbiol.* 10: 219.
- Tatè R, Ferraioli S, Filosa S, Cermola M, Riccio A, Iaccarino M, Patriarca EJ. 2004. Glutamine utilization by *Rhizobium etli*. *Mol Plant Microbe Interact.* 17(7): 720-8.
- Tatè R, Mandrich L, Spinosa MR, Riccio A, Lamberti A, Iaccarino M, Patriarca EJ. 2001. The *Rhizobium* GstI protein reduces the NH₄⁺ assimilation capacity of *Rhizobium leguminosarum*. *Mol Plant Microbe Interact.* 14(7): 823-31.
- Téllez-Sosa J, Soberón N, Vega-Segura A, Torres-Márquez ME, Cevallos MA. 2002. The *Rhizobium etli cyaC* product: characterization of a novel adenylate cyclase class. *J Bacteriol.* 184(13): 3560-8.
- Terpolilli JJ, Hood GA, Poole PS. 2012. What determines the efficiency of N(2)-fixing *Rhizobium*-legume symbioses? *Adv Microb Physiol.* 60: 325-89.

- Thies JE, Bohlool BB, Singleton PW. 1991. Subgroups of the Cowpea Miscellany: Symbiotic Specificity within *Bradyrhizobium* spp. for *Vigna unguiculata*, *Phaseolus lunatus*, *Arachis hypogaea*, and *Macroptilium atropurpureum*. *Appl Environ Microbiol.* 57(5): 1540-5.
- Thöny B, Hennecke H. 1989. The -24/-12 promoter comes of age. *FEMS Microbiol Rev.* 5(4): 341-57.
- Thöny-Meyer L, Künzler P. 1996. The *Bradyrhizobium japonicum* aconitase gene (*acnA*) is important for free-living growth but not for an effective root nodule symbiosis. *J Bacteriol.* 178(21): 6166-72.
- Thornton, H. G. 1930. The influence of the host plant in inducing parasitism in lucerne and clover nodules. *Proceedings of the Royal Society of London B, Biological Sciences* 106: 110–122.
- Tian CF, Garnerone AM, Mathieu-Demazière C, Masson-Boivin C, Batut J. 2012. Plant-activated bacterial receptor adenylate cyclases modulate epidermal infection in the *Sinorhizobium meliloti-Medicago* symbiosis. *Proc Natl Acad Sci U S A.* 2012 109(17): 6751-6.
- Timmers AC, Auriac MC, Truchet G. 1999. Refined analysis of early symbiotic steps of the *Rhizobium-Medicago* interaction in relationship with microtubular cytoskeleton rearrangements. *Development.* 126(16): 3617-28
- Timmers AC, Soupène E, Auriac MC, de Billy F, Vasse J, Boistard P, Truchet G. 2000. Saprophytic intracellular rhizobia in alfalfa nodules. *Mol Plant Microbe Interact.* 13(11): 1204-13.
- Tjepkema, JD. 1983. Oxygen Concentration Within the Nitrogen-Fixing Root Nodules of *Myrica gale* L. *Amer. J. Bot.* 70(1): 59-63.
- Tomlinson AD, Ramey-Hartung B, Day TW, Merritt PM, Fuqua C. 2010. Microbiology. *Agrobacterium tumefaciens* ExoR represses succinoglycan biosynthesis and is required for biofilm formation and motility. 156(Pt 9): 2670-81.
- Trémillon N, Morello E, Llull D, Mazmouz R, Gratadoux JJ, Guillot A, Chapot-Chartier MP, Monlezun L, Solé V, Ginisty H, Poquet I. 2012. PpiA, a surface PPIase of the cyclophilin family in *Lactococcus lactis*. *PLoS One.* 7(3): e33516.
- Tully RE, Keister DL. 1993. Cloning and Mutagenesis of a Cytochrome P-450 Locus from *Bradyrhizobium japonicum* That Is Expressed Anaerobically and Symbiotically. *Appl Environ Microbiol.* 59(12): 4136-42.

Tzfira T, Vaidya M, Citovsky V. 2004. Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. *Nature*. 431(7004): 87-92.

Uchiumi T, Ohwada T, Itakura M, Mitsui H, Nukui N, Dawadi P, Kaneko T, Tabata S, Yokoyama T, Tejima K, Saeki K, Omori H, Hayashi M, Maekawa T, Sriprang R, Murooka Y, Tajima S, Simomura K, Nomura M, Suzuki A, Shimoda Y, Sioya K, Abe M, Minamisawa K. J. 2004. Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *J Bacteriol*. 186(8): 2439-48.

Ullrich J, van Putten JP. 1995. Identification of the gonococcal *glmU* gene encoding the enzyme N-acetylglucosamine 1-phosphate uridyltransferase involved in the synthesis of UDP-GlcNAc. *J Bacteriol*. 177(23): 6902-9.

Valderrama B, Dávalos A, Girard L, Morett E, Mora J. 1996. Regulatory proteins and cis-acting elements involved in the transcriptional control of *Rhizobium etli* reiterated *nifH* genes. *J Bacteriol*. 178(11): 3119-26.

van Helden, J. 2003. Regulatory sequence analysis tools. *Nucleic Acids Res*. 31: 3593–3596.

Vance CP, Johnson LEB, Stade S, Groat RG. 1982. Birdsfoot trefoil (*Lotus corniculatus*) root nodules: morphogenesis and the effect of forage harvest on structure and function. *Canadian Journal of Botany*. 60(4): 505-518, 10.1139/b82-068.

Vance CP, Johnson LEB, Halvorsen AM, Heichel GH, Barnes DK. 1980. Histological and ultrastructural observations of *Medicago sativa* root nodule senescence after foliage removal. *Canadian Journal of Botany* 58: 295–309.

Vargas Mdel C, Encarnación S, Dávalos A, Reyes-Pérez A, Mora Y, García-de los Santos A, Brom S, Mora J. 2003. Only one catalase, *katG*, is detectable in *Rhizobium etli*, and is encoded along with the regulator OxyR on a plasmid replicon. *Microbiology*. 149(Pt 5): 1165-76.

Vasse J, de Billy F, Camut S, Truchet G. 1990. Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol*. 172(8): 4295-306.

Vercruyssen M, Fauvart M, Beullens S, Braeken K, Cloots L, Engelen K, Marchal K, Michiels J. 2011. A comparative transcriptome analysis of *Rhizobium etli* bacteroids: specific gene expression during symbiotic nongrowth. *Mol Plant Microbe Interact*. 24(12): 1553-61.

Vercruyssen M, Fauvart M, Cloots L, Engelen K, Thijs IM, Marchal K, Michiels J. 2010. Genome-wide detection of predicted non-coding RNAs in *Rhizobium etli* expressed during

free-living and host-associated growth using a high-resolution tiling array. *BMC Genomics*. 11: 53.

Vercruyse M, Fauvart M, Jans A, Beullens S, Braeken K, Cloots L, Engelen K, Marchal K, Michiels J. 2011. Stress response regulators identified through genome-wide transcriptome analysis of the (p)ppGpp-dependent response in *Rhizobium etli*. *Genome Biol*. 12(2): R17.

Villaseñor T, Brom S, Dávalos A, Lozano L, Romero D, Los Santos AG. 2011. Housekeeping genes essential for pantothenate biosynthesis are plasmid-encoded in *Rhizobium etli* and *Rhizobium leguminosarum*. *BMC Microbiol*. 11: 66.

Vineetha KE, Vij N, Prasad CK, Hassani R, Randhawa GS. 2001. Ultrastructural studies on nodules induced by pyrimidine auxotrophs of *Sinorhizobium meliloti*. *Indian J Exp Biol*. 39(4): 371-7.

Vizcaíno N, Cloeckert A, Zygmunt MS, Dubray G. 1996. Cloning, nucleotide sequence, and expression of the *Brucella melitensis omp31* gene coding for an immunogenic major outer membrane protein. *Infect Immun*. 64(9): 3744-51.

Voysest, O. 2000. Mejoramiento genético del frijol (*Phaseolus vulgaris* L.): Legado de variedades de América Latina 1930-1999. Edi. CIAT, pp 191.

Waelkens F, Voets T, Vlassak K, Vanderleyden J, van Rhijn P. 1995. The *nodS* gene of *Rhizobium tropici* strain CIAT899 is necessary for nodulation on *Phaseolus vulgaris* and on *Leucaena leucocephala*. *Mol Plant Microbe Interact*. 8(1): 147-54.

Walshaw DL, Wilkinson A, Mundy M, Smith M, Poole PS. 1997. Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in *Rhizobium leguminosarum*. *Microbiology*. 143 (Pt 7): 2209-21.

Wang L, Hames C, Schmidl SR, Stülke J. 2010. Upregulation of thymidine kinase activity compensates for loss of thymidylate synthase activity in *Mycoplasma pneumoniae*. *Mol Microbiol*. 77(6): 1502-11.

Wang Y, Chen AM, Yu AY, Luo L, Yu GQ, Zhu JB, Wang YZ. 2008. The GntR-type regulators *gtrA* and *gtrB* affect cell growth and nodulation of *Sinorhizobium meliloti*. *J Microbiol*. 46(2): 137-45.

Waters JK, Hughes BL 2nd, Purcell LC, Gerhardt KO, Mawhinney TP, Emerich DW. 1998. Alanine, not ammonia, is excreted from N₂-fixing soybean nodule bacteroids. *Proc Natl Acad Sci U S A*. 95(20):12038-42.

- Weissenmayer B, Geiger O, Benning C. 2000. Disruption of a gene essential for sulfoquinovosyldiacylglycerol biosynthesis in *Sinorhizobium meliloti* has no detectable effect on root nodule symbiosis. *Mol Plant Microbe Interact.* 13(6): 666-72.
- Westenberg DJ, Guerinot ML. 1999. Succinate dehydrogenase (Sdh) from *Bradyrhizobium japonicum* is closely related to mitochondrial Sdh. *J Bacteriol.* 181(15): 4676-9.
- Wigneshweraraj S, Bose D, Burrows PC, Joly N, Schumacher J, Rappas M, Pape T, Zhang X, Stockley P, Severinov K, Buck M. 2008. Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Mol Microbiol.* 68(3): 538-46.
- Wilkinson A, Danino V, Wisniewski-Dyé F, Lithgow JK, Downie JA. 2002. N-acyl-homoserine lactone inhibition of rhizobial growth is mediated by two quorum-sensing genes that regulate plasmid transfer. *J Bacteriol.* 184(16): 4510-9.
- Willis LB, Walker GC. 1999. A novel *Sinorhizobium meliloti* operon encodes an alpha-glucosidase and a periplasmic-binding-protein-dependent transport system for alpha-glucosides. *J Bacteriol.* 181(14): 4176-84.
- Wilson AC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, Thomas GH, Douglas AE. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol.* 19 Suppl 2: 249-58.
- Wösten MM. 1998. Eubacterial sigma-factors. *FEMS Microbiol Rev.* 22(3): 127-50.
- Xi C, Schoeters E, Vanderleyden J, Michiels J. 2000. Symbiosis-specific expression of *Rhizobium etli casA* encoding a secreted calmodulin-related protein. *Proc Natl Acad Sci U S A.* 97(20): 11114-9.
- Xie F, Cheng G, Xu H, Wang Z, Lei L, Li Y. 2011. Identification of a novel gene for biosynthesis of a bacteroid-specific electron carrier menaquinone. *PLoS One.* 6(12): e28995.
- Xu T, Ping J, Yu Y, Yu F, Yu Y, Hao P, Li X. 2010. Revealing parasite influence in metabolic pathways in *Apicomplexa* infected patients. *BMC Bioinformatics.* 11 Suppl 11: S13.
- Xu Y, Mortimer MW, Fisher TS, Kahn ML, Brockman FJ, Xun L. 1997. Cloning, sequencing, and analysis of a gene cluster from *Chelatobacter heintzii* ATCC 29600 encoding nitrilotriacetate monooxygenase and NADH: flavin mononucleotide oxidoreductase. *J Bacteriol.* 179(4): 1112-6.

- Yang K, Wang M, Metcalf WW. 2009. Uptake of glycerol-2-phosphate via the *ugp*-encoded transporter in *Escherichia coli* K-12. *J Bacteriol.* 191(14): 4667-70.
- Yao SY, Luo L, Har KJ, Becker A, Rüberg S, Yu GQ, Zhu JB, Cheng HP. 2004. *Sinorhizobium meliloti* ExoR and ExoS proteins regulate both succinoglycan and flagellum production. *J Bacteriol.* 186(18): 6042-9.
- Yost CK, Rochepeau P, Hynes MF. 1998. *Rhizobium leguminosarum* contains a group of genes that appear to code for methyl-accepting chemotaxis proteins. *Microbiology.* 144 (Pt 7): 1945-56.
- Yurgel SN, Berrocal J, Wilson C, Kahn ML. 2007. Pleiotropic effects of mutations that alter the *Sinorhizobium meliloti* cytochrome c respiratory system. *Microbiology.* 153: 399-410.
- Yurgel SN, Kahn ML. 2004. Dicarboxylate transport by rhizobia. *FEMS Microbiol Rev.* 28(4): 489-501.
- Zeller T, Moskvina OV, Li K, Klug G, Gomelsky M. 2005. Transcriptome and physiological responses to hydrogen peroxide of the facultatively phototrophic bacterium *Rhodobacter sphaeroides*. *J Bacteriol.* 187(21): 7232-42.
- Zhu L, Cheng J, Luo B, Feng S, Lin J, Wang S, Cronan JE, Wang H. 2009. Functions of the *Clostridium acetobutylicum* FabF and FabZ proteins in unsaturated fatty acid biosynthesis. *BMC Microbiol.* 9: 119.
- Zientz E, Beyaert I, Gross R, Feldhaar H. 2006. Relevance of the endosymbiosis of *Blochmannia floridanus* and carpenter ants at different stages of the life cycle of the host. *Appl Environ Microbiol.* 72(9): 6027-33.
- Zorreguieta A, Finnie C, Downie JA. 2000. Extracellular glycanases of *Rhizobium leguminosarum* are activated on the cell surface by an exopolysaccharide-related component. *J Bacteriol.* 182(5): 1304-12.