VNIVERSIDAD NACIONAL
AVTONOMA DE
MEXICO

Universidad Nacional Autónoma de México

Programa de Doctorado en Ciencias Biomédicas

Sustitución de ortólogos *argC* en *Sinorhizobium meliloti*: función y expresión

Tesis que para optar por el grado de

Doctor en Ciencias

presenta:

M. en C. Rafael Díaz Méndez

Director de tesis**:**
Dr. Jaime Mora Célis
Centro de Ciencias Genómicas

Comité tutoral:
Gloria Soberón Chávez
Instituto de Investigaciones Biomédicas, UNAM

Sergio Encarnación Guevara
Centro de Ciencias Genómicas, UNAM

Cuernavaca, Mor., Noviembre del 2015.

Resumen

Por definición, los genes ortólogos son aquellos genes que se encuentran presentes en un organismo ancestral común antes de la especiación, desempeñando generalmente la misma función y constituyendo el grupo de genes más conservados del genoma bacteriano. Sin embargo, al comparar las secuencias de los genes ortologós de organismos cercanos, existe una proporción de las proteínas codificadas por dichos genes que presentan residuos específicos en cada especie (firma de la especie, FDE), lo que podría ser resultado de la adaptación del organismo a un microambiente específico (Guerrero *et al.* 2005).

El objetivo principal del presente trabajo fue evaluar el significado funcional de la FDE del ortólogo esencial *argC*, que participa en la vía de síntesis de la arginina en Rhizobiales.

El estudio se realizó mediante la sustitución del gen ortólogo *argC en Sinorhizobium meliloti* y los genes ortólogos sustitutos provinieron de organismos cercanos del grupo Rhizobiaceae (*Rhizobium etli, Agrobacterium tumefaciens, Mesorhizobium loti* y el propio de *Sinorhizobium meliloti*).

Las secuencias codificadoras (*orf*) de los genes ortólogos *argC* mencionados fueron clonados en plásmidos y fusionados a tres promotores para su expresión regulada baja, media y constitiutiva: *pspeB (S. meliloti)*, *pargC* (*S. meliloti*) y *plac* de *E. coli* (constitutivo). Posteriormente fueron introducidos en la cepa mutante de *S. meliloti argC*$^{-}$ *y E. coli*. Cada complementación fue analizada mediante: crecimiento, expresión transcripcional, actividad enzimática de ArgC (*N*-acetil-glutamil-fosfato reductasa), niveles de RNA mensajero (mRNA), detección proteomica de la proteína ArgC y eficiencia traduccional (a partir de la cuantificación del mensajero y de la proteína producida).

Entre los resultados obtenidos destacaron los siguientes:

•Los ortólogos *argC* se desempeñaron de una forma diferencial para cada complementación, reflejando que hay diversos factores que influyen en la expresión de los genes y en la capacidad del producto ortólogo para funcionar en un fondo metabólico extraño.

•La complementación óptima fue directamente relacionada a la similitud de la secuencia respecto a la de *S. meliloti* y fue inversamente relacionada a la cantidad de cambios propios de la especie (un parámetro que denominamos la firma de la especie). El gen *argC* de *M. loti* mostró el desempeño más pobre, seguido por el de *R. etli* y *A. tumefaciens*.

•Se encontró que se produjeron diferentes números de copias de los plásmidos en que fueron clonados los genes ortólogos, y en la concentración del mRNA y de la proteína correspondiente. Esto sugiere la presencia de mecanismos celulares de regulación para optimizar la actividad enzimática, la función y concentración del producto codificado.

Los resultados sugieren que la sustitución de genes ortólogos aún en especies cercanamente relacionadas no es funcionalmente equivalente. La expresión de un gen ortólogo esencial sustituido en el fondo de una especie cercana tiene implicaciones funcionales y evolutivas diversas, que se relacionan con el cambio en las secuencias de genes conservados, y su efecto en la transcripción, traducción y actividad del producto codificado, lo que refleja mecanismos celulares de respuesta ante la expresión de un gen con adaptaciones que corresponden a un microambiente celular especie-específico.

Introducción.

En el estudio de los organismos y su evolución, una herramienta fundamental ha sido el establecimiento de una clasificación natural de las especies para establecer las relaciones filogenéticas entre éstas. Los caracteres o atributos evolutivos pueden ser desde las moléculas hasta los ecosistemas. Una herramienta valiosa ha sido la comparación de las secuencias codificadoras de los genes y la determinación de los cambios provenientes de la adaptación evolutiva.

La adaptación puede estimarse al cuantificar la velocidad de evolución mediante la relación de cambios no sinónimos a sinónimos (dN/dS) (Yang 2000) en el genoma de especies cercanas, por la tendencia (*bias)* en el uso de codones o por la diferencia en el contenido de GC en la tercera posición de los codones (%GC3)(Elhaik 2009). Sin embargo, la medida más clara (y al mismo tiempo más compleja) de la adaptación evolutiva, es el cambio funcional, ya que no todas las mutaciones producen cambios cuantificables y los que efectivamente afectan el desempeño enzimático pudieron haber sido seleccionadas positivamente.

Por otro lado, la variación fenotípica y genotípica presente en las poblaciones no puede ser atribuida únicamente a la selección natural (selección adaptativa o positiva). A nivel molecular existe una gran variación genética contenida en las proteínas y genes, que no puede ser justificada únicamente por efectos de la selección natural. Esta observación culminó con la propuesta de la teoría neutra de la evolución (Kimura 1968, Kimura 1984). De acuerdo con ésta, la mayor parte de la variación al nivel molecular no es mantenida por la selección tipo Darwiniana (selección adaptativa o positiva), sino que es el resultado de un balance entre la deriva génica y las mutaciones neutras o casi neutras (Kimura 1984, Ohta 1992). Es decir, esta teoría propone que la mayor parte de la variación a nivel molecular ocurre de manera aleatoria y no tiene importancia adaptativa.

Homología, ortología y paralogía.

En el estudio de organismos y sus genomas, una herramienta ampliamente utilizada, la constituyen los genes, los cuales son considerados caracteres moleculares susceptibles de ser comparados. Uno de los primeros conceptos utilizados en esta ciencia descriptiva para referirnos a los mismos caracteres en diferentes especies, fue el término de *homología*. La definición original de homología de Owen (Owen 1848) es:

Homólogo, el mismo órgano en diferentes animales bajo una variedad de formas y funciones…

La definición de Owen se acuñó antes de la aparición del concepto de la evolución. Actualmente la definición de Ax es la utilizada por ser más explícita, ya que implementa el concepto evolutivo del cambio serial de un caracter :

"Caracteres homólogos son caracteres en dos o más especies que retrocediendo hasta la especie troncal son el mismo caracter. Pueden haber surgido de la especie troncal sin cambios o con transformaciones evolutivas" (Ax 1987).

El concepto de homología es fundamental y constituye la base para hacer inferencias de procesos evolutivos tales como especiación, duplicación genética y tranferencia genética horizontal (HGT). A principios de los 70´s, Walter Fitch dividió el término homología en ortología y paralogía de acuerdo a dos distintos procesos evolutivos: especiación y duplicación genética (Fitch 1970, Fitch 2000; Gabaldón & Koonin 2013). De esta forma, los genes ortólogos son homólogos que se relacionan a través de un evento de especiación de un gen único ancestral presente en su último ancestro común, mientras que los genes parálogos fueron definidos como homólogos que surgen de un evento de duplicación genética. Por otra parte también existen genes

homologos relacionados por eventos de transferencia horizontal, los xenólogos (Koonin *et al.* 2001).

El significado de homología, ortología y paralogía pueden ser modificados por otros eventos evolutivos importantes tales como la transferencia horizontal de material genético (HGT) y los eventos de fusion, mismos que permitieron la construcción de redes filogenéticas (Pavlinov 2012, McCune & Schimenti 2012). Los terminos de ortología y paralogía han sido ampliamente utilizados desde los 90´s cuando se originaron las primeras comparaciones genéticas a nivel nucleotídico. A partir de entonces, la identificación de ortólogos ha sido indispensable para estudiar la evolución de las especies y sus genomas.

La ortología es uno de los conceptos más precisos para describir y comparar diferencias y similitudes genéticas entre diferentes especies, debido a que los ortólogos, conceptualmente, provienen de un gen ancestral común a las especies comparadas. La aplicación del concepto de ortología en la anotación genética ha permitido identificar ortólogos y asignarles función en organismos no caracterizados.

La transferencia o extrapolación de anotación genética funcional es un rasgo biológico importante que permite transferir la información experimental de genes de un organismo modelo a genes de un organismo no caracterizado o recién secuenciado (Storm & Sonnhammer 2002; Koonin 2005, Dolinski & Botstein 2007). La premisa que permite transferir la anotación funcional conocida a organismos no caracterizados se conoce como el modelo de suposición ortóloga (*ortholog conjecture*, OC), que establece que los genes ortólogos realizan funciones equivalentes en diferentes organismos (Koonin 2005, Dolinski & Botstein 2007, Studer & Robinson-Rechavi 2009, Altenhoff *et al.* 2012). Este modelo, de amplia aceptación, ha sido recientemente cuestionado por trabajos que indican que los genes parálogos dentro de la misma especie tienen funciones equivalentes más cercanas que genes ortólogos de diferentes organismos, filogenéticamente muy relacionados (Nehrt *et al.* 2011).

La duplicación y la Transferencia Horizontal (HGT) como fuente de innovación genética.

La duplicación genética es la principal fuente evolutiva para la aparición de nuevos genes y funciones (Hughes 1994, Force *et al*. 1999, Innan & Kondrashov 2010, Ohno 1970). Se han propuesto las diversas rutas que siguen estos genes duplicados (nuevos parálogos):

a) *Modelo de no funcionalización.* De forma simple, la duplicación produce dos parálogos los cuales son funcionalmente redundantes, por consiguiente, uno de ellos estará libre de restricciones selectivas. El nuevo parálogo, libre de restricciones, no permanecerá funcional por mucho tiempo debido a la acumulación de mutaciones deletéreas, las cuales no serán removidas por selección purificadora (Nei & Roychoudhury, 1973).

b) *Modelo de sub-funcionalización.* Este mecanismo involucra la acumulación de mutaciones neutrales que finalmente serán mantenidas por selección negativa, donde la función original se distribuye entre ambos parálogos y funcionarán de forma complementaria (Hugnes 1994, Force *et al*. 1999, Lynch & Conery 2000, Lynch & Force 2000, Dittmar & Liberles 2011).

c) *Modelo de neo-funcionalización*. Este modelo implica la aparición de una nueva función en uno de los genes duplicados, mientras que la función original se mantiene en la otra copia (Ohno, 1970, Bergthorssoon *et al.* 2007, Dittman & Liberles 2011, Nasvall *et al.* 2012, Adler *et al.* 2014).

En un trabajo realizado para dilucidar el costo de la duplicación genética y su estabilidad, se estimó que la duplicación genética tiene un costo en la adaptación celular de 0.15% por kb de ADN, y por consiguiente se propone que este factor, no contemplado en los modelos clásicos (sub-funcionalización y neo-funcionalización), puede reducir sustancialmente la probabilidad de aparición de nuevas funciones genéticas, en ausencia de presión de selección que permita retener la región genética amplificada (Adler *et al.* 2014).

Por otro lado, el contexto genómico es importante en la identificación, función y comparación de ortólogos, debido a que los genes que se encuentran cercanos tienden a interaccionar (Huynen *et al.* 2000) y su expresión es afectada significativamente por su posición genómica (Kleinjan & Van Heyningen 1998, Dewey 2011). El contexto genómico de un gen también se encuentra asociado a su tasa de cambio y modo de evolución. En general, las secuencias de ortólogos con posiciones genómicas conservadas tienden a tener más restricciones evolutivas (Koski *et al.* 2001, Burgetz *et al.* 2006, Lemoine *et al.* 2007, Notebaart *et al.* 2005, Cusack & Wolfe 2007, Jun *et al.* 2009, Wang *et al.* 2010, Guerrero *et al.* 2005, Peralta *et al.* 2011) que aquellos que no conservan su posición ancestral.

Si bien es cierto que existen varias propuestas para la aparición de nuevas funciones a partir de eventos de duplicación genética, a la fecha el mecanismo molecular para la divergencia de genes duplicados no ha sido resuelto satisfactoriamente (Kaessmann 2010, Innan & Koondrashov 2010). La elucidación del mecanismo por el cual estos genes se retienen después de la duplicación y evolucionan a nuevas funciones es uno de los problemas fundamentales en la evolución molecular (Crow & Wagner 2006; Gu *et al.* 2002, Lynch & Conery 2000; Shiu *et al.* 2006; Wagner *et al.* 2003, Conrad *et al.* 2010, Chen *et al.* 2010).

Es ampliamente aceptado que las nuevas enzimas evolucionaron a partir de duplicación genética y eventos de funcionalización (Hughes 1994, Force *et al*. 1999, Conant & Wolfe 2008, Khersonsky & Tawfik 2010; Liberles *et al.* 2010, Noda-García *et al.* 2013). Esta hipótesis basada en duplicación, podría explicar los mecanismos primitivos del origen de nuevas enzimas debido a la poca disposición de regiones codificantes y a la escasa fuente de recursos genéticos (Hughes 1994, Force *et al*. 1999, Conant & Wolfe 2008, Khersonsky & Tawfik 2010, Liberles *et al*. 2010). Sin embargo, la duplicación genética como fuerza impulsora en la adaptación, es más frecuente en Eucariontes que en Procariontes (Liberles *et al.* 2010). Recientemente se ha propuesto que la transferencia horizontal (HGT, Horizontal Gene Transfer) y no la duplicación, es el principal mecanismo por el cual los procariontes expanden su genes y familias de proteínas (Treangen & Rocha 2011). Existen pocos datos respecto al impacto de las enzimas adquiridas recientemente por HGT sobre la fisiología y metabolismo de la bacteria (Pál *et al.* 2005, Klassen 2009, Noda-García *et al.* 2013, Nowell *et al.* 2014, Baltrus *et al.* 2011, Baltrus 2013).

La HGT es un mecanismo bacteriano muy común en alfa-proteobacterias donde al menos un evento de transferencia horizontal ha ocurrido por familia de proteínas (Kloesges *et al.* 2011). Por otra parte, al menos el 80% de los genes en cada genoma han participado en eventos de HGT en algún momento de su historia (Dagan *et al.* 2008). Un ejemplo lo constituye el gen que codifica para la prolina racemasa, la cual ha sido intercambiada entre varias α-proteobacterias, incluyendo *Agrobacterium tumefaciens*, *Paracoccus denitrificans*, *Brucella suis* y *Brucella melitensis* (Nester 2000). El intercambio genético de α-proteobacterias con otras bacterias, arqueas y eucariontes ha sido ampliamente revisado (Le *et al.* 2014). De este modo las α-proteobacterias las cuales son ancestros de la mitocondria vía endosimbiosis, contribuyen al cambio de contenido genético de organismos cercanos y lejanamente relacionados.

Recientemente, se han generado nuevos hallazgos acerca de genes duplicados, que sugieren una divergencia y silenciamiento epigenético por mecanismos de metilación (Rodin & Riggs 2003, Flavell 1994, Chang & Liao 2011, Yi & Goodisman 2009). En un trabajo actual (Keller & Soojin 2014), se demostró que las regiones promotoras de regiones duplicadas recientemente se encuentran hipermetiladas, mientras que las duplicaciones menos recientes están generalmente hipometiladas. Este estudio indica que las modificaciones epigenéticas podrían ser facilitadores de la evolución de genes duplicados, debido a su efecto sobre la divergencia funcional y/o a la dependencia potencial sobre los determinantes genómicos.

Por otro lado, en otro trabajo, donde se trató de estudiar el efecto sobre la adaptación de nuevas variaciones estructurales inducidas por duplicación y HGT en 48 cepas de *Sinorhizobium*, se encontró que la mayoría de los genes adquiridos por ambas vías tienen un efecto deletéreo, y que la eliminación mediante la selección purificadora prevalece en las nuevas adquisiciones genéticas (Epstein *et al.* 2014).

Características y evolución de los rizobios.

El orden Rhizobiales es un grupo bacteriano muy versátil, que presenta amplia diversidad en sus características genómicas, metabólicas y estilos de vida. Son α-proteobacterias cuyos genomas generalmente se encuentran distribuidos en varios replicones, como cromosomas circulares o lineales, plásmidos y megaplásmidos. Los rizobios son generalmente saprofíticos, pero pueden presentar varios tipos de metabolismos (Larimer *et al*. 2004). Respecto a sus estilos de vida, se les encuentra como organismos de vida libre, simbiontes de plantas o patógenos de plantas y animales (Sällström & Andersson 2005). En este orden se encuentran los generos *Rhizobium*, *Agrobacterium*, *Bradyrhizobium*, *Mesorhizobium*, *Sinorhizobium* (o *Ensifer*) y *Azorhizobium*.

La característica más importante y objeto de estudio de este Orden es su capacidad de inducir nódulos fijadores de nitrógeno atmosférico en las raíces de las plantas de la familia Leguminosae (y en la no leguminosa *Parasponia*). Algunos rizobios también son capaces de inducir nódulos en el tallo de leguminosas, como *Sesbania* y *Aeschynomene*. Estas bacterias son bacilos de 0.5-0.9 µm de diámetro y de 1.2-3.0 µm de longitud, de tinción Gram negativa y no esporulan. Presentan flagelos perítricos o un flagelo polar o subpolar. De acuerdo a su velocidad de crecimiento se les divide en dos grandes grupos:

• *Rhizobium*: cepas de crecimiento rápido con tiempos de duplicación de 2 a 4 horas, con varios flagelos, acidificantes en diferentes medios de fuentes de carbono. Estos rizobios generan nodulación en leguminosas de zonas templadas.

• *Bradyrhizobium*: cepas de crecimiento lento, alcalinizantes de diversos medios con tiempos de duplicación de mas de 6 horas, con un solo flagelo.

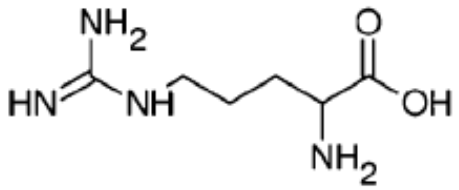
A esta clasificación se le agregó el género *Azorhizobium*, con cepas capaces de formar nódulos en tallos y raíces. Pueden fijar y asimilar nitrógeno atmosférico en cultivos puros (Willems 2006, Jordan 1984).

El ancestro común del orden de los Rhizobiales existió hace unos 400 millones de años, cuando se separaron los géneros *Rhizobium* y *Bradyrhizobium*. Esto es, 300 millones de años antes de la aparición de las plantas leguminosas (Sprent & Plaztmann 2001, Sprent & James 2007). El análisis de los genes parálogos de la glutamina sintetasa GSI y GSII (EC 6.3.1.2), que es una enzima clave en la asimilación de nitrógeno, ha permitido calibrar la divergencia entre procariontes y eucariontes, debido a que esta enzima se duplicó antes del origen de los eucariontes (Kumada *et al.* 1993) (Iwabe *et al.* 1989). La forma GSI se encuentra únicamente en procariontes, mientras que la forma GSII se encuentra únicamente en eucariontes, a excepción de los rizobios.

De acuerdo a Turner & Young 2000, los tiempos de divergencia para las rizobios de crecimiento rápido *(Rhizobium, Sinorhizobium, Mesorhizobium)* fue de 203-324 millones de años (Ma). Estos tiempos son más tempranos que la división de las plantas mono- y dicotiledoneas (156-171 Ma) y de la división de *Brassica* y las leguminosas (125-136 Ma). La posibilidad de que los rizobios empezaron a divergir antes de la aparición de las leguminosas ha sido propuesta (Young & Johnston 1989), quienes sugieren que la división de las rizobias de crecimiento rápido y lento hace 507 a 553 Ma, podría ser anterior a la existencia de las Angiospermas y las plantas terrestres.

La biosíntesis de arginina

La arginina es un aminoácido esencial y versátil en el metabolismo bacteriano y su síntesis requiere de una gran cantidad de energía y poder reductor (Fig. 5). Es uno de los aminoácidos con mayor contenido de nitrógeno, constituyente de las proteínas, con un precursor común en la síntesis de pirimidina (el carbamoilfosfato). La arginina también es un precursor de la síntesis de poliaminas, las cuales poseen una amplia gama de funciones biológicas (Igarashi & Kashiwagi 2010, Di Martino *et al.* 2013). El catabolismo de la arginina ha sido estudiado en varias bacterias como *E. coli* (Abdelal 1979, Cunin *et al.* 1986) y *Pseudomonas* (Lüthi *et al.* 1990; Itoh & Nakada 2004).



Arginina (Arg)

Fig. 5. Molécula del aminoácido arginina.

Se han definido tres vías para la biosíntesis de arginina procedentes a partir de L-glutamato (Fig. 6, Tabla 6): la vía lineal, la de reciclado y la presente en *Xanthomonas* (Lima & Menck 2008, Lu 2006). En la llamada vía lineal, presente en Enterobacterias, una molécula de acetil-CoA se consume para convertir L-glutamato en L-ornitina en cinco pasos, a través de una serie de intermediarios acetilados. En el quinto paso, ArgE (acetil-ornitina desacetilasa, EC 3.5.1.16) cataliza la hidrólisis de *N*-acetil-ornitina en ornitina y acetato (Lima & Menck 2008).

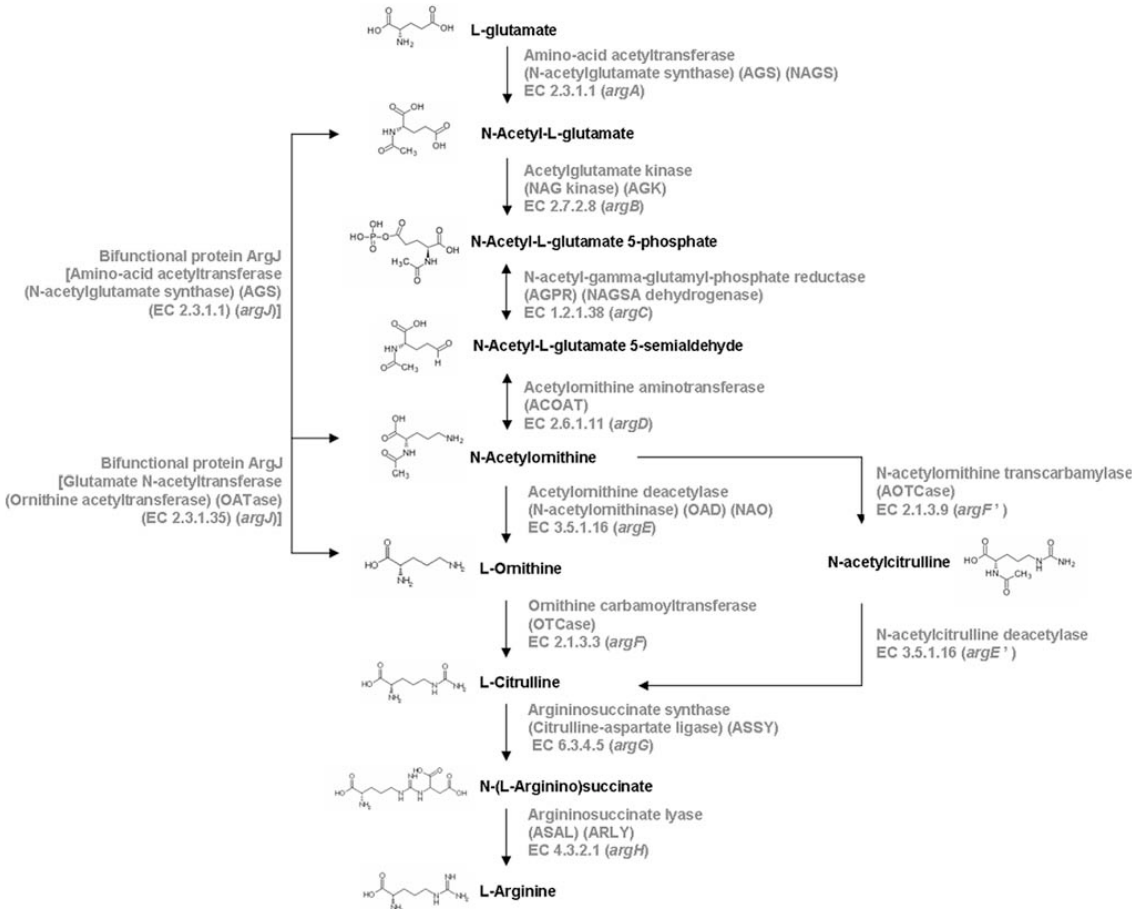


Fig. 6. Vía de la síntesis de arginina (Tomado de Lima & Menck 2008).

La vía de reciclado presente en bacterias (incluyendo *Rhizobium*) y algunos eucariontes microbianos, es una vía energeticamente más económica, ya que recicla el grupo acetilo sobre el glutamato, generando el *N*-acetil-glutamato. Esta reacción es catalizada por la ornitina acetiltransferasa (OAT; EC 2.3.1.35; *argJ*). En estos organismos, el primer paso en la vía cumple un papel anaplerótico: una vez que se sintetiza la *N*-acetil-ornitina, la acción de la *N*-acetil-glutamato sintasa (NAGS; EC 2.3.1.1; codificada por *argA*) se vuelve superflua (Cunin *et al.* 1986). La parte final de estas vías incluye tres reacciones más para llegar a la biosíntesis de L-arginina mediante el ensamble del grupo guanidino a partir de carbamoilfosfato y el grupo amino de aspartato.

Dependiendo de la vía, aparece la inhibición por retroalimentación por el producto final L-arginina sobre la *N*-acetil-glutamato sintasa (ArgA) en la ruta lineal, o sobre la enzima vía *N*-acetil-glutamato cinasa (ArgB) en la vía de reciclado. Por último, algunas bacterias poseen un ArgJ bifuncional, que pueden utilizar acetil-CoA y acetil-ornitina como compuestos donadores del grupo acetilo (Marc *et al*. 2000).

Tabla 6. Enzimas que participan en la biosíntesis y degradación de arginina.

Actividad	Gen codificante
N-acetil-glutamato sintasa	*argA*
N-acetil-glutamato cinasa	*argB*
N-acetil-glutamil fosfato reductasa	*argC*
N-acetil-ornitina transferasa	*argD*
N-acetil ornitinasa	*argE*
Ornitina-carbamoil transferasa	*argF, argI*
Arginino-succinato sintasa	*argG*
Arginino-succinasa	*argH*
Carbamoil-fosfato sintasa	*carAB*
Arginina descarboxilasa	*speA*
Agmatina ureohidrolasa	*speB*
Ornitina descarboxilasa	*speC*
S-adenosil-metionina descarboxilasa	*speE*
Lisina descarboxilasa	*cadA*

Sintenia

El termino sintenia (del griego; syn=juntos, taenia=cinta) se utiliza en genética para indicar la conservación de dos o más *loci* genéticos en replicones de diferentes especies. La identificación de regiones genéticas homólogas al comparar dos o más organismos es importante no únicamente con fines de anotación de genomas o el descubrimiento de nuevas regiones funcionales, sino por el estudio de las relaciones evolutivas entre las especies comparadas.

Desde 1995, cuando se obtuvieron los primeros genomas, los de *Haemophilus influenzae* (Fleischmann *et al.* 1995) y *Mycoplasma genitalium* (Fraser *et al.* 1995), se analizó el grado de conservación del orden de los genes ortólogos en los genomas bacterianos. A la fecha, existen varios trabajos sobre sintenia en bacterias, plantas, vertebrados y mamiferos (Li *et al.* 2013, Citerne *et al.* 2013, Cheng *et al.* 2012, Kemkemer *et al*. 2009, Rauf & Mir 2013). En Rhizobiales, nuestro grupo publicó uno de los primeros trabajos al respecto (ver Figuras 7 y 8) (Guerrero *et al.* 2005). En estos estudios se encontró que existe una relación directa entre la cercania filogenética y el grado de conservación del orden de los genes ortólogos. Los organismos mas cercanos tienen una sintenia mas conservada, y a medida de que la relación filogenética se incrementa, la sintenia se va reduciendo. Por otro lado, la sintenia también puede ser utilizada como una medida auxiliar en la identificación de ortólogos (Jun *et al.* 2009).

Otro término relacionado a la sintenia es la colinearidad, la cual se define como la conservación del orden genético en replicones de diferentes especies. Este concepto no toma en cuenta la conservación de los *loci* genéticos independientes del replicón como la sintenia, sino que hace comparaciones de dos segmentos genéticos o replicones.

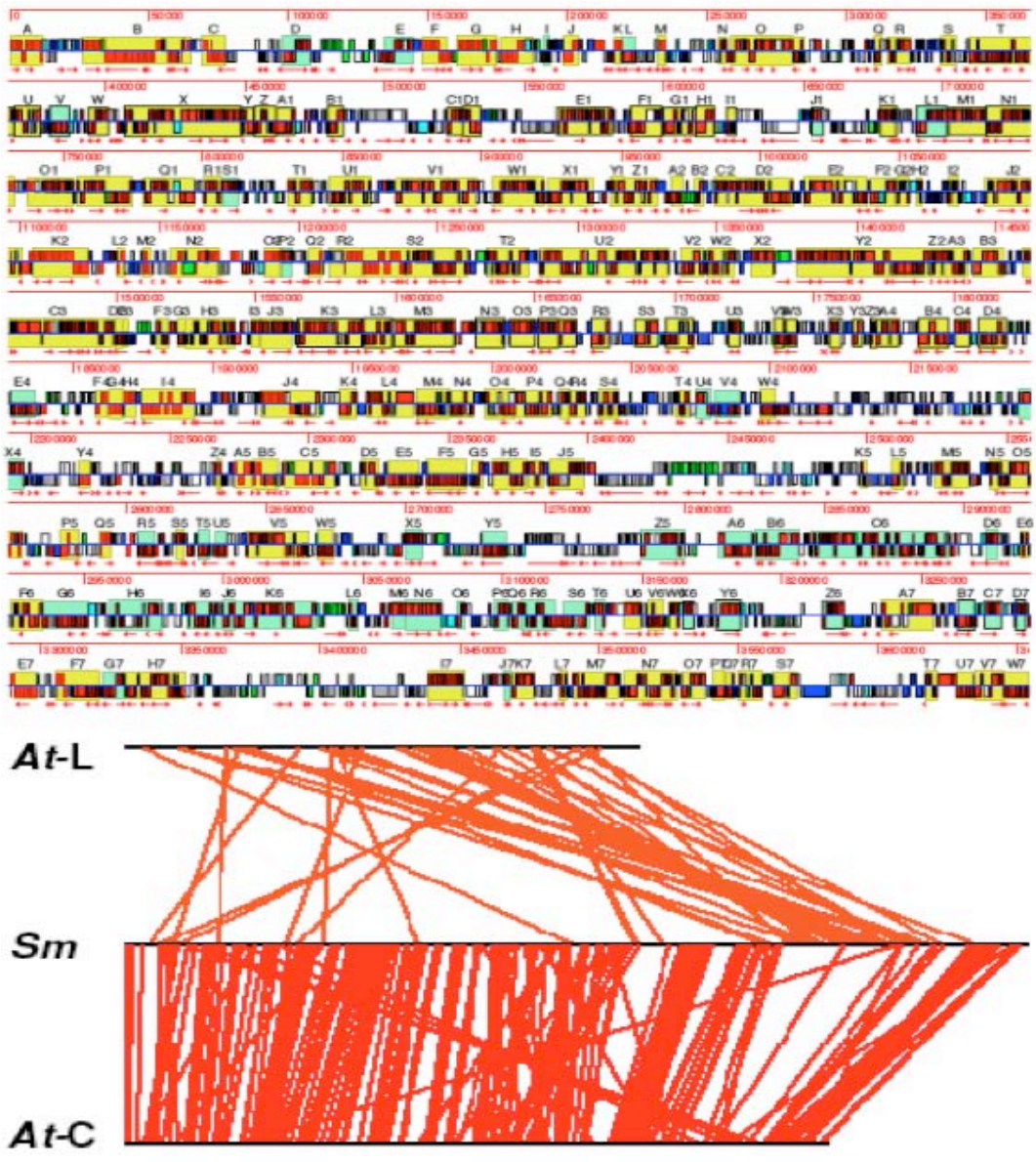


Fig. 7. Representación gráfica de las regiones de sintenia compartidas entre el cromosoma circular de *Sinorhizobium meliloti* 1021 (*Sm*) y los cromosomas de *Agrobacterium tumefaciens* C58, circular (At-C) y lineal (At-L). Arriba, histograma de sintenia del cromosoma de *Sm* en comparación con los cromosomas At-C y At-L. Las regiones de microsintenia se indican con números y letras. En cuadros amarillos las regiones microsinténicas compartidas con At-C y en verde con At-L. En barras de color rojo, los genes ortólogos con At-C y At-L. En barras de color azul oscuro y claro, los genes no sinténicos con At-C o At-L, respectivamente. Las barras grises representan genes específicos de cada especie. Las barras blancas, homólogos con otros cromosomas de Rhizobiales. La dirección de la transcripción se denota por las posición, superior o inferiror, con respecto a la línea central. La predicción

de operones se indican con flechas rojas. Escala en pares de bases (pb). Abajo, regiones de microsintenia compartidas entre el cromosoma circular de *S. meliloti* y los cromosomas circular y lineal de *A. tumefaciens*. Las líneas negras representan los replicones y las rojas la ubicación de las regiones de microsintenia (tomado de Guerrero *et al.* 2005).

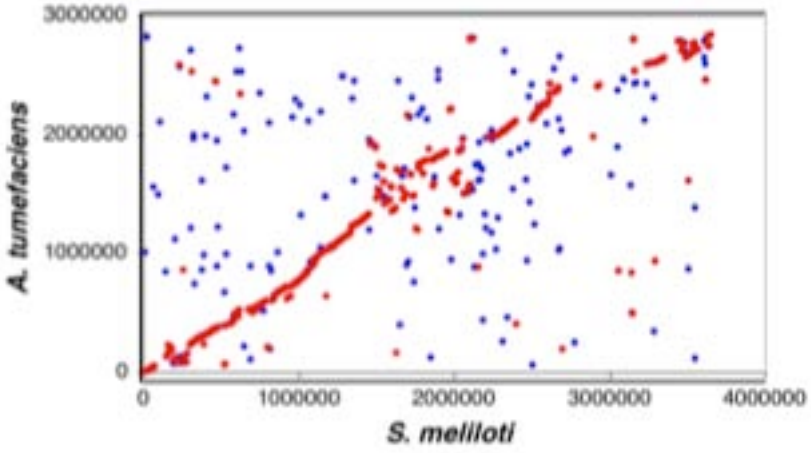


Fig. 8. Gráfica que representa la colinearidad de los cromosomas circulares de *Sinorhizobium meliloti* 1021 y *A. tumefaciens* C58. Los genes ortólogos fueron agrupados en regiones de microsintenia (puntos rojos). Los puntos de color azul son regiones no sinténicas (Guerrero *et al.* 2005).

Genes ortólogos en Rhizobiales

Basados en la disponibilidad de secuencias genómicas del grupo de Rhizobiales que nos interesa, los Rhizobiales, fue posible desarrollar un proyecto de genómica comparativa para contestar principalmente preguntas acerca de la relación entre la conservación o divergencia en los genes comunes en estos organismos y su localización cromosomal. En 2005 publicamos los resultados del proyecto (Guerrero *et al.* 2005), que incluyeron cinco genomas de bacterias Rhizobiales: *Rhizobium etli*, *Sinorhizobium meliloti, Mesorhizobium loti, Agrobacterium tumefaciens y Brucella melitensis.* Los tres primeros son simbiontes de plantas leguminosas y los dos últimos patógenos en plantas y animales, respectivamente.

Encontramos que estas cinco especies de bacterias comparten una gran proporción de sus genes ortologos, principalmente en los cromosomas. Estos genes ortólogos detectados se dividieron de acuerdo a su conservación del orden o sintenia.

La sintenia puede ser una característica genética susceptible de perderse por recombinación, rearreglos cromosomales, inserción de transposones, transferencia horizontal, etc. Sin embargo, según nuestro análisis y análisis previos, la sintenia representa un nivel de organización importante en los cromosomas bacterianos, debido, tal vez, a que los complejos y funciones celulares bacterianas requieren de segmentos de genes contiguos, con un orden conservado, para un mejor desarrollo de su función (Guerrero *et al*. 2005, Król *et al*. 2007).

El mejor ejemplo sobre la conservación del orden genético y funcional es el del operón, en el que además se modula y se coordina la expresión de sus componentes. En los análisis y datos publicados demostramos que los ortólogos sinténicos codifican para la mayor parte de las funciones esenciales, presentan mayor restricción al cambio en sus secuencias y tienen una mayor interacción funcional (formación de redes) que los genes no sinténicos (Guerrero *et al*. 2005).

En este trabajo llevado a cabo por nuestro grupo de investigación, se identificaron 159 regiones sinténicas cromosomales comunes en las cinco especies de bacterias Rhizobiales analizadas (además de *Rhizobium etli*), con 1002 genes y una tendencia evolutiva diferencial respecto de genes ortólogos sin sintenia. Donde concluimos que los genes sinténicos presentes en Rhizobiales son un grupo que evoluciona en conjunto, presenta resistencia a variaciones en su secuencia y a rearreglos genéticos.

En otro trabajo, realizado por nuestro grupo y consecutivo al estudio de ortólogos de Rhizobiales, se analizó la correlación entre la variabilidad de la secuencia de la proteínas ortólogas y sus propiedades fisicoquímicas (como polaridad, carga electrostática, estructura secundaria, volumen molecular y composición de aminoácidos). Los resultados de este análisis sugirió que las propiedades fisicoquímicas están relacionadas con el papel funcional de los ortólogos (Peralta *et al.* 2011).

Firma de la especie.

En nuestro grupo de Investigación, se propuso por primera vez el concepto de firma de la especie (FDE) para denominar a los cambios específicos observados en las secuencias de genes ortólogos. La FDE como una forma de detectar la proporción de residuos en una proteína que están divergiendo en respuesta a la adaptación funcional (Guerrero *et al.* 2005). La validez de esta medida se observó al comparar los valores de firma de la especie de 1416 ortólogos de cinco genomas con la tasa evolutiva de sustitución no sinónima de los mismos genes (Peralta *et al.* 2011). La alta correlación obtenida (0.932 a 0.953) reveló que la firma de la especie lograba capturar la tendencia evolutiva de los genes usando secuencias de aminoácidos.

El procedimiento para identificar los residuos que han variado más específicamente entre varios ortólogos, se realiza primero con un alineamiento múltiple de ClustalW, al que se aplica un algoritmo mencionado más adelante y se expresa como un porcentaje del número total de aminoácidos alineados.

El algoritmo para el cálculo de la firma de la especie se basa en lo siguiente: Las posiciones consideradas como parte del porcentaje de cambios por especie son aquellas que cumplen con la restricción $0 < firma \leq n$, donde (n) es la frecuencia mínima de residuos en cada posición del alineamiento, que varía de acuerdo a la cantidad de secuencias analizadas dada por $n = (total\ de\ secuencias)/6 + 1$, es decir un cambio cada 5 organismos; este parámetro puede ser modificado de acuerdo a la especificidad requerida. Los *gaps* y las secuencias no alineadas no son considerados y se eliminan del análisis. El porcentaje de cambio se calcula dividiendo los residuos específicos de cada especie entre la longitud alineada (Guerrero *et al.* 2005). En la Figura 9 se observa el cálculo de la FDE para ArgC.

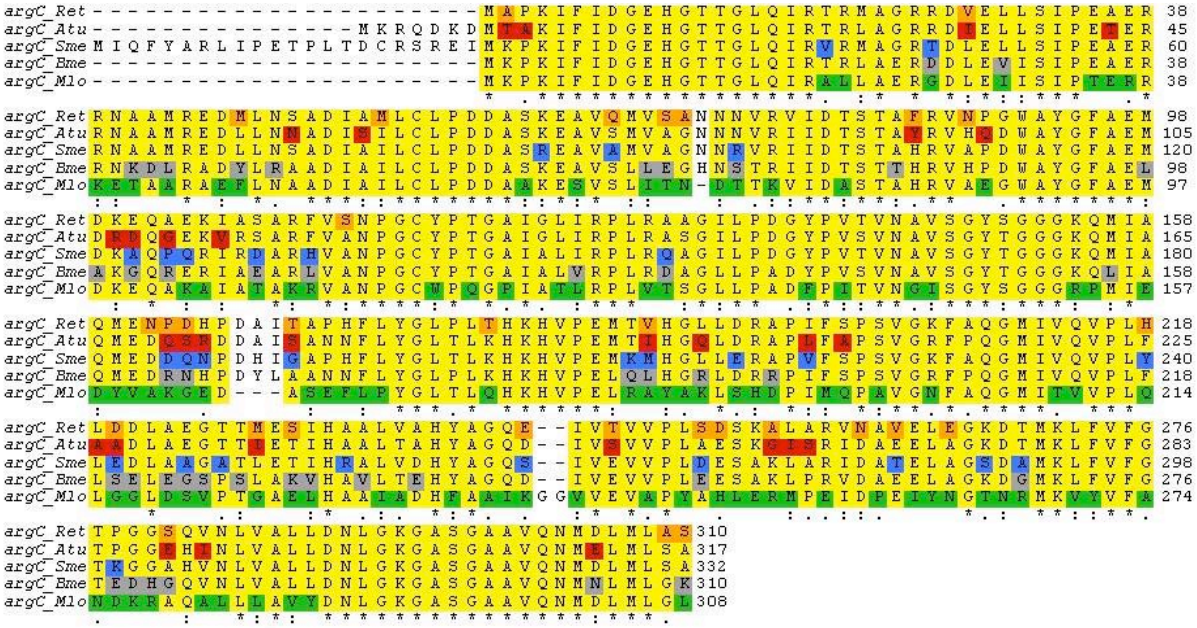


Fig. 9. Alineamiento de las proteínas ArgC de 5 especies de Rhizobiales mostrando el concepto FDE. En naranja, residuos específicos para *Rhizobium etli (Re)*, en rojo para *A*

Agrobacterium tumefaciens (At), en azul para *Sinorhizobium meliloti (Sm)*, en gris para *Brucella melitensis (Bm)* y en verde para *Mesorhizobium loti (Ml)*. 133 posiciones con residuos identicos. Residuos específicos (FDE): *Re,* 32 residuos (10.3%); At, 30 residuos (9.6%); Sm, 30 residuos (9.6%); Ml, 113 residuos (36.4%) (Díaz *et al.* 2011).

Planteamiento del trabajo

En este trabajo proponemos que la variación en secuencia de las proteínas ortólogas no son únicamente un rasgo filogenético, sino más bien representan adaptación específica a un nicho particular (Guerrero *et al.* 2005, Peralta *et al.* 2011).

Para comprobar nuestras observaciones, se sustituyeron los ortologos *argC* de Rhizobiales mediante complementación genética, con el propósito de analizar a nivel funcional el efecto del intercambio de ortologos cercanos que aunque cumplen la misma función esencial, no estan adaptados al microambiente específico del organismo receptor (*S. meliloti* y *E. coli*).

Los reemplazos de ortólogos fueron analizados a diferentes niveles, desde la transcripción, la traducción, actividad enzimática, cantidad de la enzima producida, el balance de metabolitos y la tasa de crecimiento de la bacteria.

Artículo de investigación.

Journal of Bacteriology

argC Orthologs from *Rhizobiales* Show Diverse Profiles of Transcriptional Efficiency and Functionality in *Sinorhizobium meliloti*

Rafael Díaz, Carmen Vargas-Lagunas, Miguel Angel Villalobos, Humberto Peralta, Yolanda Mora, Sergio Encarnación, Lourdes Girard and Jaime Mora

J. Bacteriol. **2011, 193(2):460. DOI: 10.1128/JB.01010-10. Published Ahead of Print 12 November 2010.**

JOURNAL OF BACTERIOLOGY, Jan. 2011, p. 460–472 Vol. 193, No. 2
0021-9193/11/$12.00 doi:10.1128/JB.01010-10

argC Orthologs from *Rhizobiales* Show Diverse Profiles of Transcriptional Efficiency and Functionality in *Sinorhizobium meliloti*▽†

Rafael Díaz, Carmen Vargas-Lagunas, Miguel Angel Villalobos,‡ Humberto Peralta, Yolanda Mora, Sergio Encarnación, Lourdes Girard, and Jaime Mora*

Programa de Genómica Funcional de Procariotes, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad, Cuernavaca, Morelos C.P. 62210, Mexico

Received 24 August 2010/Accepted 31 October 2010

Several factors can influence ortholog replacement between closely related species. We evaluated the transcriptional expression and metabolic performance of ortholog substitution complementing a *Sinorhizobium meliloti argC* mutant with *argC* from *Rhizobiales* (*Agrobacterium tumefaciens*, *Rhizobium etli*, and *Mesorhizobium loti*). The *argC* gene is necessary for the synthesis of arginine, an amino acid that is central to protein and cellular metabolism. Strains were obtained carrying plasmids with *argC* orthologs expressed under the *speB* and *argC* (*S. meliloti*) and *lac* (*Escherichia coli*) promoters. Complementation analysis was assessed by growth, transcriptional activity, enzymatic activity, mRNA levels, specific detection of ArgC proteomic protein, and translational efficiency. The *argC* orthologs performed differently in each complementation, reflecting the diverse factors influencing gene expression and the ability of the ortholog product to function in a foreign metabolic background. Optimal complementation was directly related to sequence similarity with *S. meliloti*, and was inversely related to species signature, with *M. loti argC* showing the poorest performance, followed by *R. etli* and *A. tumefaciens*. Different copy numbers of genes and amounts of mRNA and protein were produced, even with genes transcribed from the same promoter, indicating that coding sequences play a role in the transcription and translation processes. These results provide relevant information for further genomic analyses and suggest that orthologous gene substitutions between closely related species are not completely functionally equivalent.

Synteny, gene neighboring, or conservation of chromosomal gene order has been proposed to be a result of the interdependence between a gene product and its genomic context. Conservation of a syntenic block could be favored by selective pressure because the arrangement allows the sequential or coordinated expression of genes that correctly integrate important metabolic functions (41, 64, 69). By a comparative analysis of four *Rhizobiales* species (namely, *Sinorhizobium meliloti*, *Agrobacterium tumefaciens*, *Mesorhizobium loti*, and *Brucella melitensis*), we found that syntenic genes exhibit striking differences compared to nonsyntenic genes, including increased operon and network organization, high sequence conservation, diminished evolutionary rates, essential functional role, and specific phylogenetic associations (32; H. Peralta, G. Guerrero, A. Aguilar, and J. Mora, unpublished data).

Orthologs encode similar functions sharing a common ancestor and exhibiting various degrees of conservation, due in large part to functional adaptation and cellular role in different species (45). We have proposed a new parameter termed "species signature" to extract the amount of amino acid residues specific for a given species, based on multiple sequence alignment. We hypothesized that the species signature represents the proportion of a particular sequence that responds to adaptation (32); it is a useful evolutive measure because we found a high direct correlation with the nonsynonymous substitution rate (Peralta et al., unpublished). Although orthologous genes of *Rhizobiales* encode proteins with high degrees of identity among several species, the proteins differ in many respects, including isoelectric point. Small differences in protein sequence may be the result of evolutionary adaptation in response to the particular intracellular environment of a given species. In this scenario, specific amino acid changes would be selected for optimal performance in a particular genetic background. Knight et al. (40) carried out a virtual analysis of proteomes of nearly 100 organisms and showed that there was a correlation between theoretical proteomes and ecological niches; conversely, there was no correlation between phylogeny and differences observed in the theoretical proteomes. Our previous work has shown that mutation of a single gene, *aniA* (a carbon flux regulator), produced a proteomic alteration of approximately 800 proteins (16, 22), indicating that the absence or modification of a single gene can result in complex changes in global gene expression. In this context, syntenic orthologs are ideal to evaluate the functional importance of species signature in related organisms.

The *Rhizobiales* order is a versatile group of bacteria that present very interesting features, such as a huge amount of genes, abundance of genes acquired by horizontal transfer events, a symbiotic or pathogenic association with higher organisms such as plants or animals, and the ability for nodulation and nitrogen fixation in some species (9, 13, 29, 31, 50, 54,

* Corresponding author. Mailing address: Programa de Genómica Funcional de Procariotes, Centro de Ciencias Genómicas-UNAM, Apdo. postal 565-A, Cuernavaca, Morelos C.P. 62220, México. Phone: 52 777 3114661. Fax: 52 777 3175094. E-mail: jmora@ccg.unam.mx.

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

‡ Present address: Centro de Investigación en Biotecnología Aplicada-IPN, Ex Hacienda San Juan Molino, Tlaxcala C.P. 90700, Mexico.

▽ Published ahead of print on 12 November 2010.

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics[a]	Source or reference
Rhizobiales strains		
Sinorhizobium meliloti 1021^T	Wild type (type strain)	51
Rhizobium etli CFN42^T	Wild type (type strain)	62
Rhizobium etli CE3	Sm^r derivative of CFN42	53
Agrobacterium tumefaciens C58	Wild-type strain	ATCC 33970
Mesorhizobium loti MAFF303099	Wild-type strain	38
Sinorhizobium meliloti argC mutant	*S. meliloti* 1021 derivative, *argC*::ΩSp	This study
Escherichia coli strains		
DH5α	*supE44ΔlacU169*(ϕ80*lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	34
HB101	*supE44 hsdS20* (r_B^- m_B^-) *recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1*	6
Plasmids		
pCR2.1 TOPO	Cloning vector for PCR products; Ap^r Km^r	Invitrogen
pJQ200SK$^+$	Suicide cloning vector; Gm^r	60
pBBR1MCS3	Broad host range cloning vector; Tc^r	42
pBBMCS53	Δp*lacZ* pBBR1MCS5 derivative, carrying the promoterless *gus* gene; Gm^r	28
pRK2073	Conjugation helper plasmid; Sm/Sp^r	4
pRK2013	Conjugation helper plasmid; Km^r	25
pHP45ΩSp	Plasmid containing the Ω-Sp interposon; Sp^r	57
pFGP1	pBBR1MCS3 derivative with *S. meliloti speB-argC* genes	This study
pFGP2	pBBR1MCS3 derivative with *A. tumefaciens speB-argC* genes	This study
pFGP3	pBBR1MCS3 derivative with *R. etli speB-argC* genes	This study
pFGP4	pBBR1MCS3 derivative with *M. loti speB-argC* genes	This study
pFGP5	pBBMCS53 derivative p*lac::gus* transcriptional fusion	This study
pFGP6	pBBMCS53 derivative p$argC_{Sm}$::*gus* transcriptional fusion	This study
pFGP7	pBBMCS53 derivative p$speB_{Sm}$::*gus* transcriptional fusion	This study
pFGP8	pBBMCS53 derivative p*lac-*$argC_{Sm}$::*gus* transcriptional fusion	This study
pFGP9	pBBMCS53 derivative p*lac-*$argC_{At}$::*gus* transcriptional fusion	This study
pFGP10	pBBMCS53 derivative p*lac-*$argC_{Re}$::*gus* transcriptional fusion	This study
pFGP11	pBBMCS53 derivative p*lac-*$argC_{Ml}$::*gus* transcriptional fusion	This study
pFGP12	pBBMCS53 derivative p*argC-*$argC_{Sm}$::*gus* transcriptional fusion	This study
pFGP13	pBBMCS53 derivative p*argC-*$argC_{At}$::*gus* transcriptional fusion	This study
pFGP14	pBBMCS53 derivative p*argC-*$argC_{Re}$::*gus* transcriptional fusion	This study
pFGP15	pBBMCS53 derivative p*argC-*$argC_{Ml}$::*gus* transcriptional fusion	This study
pFGP16	pBBMCS53 derivative p*speB-*$argC_{Sm}$::*gus* transcriptional fusion	This study
pFGP17	pBBMCS53 derivative p*speB-*$argC_{At}$::*gus* transcriptional fusion	This study
pFGP18	pBBMCS53 derivative p*speB-*$argC_{Re}$::*gus* transcriptional fusion	This study
pFGP19	pBBMCS53 derivative p*speB-*$argC_{Ml}$::*gus* transcriptional fusion	This study
pFGP20	pBBMCS53 derivative carrying the *S. meliloti argC* gene under its promoter	This study
pFGP21	pBBMCS53 derivative carrying the *A. tumefaciens argC* gene under the control of the *S. meliloti argC* promoter.	This study
pFGP22	pBBMCS53 derivative carrying the *R. etli argC* gene under the control of the *S. meliloti argC* promoter	This study
pFGP23	pBBMCS53 derivative carrying the *M. loti argC* gene under the control of the *S. meliloti argC* promoter	This study
pFGP24	pBBMCS53 derivative p$argC_{At}$::*gus* transcriptional fusion	This study
pFGP25	pBBMCS53 derivative p$argC_{Re}$::*gus* transcriptional fusion	This study
pFGP26	pBBMCS53 derivative p$argC_{Ml}$::*gus* transcriptional fusion	This study

[a] Tc^r, tetracycline resistance; Ap^r, ampicillin resistance; Sm^r, streptomycin; Sp^r, spectinomycin resistance; Km^r, kanamycin resistance; Gm^r, gentamicin resistance.

65). In addition to a circular chromosome, some rhizobia carry secondary chromosomes or plasmids of high molecular size. Arginine is an essential amino acid in bacteria, and its synthesis requires a great deal of energy and reducing power (14, 49, 63). Arginine is the most common nitrogen storage compound and precursor of polyamine synthesis (48, 61). *argC* genes are commonly organized in operons, exemplified by the *Escherichia coli argCBH* operon (56). In *R. etli*, the *argC* mutation affects growth capacity using ammonium as the sole nitrogen source and results in failure to nodulate the common bean root (23).

We were interested in exploring whether the species signature (particular amino acid sequence variations) in orthologous syntenic genes reflects adaptations to the particular conditions of a given species, both intracellular (interactions with other enzymes and metabolites) and extracellular. The effect of heterologous complementation of an *S. meliloti argC* mutant with the corresponding ortholog from other *Rhizobiales* species (*A. tumefaciens*, *R. etli*, and *M. loti*) was compared in terms of growth, excretion of organic compounds, transcriptional and enzymatic activity, and molecular parameters such as mRNA and translation efficiency. An analysis of the putative *speB-argC* operon was also carried out to define a novel functional *argC* promoter.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Bacterial strains and plasmids used in the present study are listed in Table 1. The *S. meliloti argC* mutant was obtained

cloning an 1.3-kb *argC* SmaI-SpeI fragment into the pJQ200SK+ vector (60). The pHP45Ω-Sp plasmid (57) was digested with BamHI to obtain the Ω-Sp interposon, and it was cloned on a BglII site of the *argC* fragment. The obtained plasmid was conjugated to the *S. meliloti* 1021 wild-type strain, and double recombinants were selected on minimal medium (MM) (19) (1.2 mM K_2HPO_4, 0.8 mM $MgSO_4$, 1.5 mM $CaCl_2$, 0.01 mM $FeCl_3$), supplemented with 10 mM succinic acid and 10 mM ammonium chloride (NH_4Cl). The wild-type, *argC* mutant, and *argC* complemented strains were maintained in PY medium, as described previously (19). When necessary, antibiotics were added at the following concentrations: nalidixic acid, 20 μg/ml; streptomycin, 200 μg/ml; spectinomycin, 100 μg/ml; gentamicin, 15 μg/ml; and tetracycline, 5 μg/ml. MM was used supplemented with 10 mM succinic acid or mannose as a carbon source and 10 mM ammonium chloride (NH_4Cl) or potassium nitrate (KNO_3) as a nitrogen source. Cultures were grown aerobically at 30°C in a shaker at 200 rpm for all experiments. To confirm that the complementation phenotypes observed were due exclusively to plasmid-encoded sequences, in some cases the plasmids were cured and conjugation repeated. Similar results were obtained (data not shown).

Introduction of overlapping ends by PCR amplification using chimeric primers. Hybrid fusion genes containing *argC* genes from *S. meliloti* 1021, *A. tumefaciens* C58, *R. etli* CE3, and *M. loti* MAFF303099 under the control of different promoters were constructed by using overlap extension PCR methodology (66). Chimeric primers were designed with additional 5′ sequences to introduce homologous ends into the fragments to be fused. We used Oligo 6.0 software to construct the primer sequences, and the primers were purchased from the Unidad de Síntesis Química (IBt-UNAM).

The *S. meliloti argC* and *speB* regulatory regions containing overlapping ends corresponding to each *argC* ortholog were amplified by PCR using genomic DNA from *S. meliloti* strain 1021 as the template. Eight different PCRs were performed to obtain specific 896- and 500-bp PCR products with distinct overlapping ends. The specific primers used to generate the hybrid genes are listed in Table S1 in the supplemental material Amplification was carried out using a MasterCycler 5330 (Eppendorf, Hamburg, Germany) and AccuPrime *Pfx* DNA polymerase (Invitrogen). The reaction conditions were as follows: denaturation at 94°C for 2 min, followed by 30 cycles of 94°C for 20 s, 62°C for 30 s, and 68°C for 60 s. A final elongation step was carried out at 72°C for 10 min.

The complete *argC* coding regions of *A. tumefaciens* C58 (932 bp), *R. etli* CFN42 (932 bp), and *M. loti* MAFF303099 (926 bp) were obtained by PCR amplification using chimeric primers containing the *S. meliloti argC* promoter region overlapping ends. Total DNA from each strain was used as a template. The reaction conditions were as follows: denaturation at 94°C for 2 min, followed by 30 cycles of 94°C for 15 s and 68°C for 60 s. A final elongation step was carried out at 72°C for 10 min. For these reactions, Pfx50 DNA polymerase from Invitrogen was used. All PCR products were analyzed by agarose gel electrophoresis and then purified by using a QiaQuick gel extraction kit from Invitrogen.

To join the two DNA fragments, 1 μl of each PCR was mixed and cycled without primers. The reaction conditions were as follows: an initial denaturation step at 95°C for 2 min; followed by 20 cycles of 95°C for 1 min; annealing at 68°C for 10 s, 66°C for 5 s, 64°C for 5 s, and 62°C for 5 s; and extension at 72°C for 90 s. An additional extension was carried out at 72°C for 10 min. Portions (2 μl) of this reaction were mixed with the appropriate outer primers, and the assembled products were recovered. The reaction conditions were as follows: initial denaturation at 95°C for 2 min, followed by 30 cycles of 95°C for 15 s and 70°C for 2 min, followed by an additional extension at 72°C for 10 min. The reactions were carried out by using AccuPrime *Taq* DNA Polymerase High Fidelity (Invitrogen). The hybrid genes were cloned into pCR2.1 TOPO. All plasmids were sequenced to confirm that the fused gene was obtained with no nucleotide changes. Each set of constructs presented identical flanking sequences, including the regulatory ones, so that the only variant was the *argC* coding sequence from each organism.

Construction of plasmids for complementation experiments. Plasmids able to replicate in *S. meliloti* were constructed by using the cloning vector pBBMCS53, a derivative of pBBR1MCS5 carrying the promoterless *gus* gene (28). This permitted the complementation and the transcriptional monitoring with the same construction. The sequences of all primers used are listed in Table S1 in the supplemental material. A plasmid, pFGP20, containing a 1,889-bp PCR fragment corresponding to the *S. meliloti argC* regulatory and coding regions, was constructed. Total DNA isolated from *S. meliloti* strain 1021 was used as the template for PCR with the primers SmUpXba and Lw_H1_PSm. The resulting fragment was cloned into pCR2.1 TOPO for sequencing (pTOPO::p$argC_{Sm}$-$argC_{Sm}$). An 896-bp regulatory region and the complete *argC* coding region were cloned into the pBBMCS53 plasmid as a 1,937-bp XbaI fragment. The 1,906-bp EcoRI hybrid fragment from TOPO::p$argC_{Sm}$-$argC_{At}$ was cloned into pBBMCS53 to generate pFGP21. To generate pFGP22, the 1,924-bp XbaI-BamHI fragment from TOPO::p$argC_{Sm}$-$argC_{Re}$ was cloned into pBBMCS53. The XbaI-BamHI fragment from TOPO::p$argC_{Sm}$-$argC_{Ml}$ was cloned into the same vector to generate pFGP23. The resulting recombinant plasmids were analyzed by restriction enzyme digestion and nested PCR to confirm that they contained the proper hybrid gene. Another complete set of plasmids was obtained, containing only *argC* genes on pBBR1MCS3 vector, under each of the mentioned promoters in order to be used in mRNA and proteomic ArgC detection and quantification.

Construction of β-glucuronidase transcriptional fusions. To analyze the level of expression of each used promoter and *argC* orthologs or their segments, we constructed β-glucuronidase (*gus*) transcriptional gene fusions. The recombinant plasmids described above were used as templates for PCR with the primers SmUpXba and FusC_gus_R. Amplified fragments of the correct size were cloned into pCR2.1 TOPO. Four additional plasmids were constructed carrying the 896-bp regulatory region of *S. meliloti argC* and the complete *argC* gene sequence of each species (lacking the last 50 nucleotides) fused with the *gus* gene. This cassette was inserted into pBBMCS53 as an XbaI-KpnI fragment. Plasmids pFGP12, pFGP13, and pFGP14 contained an 1,828-bp XbaI-KpnI fragment that corresponded to *argC* genes from *S. meliloti*, *A. tumefaciens*, and *R. etli* under the *S. meliloti argC* promoter, respectively. A plasmid containing the *M. loti argC* gene under the control of the *S. meliloti argC* promoter was constructed by cloning an 1,822-bp XbaI-KpnI fragment in pBBMCS53 (pFGP15).

Conjugative transfer of plasmids from *E. coli* to *Rhizobium* was done by triparental mating, using either pRK2013 (25) or pRK2073 (4) as a helper. For determination of plasmid profiles, a modified Eckhardt procedure was used (36).

Determination of plasmid copy number. The plasmid copy number was assessed in p*speB* promoter-plasmids, by PCR, using *tetA* gene from plasmids and *rpoA*, a chromosomal gene, to calibrate the results. Complemented strains were grown for 8 h with shaking (200 rpm) at 30°C in MM containing succinate-ammonium. The initial inoculation was at an optical density at 540 nm (OD_{540}) of 0.05. An aliquot of the cell culture (2 ml) was removed, and the cells were washed with sterile, deionized water. The cells were lysed by heat treatment at 95°C for 20 min. The lysate was diluted 500-fold and used as a template for PCR amplification of the *tetA* gene (for quantification of plasmid) and *rpoA* gene (as an endogenous, single-copy control). Amplification and detection of DNA by real-time PCR was performed by using the ABI Prism 7300 sequence detection system (Applied Biosystems) in optical-grade 96-well plates. Triplicate samples were routinely used for the determination of DNA by real-time PCR. The reaction conditions for amplification of DNA were as follows: 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 data analysis was carried out by using sequence detection software (version 1.6.3; Applied Biosystems).

RACE. The transcriptional start site of the *S. meliloti argC* gene was identified by using the 5′ rapid amplification of cDNA ends (5′RACE) system (version 2.0; Invitrogen). cDNA was synthesized from 1 to 5 μg of total RNA with the complementary primer argCGSP-1 (corresponding to nucleotides 415 to 430 of the *argC* coding region). After first-strand cDNA synthesis, mRNA was removed by treatment with RNase and then the cDNA was purified by using a SNAP column (Invitrogen). Terminal deoxynucleotidyltransferase and dCTP were used to add homopolymeric tails to the 3′ ends of the cDNA. Nested PCR was carried out by using the gene-specific primer argCGSP-2 (corresponding to nucleotides 340 to 355) and the abridged anchor primer (AAP) GGCCACGCGTCGACTA GTACGGGIIGGGIIGGGIIG provided with the kit. The amplified fragment was cloned into pCR2.1 TOPO for sequence analysis.

RNA isolation, cDNA probe obtention, and real-time RT-PCR assays. RNA extraction was performed by using an RNeasy kit according to the manufacturer's instructions (Qiagen, Inc., Valencia, CA). Samples were treated with DNase I prior to reverse transcription-PCR (RT-PCR) using the protocol for preparation of DNA-free RNA (Fermentas). cDNA was synthesized from purified total RNA by RT using random primers and reverse transcriptase from the RevertAid H Minus First Strand cDNA synthesis kit (Fermentas). For real-time RT-PCR, TaqMan Universal PCR Mastermix, primer mixes containing a FAM reporter probe, and the gene-specific TaqMan MGB probe (6-FAM dye-labeled) (TaqMan gene expression assay) were obtained from Applied Biosystems. RT-PCR mixtures (25 μl) contained template cDNA, 2× TaqMan Universal PCR Mastermix buffer (12.5 μl), and forward and reverse primers. The optimal primer and probe concentrations and amplification conditions were determined. The forward and reverse *argC* primers and TaqMan probe for each species were added at concentrations of 5 and 2.5 pmol/μl, respectively.

The *S. meliloti rpoA* gene encodes the α subunit of RNA polymerase (RNAP) and is stably transcribed throughout different growth stages. Expression of *rpoA* was used as an endogenous control (normalizer) at concentrations of 1 and 2 pmol/μl. The forward and reverse *gus* primers and TaqMan *gus* probe were

added at concentrations of 5 pmol/μl each. Reactions were analyzed by using an ABI Prism 7300 sequence detector (7300 System SDS, v1.3.0 software; Applied Biosystems). The reaction conditions were as follows: 2 min at 50°C, 10 min of polymerase activation at 95°C, and then 40 cycles of 95°C for 15 s and 55°C for 60 s. Each assay included (in triplicate) a standard curve of six serial dilutions of cDNA. The baseline was corrected by using the default adaptive baseline algorithm, and the threshold was manually adjusted by visual analysis of log-scale amplification plots.

Linear regression analysis was used to calculate the standard curve, estimate the r^2 and slope, and calculate the PCR efficiency (%). After correcting the cycle threshold values (C_T) for amplification efficiency, the expression levels were normalized to the endogenous control, and relative quantification of gene expression was obtained by the comparative C_T method ($2^{-\Delta\Delta CT}$) (47).

Measurement of β-glucuronidase activity. *S. meliloti* cultures were grown to mid-exponential phase in PY medium. Cells were collected by centrifugation at 6,000 × *g* in a Sorvall SS34 rotor, washed with sterile MM, and then concentrated 100-fold. Portions (50 ml) of MM containing 10 mM succinate and 10 mM ammonium chloride as carbon and nitrogen sources, respectively, were inoculated with the cell suspension at an initial OD_{540} of 0.05. Cultures were grown with shaking (200 rpm) for 8 h at 30°C. Quantitative β-glucuronidase activity was measured in a 1.0-ml culture sample using 4-2-nitrophenyl β-D-glucuronide (pNPG) as a substrate (28). The data were normalized to total cell protein concentration obtained by the Lowry method. The specific activities are expressed as nanomoles of *p*-nitrophenol (pNP) per minute per milligram of protein (nmol/min/mg of protein).

Determination of *N*-acetyl-γ-glutamyl phosphate reductase (ArgC) activity. Bacteria were grown in MM. Cell cultures in logarithmic growth phase (8 h for MM with succinate-ammonium or 24 h for MM with mannose-nitrate) were collected by centrifugation and then washed. The cell pellets were resuspended in 5 mM Tris-HCl buffer (pH 7.5), and then the cells were disrupted by sonic oscillation. The supernatant was treated with ammonium sulfate, and the precipitate was discarded. Material that was precipitated by the addition of more ammonium sulfate was collected and dissolved in 5 mM Tris-HCl buffer (pH 7.4), containing 0.6 mM 2-mercaptoethanol. The protein solution was loaded on a Microcon YM-30 column, and the soluble cell extract was collected and used for the measurement of ArgC. The enzymatic activity was measured spectrophotometrically by the disappearance of NADPH in the presence of *N*-acetyl glutamate and ATP. The assay was initiated by adding ArgB (*N*-acetyl glutamate kinase, purified in our laboratory) to the reaction mix since *N*-acetyl glutamyl phosphate, the product of ArgB and substrate of ArgC, is highly unstable and not commercially available. Reaction rates were corrected for NADPH consumption, measured in the absence of added *N*-acetyl glutamate. The specific activity was expressed as nanomoles of NADP produced per minute per milligram of protein. The protein concentration was determined by the Lowry method.

Determination of SpeB activity. For the measurement of agmatine ureohydrolase (SpeB) activity, 150-ml cultures were grown (either for 8 h in MM supplemented with succinate-ammonium or for 24 h in MM supplemented with mannose-nitrate), and the cells were collected by centrifugation. The pellets were washed and then resuspended in 1.0 ml of ice-cold SpeB reaction buffer. The cells were disrupted, and then the cell debris was removed by centrifugation. The supernatant was collected and used to assay SpeB activity. The reaction was stopped by the addition of a perchloric acid solution. The ammonia content in the sample was measured by using a diagnostic urea/ammonia determination procedure (Boehringer Mannheim). The specific activity was expressed as micrograms of ammonia produced per milligram of protein. The protein concentration was determined by the Lowry method.

ArgC proteomic identification and quantification. Bacterial proteins were obtained by sonicating cell cultures (grown 8 h in MM supplemented with succinate-ammonium) for 5 cycles of 1 min each at 4°C in a Vibra Cell (Soniprep150, MSE Sanyo) in the presence of a protease inhibitor (Complete tablets; Roche Diagnostics GmbH, Mannheim, Germany). To further limit proteolysis, protein isolation was performed by using phenol extraction (35). To solubilize and obtain completely denatured and reduced proteins, pellets were dried and resuspended as previously described (21). Prior to electrophoresis, the samples were mixed with 7 M urea, 2 M thiourea, 4% CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}, 2 mM tributhyl phosphine (TBP), 2% ampholytes, and 60 mM dithiothreitol. Sample preparation, analytical and preparative two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), and image analysis were performed as previously described (20). pH gradients were obtained using a 2D-PAGE standard (Sigma). For separation in the first dimension, ~500 mg of total protein was loaded. The gels were stained with Coomassie Blue R-250, and protein spots were detected at a 127-by-127-mm resolution using a PDI Image Analysis System and PD-Quest software (Protein Databases, Inc., Huntington Station, NY). Selected spots were excised manually and prepared for mass spectroscopy (MS), as previously described (21). Purified ArgC protein was subjected to enzymatic digestion, and then the resulting peptide fragments were analyzed by complementary MS methods (matrix-assisted laser desorption ionization–time of flight–MS and liquid chromatography-electrospray ionization-tandem MS). ArgC quantification was performed by dot pixel evaluation with PD-Quest software and extrapolated to the total gel load. All experiments were performed three times. Mass spectra were obtained using a Bruker Daltonics Autoflex (Bruker Daltonics, Bellerica, MA) operated in the delayed extraction and reflectron mode. Spectra were externally calibrated using a peptide calibration standard (Bruker Daltonics standard 206095). Peak lists of the tryptic peptide masses were generated and searched against the NCBI nr databases or the Rhizobase (http://bacteria.kazusa.or.jp/rhizo/) using the Mascot search program (Matrix Science, Ltd., London United Kingdom).

Determination of excreted organic acids and amino acids. For extracellular organic acids and amino acids determination, cells were grown for 8 h in MM with succinate-ammonium or 24 h for MM with mannose-nitrate, and 25-ml cultures were subjected to centrifugation, and then the supernatant, free of bacteria, was lyophilized. Amino acids and organic acids were separated and quantified by using reversed-phase high-performance liquid chromatography with a Symmetry C18 column (5 μm, 3.9 by 150 mm; Waters). The data are expressed as nmol/mg of protein and μmol/mg of protein, respectively.

Bioinformatics methods. Codon adaptation index (CAI) calculations were performed with the CAI program in the EMBOSS (European Molecular Biology Open Software Suite) package available at www.emboss.org. The species signature was calculated as the number of specific amino acid residues in the encoded protein that were unique or least abundant at each position based on multiple alignments, and was expressed as a percentage of protein length. Species signatures were calculated with a perl custom program (available on request) based on multiple alignments performed with CLUSTAL W (70). For phylogeny, multiple alignment of translated amino acid sequences was carried out by using MUSCLE v3.7 (17) and the approximate-likelihood rate method with PhyML v3.0 (1, 33) on the Phylogeny.fr server (15). A bootstrap was applied with 100 replicates, and a WAG model for amino acid substitution was used. The tree was drawn with TreeView1.6.6 (Roderic Page). ArgC sequences for species signature evaluation and phylogenetic analysis were obtained from GenBank using the following access numbers: *A. tumefaciens* C58 (NP_354256), *A. vitis* S4 (YP_002549205), *Bradyrhizobium* ORS278 (YP_001205564), *B. melitensis* 16 M (NP_540088), *E. coli* K-12 W3110 (AP_003852), *M. loti* MAFF303099 (Q982X3), *R. etli* CFN42 (ABC90369), and *S. meliloti* 1021 (NP_385346). Obtained *argC* sequences were registered in the GenBank and assigned the following provisional access numbers: *M. loti* strains R88B (HM753558), N86A99 (HM753564), VTI (HM753565), CJ4 (HM753561), CJ2 (HM753562), R12C (HM753559), R7A (HM753560), and CJ6 (HM753563). *S. meliloti* strains Ts48 (HM753568), Cx2 (HM753573), Ter3 1 (HM753570), Mac3 5 (HM753572), REF21 (HM753574), Cx48 (HM753571), SauI 5 (HM753569), Cx26 (HM753575), Tx 30 (HM753567), and ES2 (HM753566). *A. tumefaciens* strains IAM12048 (HM753578), IAM13129 (HM753576), IAM13549 (HM753577), and IAM14040 (HM753579). *R. etli* strains Kim5 (HM753582), CIAT652 (HM753583), GR56 (HM753580), CIAT151 (HM753584), and Brasil5 (HM753581).

RESULTS

Structural analysis of *argC* genes and their protein sequences. (i) Species signatures and phylogenetic relationships of *argC* from *Rhizobiales*. Sequence differences among orthologs could be the result of evolutionary adaptation to a specific cellular environment. The extent of these specific changes has been termed the "species signature" (32). In order to assess the functional importance of species signature in *Rhizobiales*, we focused on a syntenic gene, *argC*, which encodes an *N*-acetyl-γ-glutamyl phosphate reductase that is essential for arginine biosynthesis. Alignment and comparison of the amino acid sequences of ArgC from five species—*R. etli*, *A. tumefaciens*, *S. meliloti*, *B. melitensis*, and *M. loti*—showed that there were 133 positions with identical residues, and 32 (10.3%), 30 (9.6%), 30 (9.6%), and 113 (36.4%) residues, respectively, that were unique and are considered species sig-

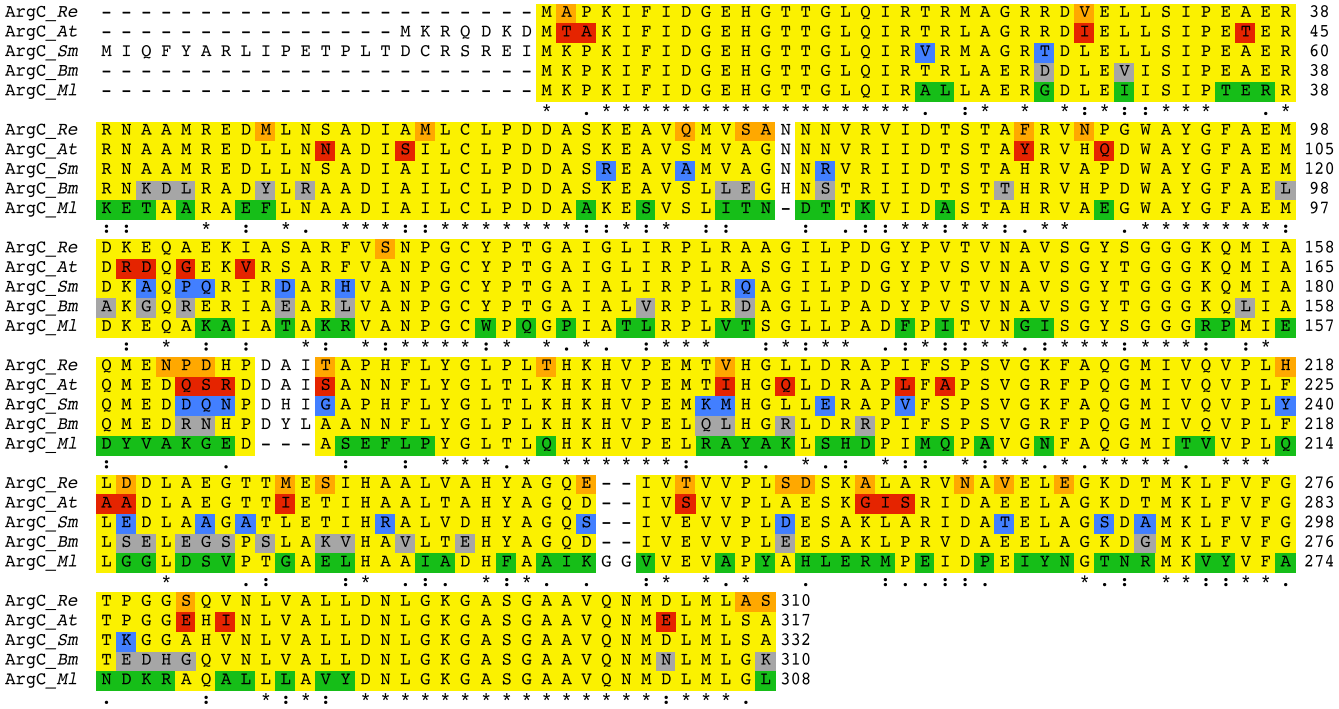


FIG. 1. Species signature of ArgC from *Rhizobiales*. A multiple alignment was carried out using CLUSTAL W. Species signatures were determined with a perl custom script. Yellow blocks indicate identical residues in at least two species. Colored residues for each organism denote species signature as follows: orange, *R. etli*; red, *A. tumefaciens*; blue, *S. meliloti*; gray, *B. melitensis*; and green, *M. loti*. Residues not in blocks are gaps in some organism or nonhomologous sequences.

natures (Fig. 1); the isoelectric points were 5.46, 5.39, 6.0, and 5.4, respectively.

To determine the degree of conservation of the ArgC protein among members of the same species of *Rhizobiales*, the *argC* coding regions of different *Sinorhizobium*, *Rhizobium*, *Agrobacterium*, and *Mesorhizobium* strains were amplified by PCR and then sequenced. The translated sequences showed 98 to 99% similarity among species. A phylogenetic tree was constructed to assess the relationships among *Rhizobiales* species based on the ArgC sequences (Fig. 2). *Rhizobium-Agrobacterium* species exhibited the closest relationships, followed by *Sinorhizobium* strains; *M. loti* sequences were located on a more distant branch. This species relationship was in general concordance with other vertically inherited markers, such as DnaK, GyrB, and AtpD (data not shown).

(ii) Genomic context analysis and identification of the *argC* promoter. The genomic organization of *argC* in *Rhizobiales* was analyzed by using a publicly available annotated database (http://www.ncbi.nlm.nih.gov/genomes/MICROBES). In the circular chromosome in each *Rhizobiales* species analyzed, the *argC* gene are contiguous to upstream *speB* (agmatine ureohydrolase), which participates in the synthesis of putrescine, part of the polyamine synthesis pathway (e.g., spermine and spermidine) involved in several cellular and metabolic functions (37). Previous *in silico* analyses have indicated that *argC* and *speB* are organized in an operon (52, 58).

As an initial approach, the complete putative *speB-argC* operon of each species was cloned and used for complementation analysis of an *S. meliloti argC* mutant (*argC*::ΩSp). The enzymatic activity of ArgC and SpeB expressed from the putative *speB-argC* operons of *S. meliloti*, *A. tumefaciens*, *R. etli*, and *M. loti* was evaluated in the *S. meliloti argC* mutant strain. We found differences in ArgC and SpeB activities (data not shown), which conducted us to consider independent promoters for each gene. To identify the putative *argC* promoter, the *speB-argC* segment-containing plasmids were modified by deletion of 150 bp of *speB* gene using XbaI-KpnI restriction sites (data not shown). Disruption of *speB* did not compromise the ability of the plasmids to complement the mutant phenotype (data not shown).

To verify that there was an independent promoter region for *argC* located within the *speB* coding region, we identified the transcriptional start site of *S. meliloti argC* by 5′RACE using total RNA purified from wild-type strain cells. The *argC* transcriptional start site was located 125 nucleotides upstream of the putative translational start codon, and a putative −10 box characteristic of sigma 70-dependent promoters was identified (see Fig. S1 in the supplemental material).

To determine the transcriptional activity of the putative *argC* promoter of *Rhizobiales*, *argC* promoter-β-glucuronidase (*gus*) gene fusions were constructed, and their activity was evaluated in *S. meliloti* and *M. loti*. As a control, each species-specific promoter-*gus* construct was also evaluated in its native background. As seen in Fig. 3, the activities of the *S. meliloti*, *A. tumefaciens*, and *R. etli argC* promoters were similar in their native backgrounds and in that of *S. meliloti.* Also, the *M. loti argC* promoter had very low activity in the latter background. However, the activities were more induced in the *M. loti* background compared to the *M. loti argC* promoter activity. Thus, the newly identified *argC* promoter was active in all strains, and

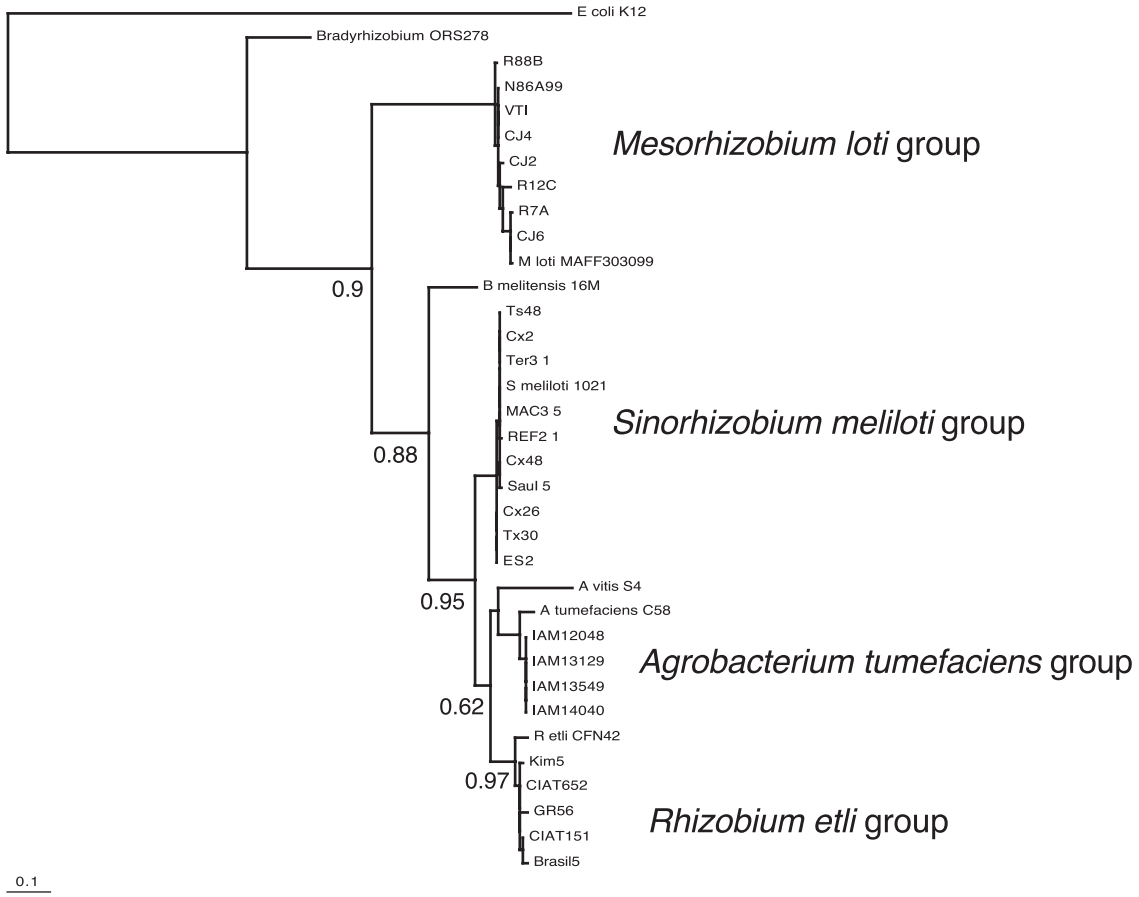


FIG. 2. Phylogenetic analysis of ArgC from *Rhizobiales*. The alignment was done using the MUSCLE program. Phylogeny was generated by the PhyML method. The tree was drawn with TreeView. Bootstrap analysis was performed with 100 replicates. Supporting values are shown in the main branches. The *E. coli* sequence was taken as the outgroup. The bar represents substitutions per residue.

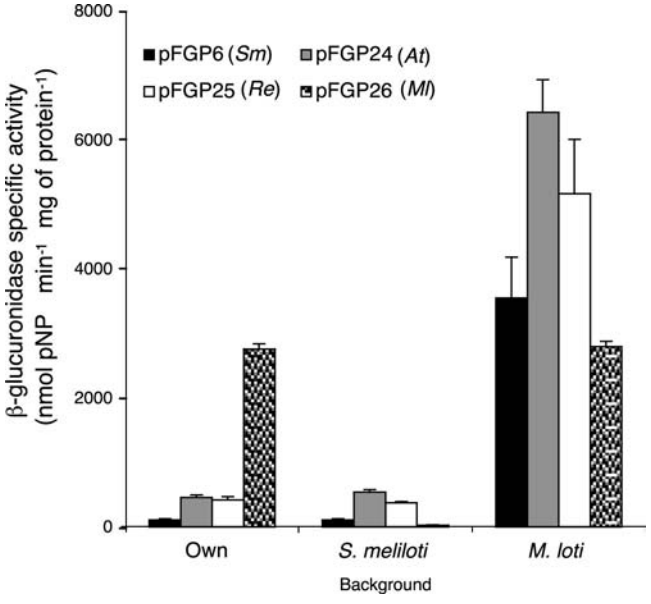


FIG. 3. Transcriptional analysis of *argC* promoter activity. Expression of the indicated *argC* promoter-*gus* (β-glucuronidase) gene fusions is depicted as follows: first group, in the genetic background of the indicated species; second group, in *S. meliloti*; and third group, in *M. loti*. Bars: black, *S. meliloti*; gray, *A. tumefaciens*; empty, *R. etli*; patterned, *M. loti*.

the *M. loti argC* promoter was weakly active in the *S. meliloti* background.

The species signatures of *Rhizobiales* ArgC orthologs ranged from 10 to 36%, and phylogenetic analysis revealed that *R. etli* was closely related to *A. tumefaciens*. Unique to *Rhizobiales*, the *argC* gene is linked to *speB* (involved in putrescine synthesis), and in the case of *S. meliloti* it was demonstrated that *argC* has its own active promoter inside the coding sequence of *speB*. *M. loti argC* promoter activity was very low in the *S. meliloti* background compared to its native background.

Expression analysis of *argC* regulated by different promoters. The aim of the present study was to determine the complementation efficiency of *argC* orthologs from related species in an *S. meliloti argC* mutant. The level of mRNA synthesis is dependent on promoter activity; thus, we were interested in the expression patterns of the Rhizobiales *argC* genes under the control of the same promoter, either the *S. meliloti speB* (p*speB*) promoter, the *S. meliloti argC* (p*argC*) promoter, or the *E. coli lac* (p*lac*) promoter.

The level of expression of β-glucuronidase under the control of the *lac* promoter was three times higher than the *S. meliloti argC* promoter (compare the first and second lines in Table S2 in the supplemental material). In contrast, the *S. meliloti speB* promoter exhibited very low activity under the same conditions

TABLE 2. Effect of *argC* gene sequences on transcriptional expression

Sequence in fusion	Mean β-glucuronidase sp act ± SD[a]			
	S. meliloti	*A. tumefaciens*	*R. etli*	*M. loti*
ATG	233 ± 15	210 ± 28	230 ± 39	249 ± 17
Partial CDS (100 bp)	366 ± 23	284 ± 30	206 ± 25	637 ± 131
Partial CDS (880 bp)	439 ± 51	319 ± 48	270 ± 48	591 ± 103
Complete CDS	145 ± 6	280 ± 30	30 ± 3	153 ± 7

[a] Values are expressed as nmol of *p*-nitrophenol min^{-1} mg of $protein^{-1}$. Data represent the means of two replicates from three independent experiments. The expression under p*argC*$_{Sm}$ was determined.

(26- and 8-fold less than p*lac* or p*argC*, respectively) (third line in Table S2 in the supplemental material). These results indicated that in *S. meliloti*, p*lac* is a strong promoter, whereas p*argC* and p*speB* are moderate and weak promoters, respectively. For a comparative analysis, we used these promoters to drive the expression of *argC* from *A. tumefaciens*, *S. meliloti*, *R. etli*, and *M. loti*. The levels of transcriptional activity from the different promoters were similar to those obtained above (data not shown).

To determine the transcriptional activity through the *argC* sequences, we analyzed the expression of transcriptional fusions either containing the complete *argC* sequence, containing a partial *argC* sequence (either with 880 or 100 bp), or lacking the coding region completely (containing only the ATG codon) under the control of the *S. meliloti argC* promoter. Constructs were analyzed in an *S. meliloti* wild-type strain background. Deletion of *argC* gene portions had variable effects on transcription rates, resulting in increased reporter gene expression in the first 100 to 880 bp of the gene (Table 2). The effect was more drastic with *M. loti argC* (an ~2.5-fold increase compared to the start codon alone), followed by *S. meliloti* (2-fold increase) and *A. tumefaciens* (50% increase). The exception was the *R. etli argC*, whose activity was similar for the truncated sequences and the start codon alone. Interestingly, expression was drastically reduced when the complete *argC* gene sequence was evaluated, indicating sequences with a stalling effect on transcription. The expression of *R. etli argC* was nearly constant with shorter versions, but with the complete gene sequence the effect was very strong. These results suggested that, at least in plasmid, the *argC* coding region itself can influence gene expression, in agreement with other recent results about gene sequences modulating their transcription-translation rates (27, 43). Codon adaptation index (CAI) values were calculated for each *argC* gene in the *S. meliloti* background and yielded values ranging from 0.76 to 0.8.

Plasmid copy number also can play an effect in expression of genes; therefore, we assessed *argC* gene copy using plasmid with p*speB* promoter and chromosomal *rpoA* as a control. The control plasmid (vector alone) was present in cells at approximately two plasmid copies per chromosome (1.8 ± 0.17), as were the *R. etli* (2.1 ± 0.18), *A. tumefaciens* (2.3 ± 0.1), and *S. meliloti* (2.9 ± 0.11) gene-containing plasmids. In contrast, the *M. loti* gene-containing plasmid was present at 9.2 ± 0.23 plasmid copies per chromosome.

Based on *gus* reporter gene expression analysis, we were able to characterize three promoters of various strengths: weak (*S. meliloti* p*speB*), intermediate (*S. meliloti* p*argC*), and strong (*E. coli* p*lac*). Differential transcriptional regulation of complete and partial gene sequences suggested that *argC* coding sequences may modulate their transcriptional rates. An additional mechanism exists that modulated the plasmid copy number in the case of *M. loti argC* ortholog complementation.

Complementation analysis of *argC* from *Rhizobiales* under different promoters. (i) Physiological analysis: growth, duplication times, and excretion of metabolites. The optimal growth conditions for *S. meliloti* wild-type strain was MM containing succinate-ammonium, and typical doubling times were 3 h (Fig. 4). We also investigated suboptimal growth conditions in order to differentiate complemented strains. In MM containing mannose-nitrate, the wild-type strain exhibited delayed growth, with a doubling time of 6 h. Under either set of conditions, the *argC* mutant strain failed to grow. The best complementation was achieved with *argC* under the control of p*argC* (Fig. 4A and B). Under that promoter in both MM growth conditions, all complemented strains grew at similar rates and to levels similar to that of the wild-type strain, and similar duplication times were obtained. Under the control of p*lac*, under both sets of conditions, the growth of all complemented strains was similar to that of the *S. meliloti* wild-type strain (Fig. 4C and D). The *A. tumefaciens*-complemented strain exhibited delayed growth, with increased duplication times of 4 and 8 h in succinate-ammonium and mannose-nitrate, respectively. Under the control of p*speB* (from *S. meliloti*), the complemented strains exhibited reduced growth under both sets of conditions. The *R. etli*- and *A. tumefaciens*-complemented strains exhibited reduced growth, increased duplication times (8 h), and long lag phases (Fig. 4E and F); the *M. loti*-complemented strain failed to grow. Only *S. meliloti*-complemented strain growth was similar to that of the wild type.

Given the limited growth of the complemented strains when *argC* was placed under the control of p*speB* (Fig. 4E and F), we investigated the concentration of metabolites excreted into the culture medium, possibly as a response to a faulty metabolism. In MM containing succinate-ammonium, the *argC* mutant strain excreted 6 to 15 times more glutamate and glycine compared to the wild-type strain, 15 times more α-ketoglutarate, and consumed only a fraction of the succinate consumed by the wild-type strain (Table 3). These levels were restored to wild-type levels when the mutant strain was grown in MM supplemented with arginine. The *S. meliloti*-complemented strain exhibited the most wild-type-like metabolic profile, while the *R. etli*- and *A. tumefaciens*-complemented strains excreted slightly more α-ketoglutarate and consumed a little less succinate, than the rest of complemented strains (Table 3). The *M. loti*-complemented strain showed the highest levels of excreted metabolites, even compared to the *argC* mutant strain (twice the level of glutamate and more than 50% more α-ketoglutarate compared to the *argC* mutant). In MM containing mannose-nitrate, higher glutamate and glycine levels were observed for the mutant strain, compared to MM containing succinate-ammonium, as well as a reduction in the excreted levels with the addition of arginine. The complemented strains showed differential behavior compared to MM with succinate-ammonium. The *S. meliloti*- and *R. etli*-complemented strains showed 10-fold reduction of glutamate excretion, while the *A. tumefaciens*- and *M. loti*-complemented strains showed a slight increase. For glycine excretion, a severalfold increase was

FIG. 4. Se incluye figura sin desviaciones estandár, para distinguir cada símbolo correspondiente a las cepas.

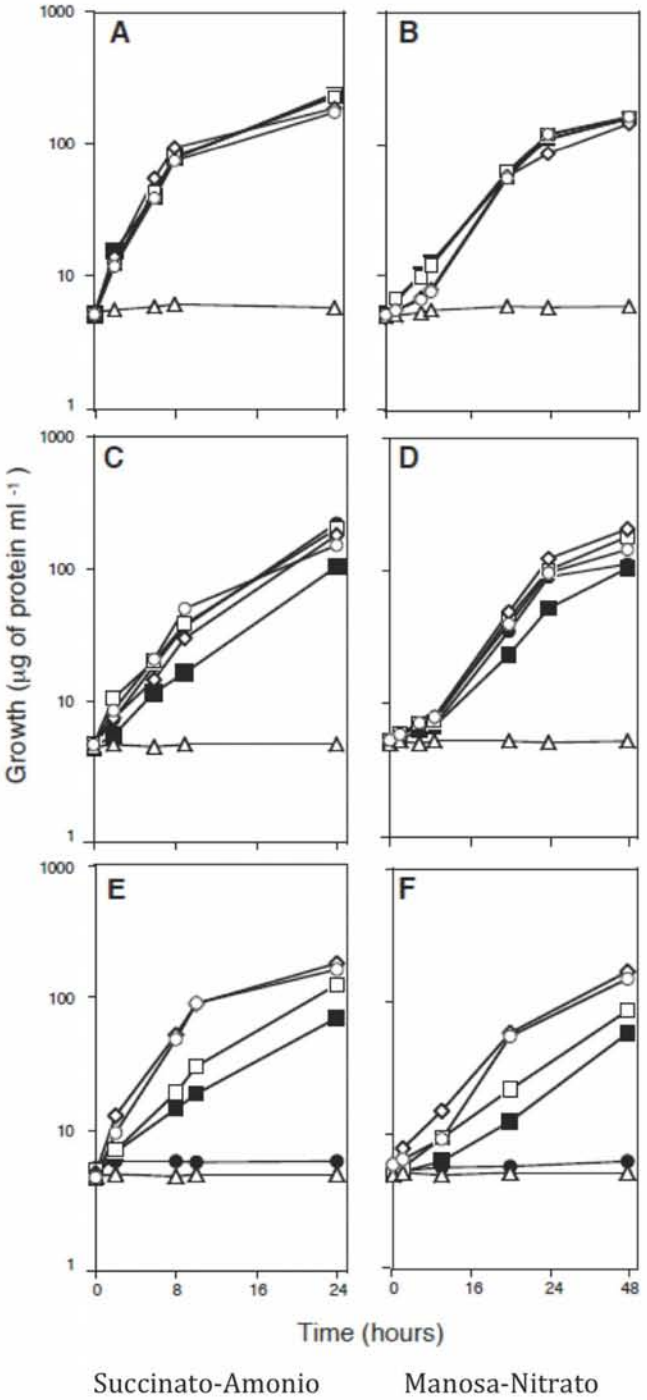


FIG. 4. Growth curves of *S. meliloti argC*-complemented strains. The left panels depict growth in MM supplemented with 10 mM succinic acid and 10 mM ammonium chloride. The right panels depict growth in MM supplemented with 10 mM mannose, 10 mM potassium nitrate. (A and B) *argC* under the control of the *S. meliloti argC* promoter; (C and D) *argC* under the control of the *E. coli lac* promoter; (E and F) *argC* under the control of the *S. meliloti speB* promoter. Each panel represents the *argC* mutant strain complemented with *argC* gene from *S. meliloti* (◆), *A. tumefaciens* (■), *R. etli* (□), and *M. loti* (●); the *S. meliloti* wild-type strain (○); and the *argC* mutant strain (▲).

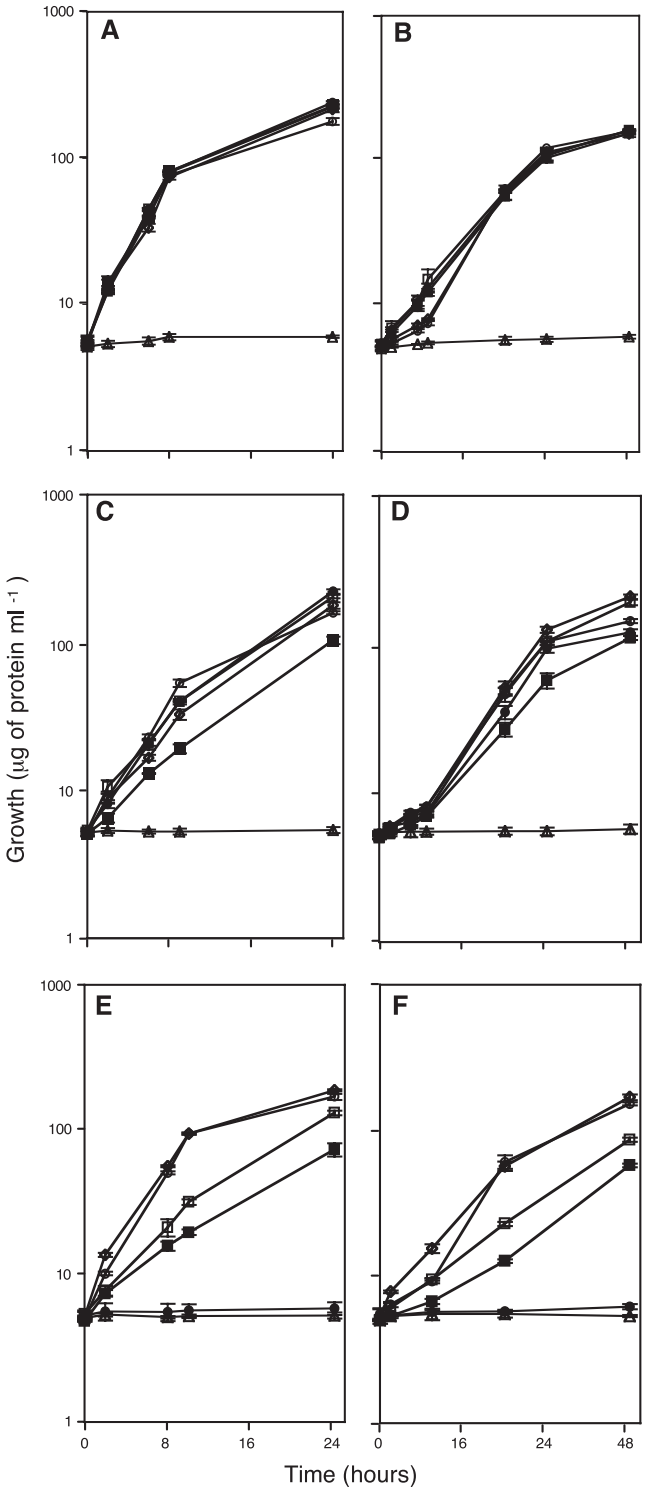


FIG. 4. Growth curves of *S. meliloti argC*-complemented strains. The left panels depict growth in MM supplemented with 10 mM succinic acid and 10 mM ammonium chloride. The right panels depict growth in MM supplemented with 10 mM mannose, 10 mM potassium nitrate. (A and B) *argC* under the control of the *S. meliloti argC* promoter; (C and D) *argC* under the control of the *E. coli lac* promoter; (E and F) *argC* under the control of the *S. meliloti speB* promoter. Each panel represents the *argC* mutant strain complemented with *argC* gene from *S. meliloti* (◆), *A. tumefaciens* (■), *R. etli* (□), and *M. loti* (●); the *S. meliloti* wild-type strain (○); and the *argC* mutant strain (▲).

observed for all of the complemented strains and was drastic with the *M. loti* gene. In the case of α-ketoglutarate, there was a slight reduction for the *S. meliloti* gene and the reduction was approximately 50% in the rest of complemented strains.

These results clearly demonstrated that the *argC* promoters and gene sequences were important determinants of growth capacity in the complemented strains. The newly identified *argC* promoter was optimum for all strains, followed by the strong, constitutive *lac* promoter. There were clear differences among the phenotypes of strains complemented with *A. tumefaciens*, *R. etli*, and *M. loti argC* under the control of p*speB*, none of which fully complemented the *argC* mutant phenotype.

(ii) Biochemical characterization of complemented strains: ArgC and SpeB enzymatic activity. We evaluated the *N*-acetyl-γ-glutamyl phosphate reductase (ArgC) of strains containing plasmids with promoter-*argC* constructions and compared the results to the growth phenotype. The wild-type strain showed reduced activity and, compared to the *S. meliloti*-complemented strain, the multicopy effect was apparent again, in diverse proportions, with the exception of construct under p*speB* promoter (Table 4). Complemented strains grown in MM supplemented with mannose-nitrate showed a reduction in ArgC activity in almost all cases, except for *S. meliloti* under p*lac* and *A. tumefaciens* in the *speB-argC* construct. The tendency toward reduced activity was possibly because these compounds are poor carbon and nitrogen sources for *S. meliloti*.

Under the control of p*argC*, in MM containing succinate-ammonium, the *S. meliloti*-complemented strain had the highest enzymatic activity, and the *M. loti*-complemented strain had the lowest. In MM containing mannose-nitrate, similar proportions among the strains were observed (Table 4). These activity levels were sufficient for all of the strains to achieve optimal growth (see Fig. 4A and B).

Under both growth conditions, p*lac* conferred high activity levels, most notably in the *S. meliloti*-complemented strain (Table 4). In this case, the growth delay of the *A. tumefaciens*-complemented strain did not correlate with diminished activity. Under the control of p*speB*, there was reduced ArgC activity in all strains (Table 4). In the *M. loti*-complemented strain, this level of activity was insufficient to support growth (see Fig. 4E and F). In addition, although both *R. etli*- and *A. tumefaciens*-complemented strains exhibited similar activity levels in terms of complementation of the *argC* mutant phenotype in MM containing succinate-ammonium compared to the *S. meliloti*-complemented strain, they did not grow well and had long lag phases in MM containing mannose-nitrate.

Strains containing the construct *speB-argC* (with each gene expressed from its own promoter) showed similar growth compared to wild-type (not shown) and intermediate ArgC activity levels (Table 4)—with the exception of the *M. loti*-complemented strain, which did not grow in both MM conditions and showed very reduced ArgC activity (Table 4). The activity of SpeB was also evaluated for strains with the *speB-argC* construct, and similar levels of activity were found among the complemented strains in both growth conditions (Table 4).

Plasmids with *S. meliloti* and *M. loti argC* genes also were introduced into the *S. meliloti* wild-type strain, and no differences in ArgC activity were observed in MM supplemented with succinate-ammonium, with the exception of *M. loti*-complemented strain under p*speB*, that presented one-third reduc-

TABLE 3. Metabolite excretion of *S. meliloti argC* complemented strains under the *S. meliloti* p*speB* promoter

Strain	Metabolite excretion (mean ± SD)[a]						
	MM succinate-ammonium (8 h)				MM mannose-nitrate (24 h)		
	Glu*	Gly*	α-Kg†	Succ†	Glu*	Gly*	α-Kg†
S. meliloti							
1021 wild type	1.6 ± 0.7	0.1 ± 0.1	13.6 ± 1.0	226.6 ± 25.2	0.1 ± 0.2	12.6 ± 2.0	10.4 ± 0.8
argC mutant	7.1 ± 3.3	1.5 ± 0.1	228.6 ± 74.0	3356.6 ± 1201.2	24.6 ± 0.5	396.4 ± 78.2	129.2 ± 36.3
argC mutant (+1 mM Arg)	0.2 ± 0.1	0.4 ± 0.2	11.7 ± 0.7	176.8 ± 1.2	0.4 ± 0.3	3.6 ± 0.8	9.2 ± 4.8
argC mutant complemented strains							
pFGP16 (*S. meliloti*)	1.0 ± 0.2	0.8 ± 0.1	13.3 ± 0.9	235.6 ± 23.0	0.1 ± 0.5	54.4 ± 13.0	11.3 ± 1.2
pFGP17 (*A. tumefaciens*)	6.5 ± 0.3	0.6 ± 0.1	68.5 ± 3.9	1246.0 ± 93.9	11.4 ± 2.4	183.4 ± 74.9	40.6 ± 4.6
pFGP18 (*R. etli*)	1.1 ± 0.3	0.5 ± 0.1	53.7 ± 3.8	854.8 ± 38.7	0.1 ± 0.2	168.2 ± 17.9	19.7 ± 2.7
pFGP19 (*M. loti*)	15.9 ± 2.4	1.6 ± 0.4	328.4 ± 11.6	5217.9 ± 1841.0	18.4 ± 0.3	898.0 ± 10.6	151.6 ± 8.4

[a] Glu, glutamate; Gly, glycine; α-Kg, α-ketoglutarate; Succ, succinate. *, Expressed as nmol mg of protein^{-1}; †, expressed as μmol mg of protein^{-1}.

tion in the activity (not shown). This was possibly due to the formation of slightly inefficient ArgC hybrid multimers that reduced native activity.

With the exception of strains carrying the p*speB* promoter in MM containing succinate-ammonium, there was a good agreement between growth curves and ArgC activity. Optimal growth and high ArgC activity were obtained with p*argC* and p*lac*. In the case of p*speB*, the *M. loti*-complemented strain showed the lowest complementation efficiency and a no-growth phenotype. *R. etli*- and *A. tumefaciens*-complemented strains under the p*speB* promoter exhibited similar enzymatic activities compared to the *S. meliloti*-complemented strain; however, they showed severe growth delay.

(iii) Molecular analysis: *argC* transcript levels, direct ArgC quantification, and translational index. Given the results of the analysis of ArgC activity and growth phenotype, we next investigated important factors influencing the ability of orthologous *argC* genes to complement the *S. meliloti argC* mutant phenotype, such as mRNA production and translational efficiency in strains containing the pBBMCS53 vector, carrying the *argC* and *gus* genes, which permitted the complementation and monitoring of transcription simultaneously (see Materials and Methods). The mRNA level in the wild-type strain was used as the reference and given a value of 1.0. The ArgC protein was quantified directly by two-dimensional gel electrophoresis and mass spectrometry. The mRNA levels and the amount of protein detected by proteomic identification were used to calculate a translational efficiency index (proteomic protein/mRNA).

The wild-type strain exhibited nearly constant values for mRNA, proteomic protein, and translational index in each set of assays (Table 5). In general, the *S. meliloti*-complemented

TABLE 4. ArgC and SpeB specific activities of *S. meliloti argC* complemented strains under the control of different promoters

Protein and conditions	Mean ArgC and SpeB sp act ± SD[a] of *argC* mutant complemented strains[a]				
	WT	*S. meliloti*	*A. tumefaciens*	*R. etli*	*M. loti*
ArgC					
p*argC* (pFGP12 to pFGP15)					
Succinate-ammonium	16.6 ± 0.4	134.0 ± 2.6	80.1 ± 2.3	83.1 ± 1.4	13.3 ± 0.8
Mannose-nitrate	16.3 ± 0.4	68.8 ± 0.9	28.7 ± 1.4	32.3 ± 1.9	7.1 ± 0.3
p*lac* (pFGP8 to pFGP11)					
Succinate-ammonium	16.7 ± 0.3	405.4 ± 3.4	49.0 ± 1.5	59.5 ± 0.9	60.6 ± 0.7
Mannose-nitrate	16.4 ± 0.1	432.4 ± 20.7	81.7 ± 2.8	32.1 ± 3.1	81.9 ± 8.7
p*speB* (pFGP16 to pFGP19)					
Succinate-ammonium	16.7 ± 0.3	20.8 ± 0.4	19.3 ± 0.5	18.9 ± 1.3	0.8 ± 0.2
Mannose-nitrate	16.1 ± 0.3	25.0 ± 0.1	16.6 ± 0.1	17.7 ± 0.1	0.9 ± 0.1
speB-argC (pFGP1 to pFGP4)[b]					
Succinate-ammonium	14.5 ± 0.3	102.6 ± 7.4	107.5 ± 3.9	38.0 ± 6.6	0.8 ± 0.2
Mannose-nitrate	16.2 ± 0.1	57.3 ± 10.3	115.4 ± 1.7	35.2 ± 0.8	1.0 ± 0.1
SpeB[c]					
speB-argC (pFGP1 to pFGP4)					
Succinate-ammonium	6.5 ± 0.1	39.9 ± 6.1	53.2 ± 2.3	55.1 ± 1.9	62.2 ± 1.2
Mannose-nitrate	9.2 ± 0.1	46.6 ± 3.4	43.7 ± 3.4	38.9 ± 1.6	21.5 ± 6.6

[a] Expressed as nmol of NADP min^{-1} mg of protein^{-1}, except as noted (see footnote *c*). WT, *S. meliloti* 1021 wild-type strain.
[b] That is, constructs containing *speB* and *argC* under their own promoters.
[c] Expressed as μg of ammonia mg of protein^{-1}.

TABLE 5. mRNA, protein quantification, and translational efficiency index for *S. meliloti argC* mutant complemented strains

Promoter and parameter[a]	*argC* mutant complemented strains[b]				
	WT	*S. meliloti*	*A. tumefaciens*	*R. etli*	*M. loti*
p*argC* (pFGP12 to pFGP15)					
mRNA	1	9.4 ± 0.5	51.1 ± 4.8	127.6 ± 22.4	19.9 ± 0.4
Proteomic protein	9.8 ± 0.5	771.2 ± 91.1	442.7 ± 20.8	331.6 ± 29.9	273.6 ± 58.8
TEI	9.8	82.0	8.7	2.6	13.8
p*lac* (pFGP8 to pFGP11)					
mRNA	1	121.4 ± 11.8	1,759.2 ± 68.2	1,596.7 ± 69.7	1,598.7 ± 42.9
Proteomic protein	9.5 ± 0.1	1,834.2 ± 100.3	1,399.5 ± 213.2	91.7 ± 26.9	1,696.1 ± 303.1
TEI	9.5	15.1	0.8	0.1	1.1
p*speB* (pFGP16 to pFGP19)					
mRNA	1	4.9 ± 0.6	53.1 ± 2.9	35.0 ± 1.1	162.9 ± 9.5
Proteomic protein	9.5 ± 0.1	25.1 ± 0.1	65.5 ± 2.8	72.9 ± 7.9	67.4 ± 5.2
TEI	9.5	5.1	1.2	2.1	0.4

[a] mRNA values were normalized relative to the abundance of mRNA in the wild-type strain. Proteomic protein was quantified by using spectrometry and dot pixels. TEI, translational efficiency index (proteomic protein/mRNA).
[b] WT, *S. meliloti* 1021 wild-type strain. Cells were grown for 8 h in MM supplemented with 10 mM succinic acid and 10 mM ammonium chloride.

strain obtained the best translational efficiency index values in each condition compared to the rest of the strains. Also, despite the fact that *argC* orthologs were controlled by the same promoter in each set, different amounts of mRNA and protein were produced, confirming the findings regarding the transcriptional effect of *argC* coding sequences detected with *gus* fusions (see Table 2).

In the case of expression driven by p*argC* promoter, the *S. meliloti*-complemented strain had 10 and 78 times more mRNA and ArgC protein production, respectively, than the wild-type strain, and an 8-fold increase in translational index (Table 5). The levels of mRNA in the other complemented strains were in the range of 20- to 140-fold higher than in the wild type, and yet the translational index in these strains was lower, with the *R. etli*-complemented strain having the lowest translational index. Despite these variations, all strains exhibited similar growth efficiency.

When strains were controlled by p*lac* promoter they exhibited highly elevated levels of mRNA. In the *S. meliloti*-complemented strain, there was an almost 110-fold increase in mRNA compared to the wild-type strain; mRNA levels in the other complemented strains increased more than 1,500-fold (Table 5). A similar trend in high protein levels was observed, with the exception of the *R. etli*-complemented strain, which also had the lowest translational index in this set. The *S. meliloti*-complemented strain demonstrated the highest protein level, but because its mRNA was so much lower, it had the highest translational index.

In the last set, strains with *argC* expression under the p*speB* promoter exhibited the lowest values of mRNA and proteomic protein production. The complemented strains showed mRNA increases of 5-fold (*S. meliloti*), 7-fold (*R. etli*), 11-fold (*A. tumefaciens*), and 30-fold (*M. loti*) compared to the wild-type strain. However, the last three strains showed similar levels in proteomic protein (a 7-fold increase compared to the wild type) and very low translational index values.

Another complete set of constructs was obtained with pBBR1MCS3 and *argC* alone (without *gus* reporter). Similar values and proportions among the strains were found in comparison to the set of strains described above (data not shown). However, strains with pBBR1MCS3 showed slightly increased duplication times, and some reduced ArgC activity (not shown). It is possible that *gus* fused to *argC* genes produced some type of stabilization.

The described molecular parameters, together with the translation/transcription index, provided relevant information about the functional performance of the complemented strains, which demonstrated that complementation of ArgC with the *S. meliloti argC* gene in its own background showed the highest translational index (Table 5), followed by those with the *R. etli* and *A. tumefaciens* orthologs.

DISCUSSION

The species signature was intended to extract the amount of particular specific amino acid residues in positions of multiple alignments of the orthologous protein sequences (32). In this context, the species signature denotes the change in physicochemical properties of the proteins and displays certain particular characteristics of the orthologous products. It is relevant because the sequence differences of the orthologs are translated into amino acid residues; we wanted to establish whether these differences can be studied by ortholog complementation and functional performance when compared in the same cellular background and related to wild-type complementation and wild-type strain.

We studied the *argC* gene because it forms part of an importantly compromised pathway that is central to amino acid and protein synthesis; also, arginine has the highest nitrogen content among the amino acids, its synthesis demands a huge amount of energy, and it is a precursor for polyamine synthesis (14, 48, 63, 68). The *S. meliloti argC* gene was mutated in the chromosome, and then the complementation was obtained either with its own gene or with the ortholog from closely *Rhizobiales* species; they were cloned on the same vector and expressed under the same promoter to study the ability to recover the arginine prototrophy.

Three promoters were used in order to achieve diverse ex-

pression degrees, and we found strong (p*lac*), medium (p*argC*), and weak (p*speB*) transcriptional levels (see Table S2 in the supplemental material). That from *speB* was used because the gene is upstream to *argC* in *Rhizobiales* and participates in polyamine synthesis (e.g., arginine utilization), which is important in diverse metabolic processes but not so abundant as amino acids (68). With these promoters we searched expression levels similar to those of the wild-type strain. In the case of *argC* complementation under the p*speB* promoter, we obtained the lowest ArgC specific activity and major differences, in regard to the optimal maintenance of cellular growth: the gene sequences from *R. etli* and *A. tumefaciens* showed reduced growth ability and that from *M. loti* showed no growth (Fig. 4E and F).

Interestingly, diverse factors altered the expression of the *argC* orthologous sequences and their complementation ability in several levels. In the first level, there was a differential multicopy effect due to plasmid vectors, because under the *speB* promoter we found two to three plasmid copies with *S. meliloti*, *R. etli*, and *A. tumefaciens* sequences, but nine copies with the *M. loti* gene. The most divergent sequence from *S. meliloti* was precisely that from *M. loti*, and the complemented strain carrying this gene presented marked inability to recover arginine synthesis and restore cell growth (Fig. 4E and F). The increased plasmid copy number was possibly related to a gene dosage compensation to favor abundance of the gene, as observed in other bacteria (67) and eukaryotes (18). Lind et al. (46), in a recent report dealing with orthologous replacement of ribosomal proteins in *Salmonella enterica* serovar Typhimurium, found that gene amplification is a compensatory mechanism to restore fitness decay. In our results, despite increase in gene copy number, growth was not fully recovered. Lind et al. showed that gene amplification is important for horizontally transferred genes and the evolution of new functions from gene duplication.

In the next level—transcription of the orthologous genes—also was affected. Theoretically, different genetic sequences located downstream of the same promoter will have the same level of transcriptional activity. However, the *argC* coding sequence itself appeared to modulate transcription to a certain extent (Tables 2 and 5). This may be due to codon usage (but in this case similar CAI values were obtained, and thus codon usage apparently did not play a role in complementation efficiency), codon distribution (8), interaction of translation factors with RNA polymerase and ribosome (7, 59), and the formation of mRNA secondary structures that can alter the speed of the transcriptional complex. Other authors have made the same observation, at the level of transcription-translation processes (43). In addition, as reported recently, important transcription modulation occurs in the first and last 100 nucleotides of the genes (71). A similar effect was clearly observed in transcriptional fusions with segments of genes (Table 2) and also when expression was directly quantified by RT-PCR (Table 5), because under each of the three promoters different amounts of synthesized mRNA were found for each *argC* gene.

The previously described effects on transcription-translation processes, altogether with coding sequence differences of the orthologs (the species signature), consequently affected enzymatic activity of ArgC (Table 4), having a physiological repercussion on growth rate (Fig. 4) and metabolite excretion to the culture medium (Table 3). The diminished growth rate is related to intermediary metabolites that cannot be completely assimilated and are excreted in huge amounts (19). To perform an in-depth analysis of what was occurring in the complementation assay, we decided to measure molecular parameters such as direct mRNA and protein quantification (Table 5) to obtain a translational efficiency index; these parameters helped to clarify the transcriptional-translational ability of *argC* genes cloned on plasmids.

The *S. meliloti* wild-type strain showed the best translational efficiency, with almost constant values; this reproducibility supported the reliability of our approach (Table 5). In the complementation with the *S. meliloti* gene under the p*speB* promoter, the copy number effect was evident, since a 5-fold increase in the mRNA amount was found, compared to the wild-type strain. However, this value did not correlate directly with the copy number found (i.e., 2.9), so possibly something else influenced the transcription rate, such as the supercoiling state of the replicon (24).

Under the p*lac* promoter, the highest amounts of transcripts were obtained, but not the maximum translational efficiency indexes. It was apparent that there is a translational limit in the cell. In this case, the *R. etli*-complemented strain obtained the lowest index, however, it was sufficient to promote an adequate growth recovery (Fig. 4C and D). Under the p*speB* promoter, despite the fact that the *M. loti*-complemented strain showed more gene copies, that amount was not enough to sustain cell growth, possibly because cells had a low translation efficiency index (Table 5 and Fig. 4E and F). It is important to note the lack of correlation between protein synthesis and enzymatic activity, which was very marked in the case of p*lac* constructs. We believe that there exists a physiological limitation to reaching high levels of enzymatic activity, a limitation that was exerted mainly in the case of *Rhizobiales* complementing sequences (because the *S. meliloti* sequence indeed showed a high level of ArgC activity under p*lac*). Several other enzymatic and physiologic factors may participate and alter the correlation between transcription and translation activity (such as stoichiometry, cofactors, regulators, enzymes of the pathway, energy flux, etc.). Another consideration is that activity was measured *in vitro* and that this does not completely reflect the physiologic alteration of *argC* ortholog expression. We thus consider that cellular growth was a more faithful measure of what occurred with ortholog complementation.

The ArgC proteins from *S. meliloti*, *R. etli*, and *A. tumefaciens* displayed similar amounts of species signature differences (ca. 10%), which was indicative of their phylogenetic proximity. Sequences from *B. melitensis* and *M. loti* were evolutionarily more distant (Fig. 2), and the species signatures were higher (36% for *M. loti*) (Fig. 1). We sequenced the *argC* gene from several strains of each species (with the exception of *Brucella melitensis*), and the phylogenetic relationships between these sequences helped to define clear species clusters (Fig. 2). The genomic organization of *argC* genes in *Rhizobiales* was very unusual, since this gene was closely linked to *speB* (from the polyamine synthesis pathway). Possibly, these genes have independent promoters and mechanisms of transcriptional regulation, with the ability to respond appropriately and effectively to particular regulatory signals.

Transcriptional activation by species-specific *argC* promot-

ers was also characterized in the native background and in *S. meliloti* and *M. loti*. In the native backgrounds, the lowest activity was seen in *S. meliloti*, the highest was seen in *M. loti*, and the other species were intermediate (Fig. 3). Interestingly, the *argC* promoter from *M. loti* was not active in *S. meliloti*; in contrast, the *argC* promoters of all of the strains were highly active in *M. loti* (Fig. 3). A partial growth defect was also apparent in the *A. tumefaciens*- and *R. etli*-complemented strains when the expression of *argC* was driven by the *speB* promoter (Fig. 4E and F). This partially defective growth phenotype correlated with a slight increase of α-ketoglutarate excretion to the culture medium and less utilization of succinate in comparison with the wild-type strain (Table 3). A similar phenotype was observed for a GOGAT mutation in *R. etli* (10).

Orthologous genes encode the same function (although some differences have been described for the HoxA3 ortholog in mice and zebrafish) (11), but they may also contain sequence characteristics that reflect specific adaptations. The specific sequence differences in a gene alter the physicochemical characteristics of the encoded proteins and may affect protein function. Factors such as transcription and translation rates, enzyme kinetics, transcriptional regulation, intracellular environment, availability of substrates, and interacting metabolites, proteins, and ions could also affect performance (30, 39, 55, 72).

Species adaptation is a slow process. Neutral drift (point mutations) and natural selection produce numerous changes in a gradual manner, each change having a small effect on fitness. However, some changes can have vast effects on an organism and tend to outcompete small-effect mutations. It has been generally accepted that any attempt to discern these changes experimentally is futile (12). Despite the fact that "species" differentiation in bacteria is considered imprecise, and its existence is even debatable (26, 44), homologous recombination and point mutations can gradually convert a homogeneous bacterial population into different, nonrecombinable populations. In this way, small (neutral and almost neutral) and large (non-neutral) genetic changes could eventually result in a new species. It remains to be seen whether the elemental mechanisms for adaptation, considered microevolution, can reconstruct the higher processes of macroevolution, including speciation and the emergence of biodiversity (12). The appearance of a new phenotype, such as citrate utilization in the presence of oxygen by an *E. coli* population in Lenski's long-term experiment, is something close to speciation (2, 5). However, even random variation (bet hedging) can enhance long-term fitness by increasing the likelihood of individuals to express an adaptive phenotype (3). There is a general lack of knowledge of how ortholog replacement may affect cell functioning and fitness. In this regard, the present study is a valid approach to sorting out the diverse consequences of expression in one species of orthologous sequences from other related species, in which specific amino acid changes may have been derived through adaptation.

In the present study we have demonstrated that an in-deep complementation analysis can uncover unusual complexity. We evaluated the ability of ArgC orthologs from closely related Rhizobiales species containing specific sequence differences to physiologically complement the *S. meliloti argC* mutant phenotype. We observed many factors participating in the complementation efficiency of each ortholog. In addition, we identified a novel promoter for *argC* in Rhizobiales, the presence of which was not predicted by existing genomic information. In agreement with recent reports (27, 43), we demonstrated that coding sequences participate in the regulation of transcription and translational efficiency, perhaps modulating mRNA stability through the formation of secondary structures. Altogether, these are important findings that open new research perspectives.

ACKNOWLEDGMENTS

This study was partially supported by the Dirección General de Asuntos del Personal Académico (DGAPA-Universidad Nacional Autónoma de México), grants IN215307 and IN212710, and by the Ph.D. Program in Biomedical Sciences-UNAM.

We thank Mario Alberto Flores for help in the construction of the *S. meliloti argC* mutant; Michael Dunn for ArgB purification; Sandra Contreras, Miguel Elizalde, Magdalena Hernández, and Gabriel Martínez-Batallar for metabolite and proteome determinations; Gabriela Guerrero and Alejandro Aguilar for preparation of graphs; Ricarda Rivero, Yunuen Acevedo, and Alma Reyes (CCG-UNAM) for technical assistance; and Claudia Silva (IBt-UNAM) and Clive Ronson (Otago University, Otago, New Zealand) for the donation of strains.

REFERENCES

1. **Anisimova, M., and O. A. Gascuel.** 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. Syst. Biol. **55:**539–552.
2. **Barrick, J. E., S. Y. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim.** 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461:**1243–1247.
3. **Beaumont, H. J. E., J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey.** 2010. Experimental evolution of bet hedging. Nature **462:**90–93.
4. **Better, M., and D. R. Helinski.** 1983. Isolation and characterization of the *recA* gene of *Rhizobium meliloti*. J. Bacteriol. **155:**311–316.
5. **Blount, Z. D., C. Z. Borland, and R. E. Lenski.** 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **105:**7899–7906.
6. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:**459–472.
7. **Burmann, B. M., K. Schweimer, X. Luo, M. C. Wahl, B. L. Stitt, M. E. Gottesman, and P. Rösch.** 2010. A NusE:NusG complex links transcription and translation. Science **328:**501–504.
8. **Cannarozzi, G., N. N. Schraudolph, M. Faty, P. von Rohr, M. T. Friberg, A. C. Roth, P. Gonnet, G. Gonnet, and Y. Barral.** 2010. A role for codon order in translation dynamics. Cell **141:**355–367.
9. **Carvalho, F. M., R. C. Souza, F. G. Barcellos, M. Hungria, and T. R. Vasconcelos.** 2010. Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order *Rhizobiales*. BMC Microbiol. **10:**37.
10. **Castillo, A., H. Taboada, A. Mendoza, B. Valderrama, S. Encarnación, and J. Mora.** 2000. Role of GOGAT in carbon and nitrogen partitioning in *Rhizobium etli*. Microbiology **146:**1627–1637.
11. **Chen, L., P. Zhao, L. Wells, C. T. Amemiya, B. G. Condie, and N. R. Manley.** 2010. Mouse and zebrafish Hoxa3 orthologues have nonequivalent in vivo protein function. Proc. Natl. Acad. Sci. U. S. A. **107:**10555–10560.
12. **Chouard, T.** 2010. Revenge of the hopeful monster. Nature **463:**864–867.
13. **Crossman, L. C., S. Castillo-Ramírez, C. McAnnula, L. Lozano, G. S. Vernikos, J. L. Acosta, Z. F. Ghazoui, I. Hernández-González, G. Meakin, A. W. Walker, M. F. Hynes, J. P. Young, J. A. Downie, D. Romero, A. W. Johnston, G. Dávila, J. Parkhill, and V. González.** 2008. A common genomic framework for a diverse assembly of plasmids in the symbiotic nitrogen fixing bacteria. PloS One **3:**e2567.
14. **Cunin, R., N. Glansdorff, A. Piérard, and V. Stalon.** 1986. Biosynthesis and metabolism of arginine in bacteria. Microbiol. Rev. **50:**314–352.
15. **Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.-F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.-M. Claverie, and O. Gascuel.** 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36:**W465–469.
16. **Dunn, M. F., G. Araíza, S. Encarnación, M. C. Vargas, and J. Mora.** 2002. Effect of *aniA* (carbon flux regulator) and *phaC* (poly-β-hydroxybutyrate synthase) mutations on pyruvate metabolism in *Rhizobium etli*. J. Bacteriol. **184:**2296–2299.
17. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797.

18. **Edger, P. P., and J. C. Pires.** 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chrom. Res. **17:** 699–717.
19. **Encarnación, S., M. Dunn, K. Willms, and J. Mora.** 1995. Fermentative and aerobic metabolism in *Rhizobium etli*. J. Bacteriol. **177:**3058–3066.
20. **Encarnación, S., Y. Guzmán, M. F. Dunn, M. Hernández, M. D. C. Vargas, and J. Mora.** 2003. Proteome analysis of aerobic and fermentative metabolism in *Rhizobium etli* CE3. Proteomics **3:**1077–1085.
21. **Encarnación, S., M. Hernández, G. Martínez-Batallar, S. Contreras, M. D. C. Vargas, 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. Proc. Online **7:**117–135.
22. **Encarnación, S., M. C. Vargas, M. F. Dunn, A. Dávalos, G. Mendoza, Y. Mora, and J. Mora.** 2002. AniA regulates reserve polymer accumulation and global protein expression in *Rhizobium etli*. J. Bacteriol. **184:**2287–2295.
23. **Ferraioli, S., R. Taté, E. Caputo, A. Lamberti, A. Riccio, and E. J. Patriarca.** 2001. The *Rhizobium etli argC* gene is essential for arginine biosynthesis and nodulation of *Phaseolus vulgaris*. Mol. Plant-Microbe Interact. **14:**250–254.
24. **Figueroa, N., and L. Bossi.** 1988. Transcription induces gyration of the DNA template in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **85:**9416–9420.
25. **Figurski, D. H., and D. R. Helinski.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. U. S. A. **76:**1648–1652.
26. **Fraser, C., W. P. Hanage, and B. G. Spratt.** 2007. Recombination and the nature of bacterial speciation. Science **315:**476–480.
27. **Fredrick, K., and M. Ibba.** 2010. How the sequence of a gene can tune its translation. Cell **141:**227–229.
28. **Girard, L., S. Brom, A. Dávalos, O. López, M. Soberón, 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.
29. **González, V., R. I. Santamaría, P. Bustos, I. Hernández-González, A. Medrano-Soto, G. Moreno-Hagelsieb, S. C. Janga, M. A. Ramírez, V. Jiménez-Jacinto, J. Collado-Vides, and G. Dávila.** 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc. Natl. Acad. Sci. U. S. A. **103:**3834–3839.
30. **Goryanin, I. I., G. V. Lebedeva, E. A. Mogilevskaya, E. A. Metelkin, and O. V. Demin.** 2006. Cellular kinetic modeling of the microbial metabolism. Methods Biochem. Anal. **49:**437–488.
31. **Guerra, H.** 2007. The brucellae and their success as pathogens. Crit. Rev. Microbiol. **33:**325–331.
32. **Guerrero, G., H. Peralta, A. Aguilar, R. Diaz, M. A. Villalobos, A. Medrano-Soto, and J. Mora.** 2005. Evolutionary, structural and functional relationships revealed by comparative analysis of syntenic genes in Rhizobiales. BMC Evol. Biol. **5:**55.
33. **Guindon, S., and O. Gascuel.** 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:**696–704.
34. **Hanahan, D., J. Jessee, and F. R. Bloom.** 1991. Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol. **204:**63–113.
35. **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.
36. **Hynes, M. F., and N. F. McGregor.** 1990. Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. Mol. Microbiol. **4:**567–574.
37. **Igarashi, K., and K. Kashiwagi.** 2000. Polyamines: mysterious modulators of cellular functions. Biochem. Biophys. Res. Commun. **271:**559–564.
38. **Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, and S. Tabata.** 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7:**331–338.
39. **Khersonsky, O., and D. S. Tawfik.** 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. **79:**11.1–11.35.
40. **Knight, C. G., R. Kassen, H. Hebestreit, and P. B. Rainey.** 2004. Global analysis of predicted proteomes: functional adaptation of physical properties. Proc. Natl. Acad. Sci. U. S. A. **101:**8390–8395.
41. **Korbel, J. O., L. J. Jensen, C. von Mering, and P. Bork.** 2004. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. Nat. Biotechnol. **22:**911–917.
42. **Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop II, and K. M. Peterson.** 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene **166:**175–176.
43. **Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin.** 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. Science **324:** 255–258.
44. **Lawrence, J. G.** 2002. Gene transfer in bacteria: speciation without species? Theoret. Pop. Biol. **61:**449–460.
45. **Lemoine, F., O. Lespinet, and B. Labedan.** 2007. Assessing the evolutionary rate of positional orthologous genes in prokaryotes using synteny data. BMC Evol. Biol. **7:**237.
46. **Lind, P. A., C. Tobin, O. G. Berg, C. G. Kurland, and D. I. Andersson.** 2010. Compensatory gene amplification restores fitness after inter-species gene replacements. Mol. Microbiol. **75:**1078–1089.
47. **Livak, K. J., and T. D. Schmittgen.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods **25:**402–408.
48. **Llácer, J. L., I. Fita, and V. Rubio.** 2008. Arginine and nitrogen storage. Curr. Opin. Struct. Biol. **18:**673–681.
49. **Lu, C. D., Z. Yang, and W. Li.** 2004. Transcriptome analysis of the ArgR regulon in *Pseudomonas aeruginosa*. J. Bacteriol. **186:**3855–3861.
50. **Masson-Boivin, C., E. Giraud, X. Perret, and J. Batut.** 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol. **7:**458–466.
51. **Meade, H. M., S. R. Long, G. B. Ruvkun, S. E. Brown, and F. M. Ausubel.** 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn*5* mutagenesis. J. Bacteriol. **149:**114–122.
52. **Moreno-Hagelsieb, G., and J. Collado-Vides.** 2002. A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics **18:**S329–S326.
53. **Noel, K. D., A. Sánchez, L. Fernández, J. Leemans, and M. A. Cevallos.** 1984. *Rhizobium phaseoli* mutants with transposon Tn*5* insertions. J. Bacteriol. **158:**148–155.
54. **Pappas, K. M.** 2008. Cell-cell signaling and the *Agrobacterium tumefaciens* Ti plasmid copy number fluctuations. Plasmid **60:**89–107.
55. **Phair, R. D.** 1997. Development of kinetic models in the nonlinear world of molecular cell biology. Metabolism **46:**1489–1495.
56. **Piette, J., R. Cunin, A. Boyen, D. Charlier, M. Crabeel, F. Van Vliet, N. Glansdorff, C. Squires, and C. L. Squires.** 1982. The regulatory region of the divergent *argECBH* operon in *Escherichia coli* K-12. Nucleic Acids Res. **10:**8031–8048.
57. **Prentki, P., and H. M. Krisch.** 1984. In vitro insertional mutagenesis with a selectionable DNA fragment. Gene **29:**303–313.
58. **Price, M., K. Huang, E. Alm, and A. Arkin.** 2005. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucleic Acids Res. **33:**880–892.
59. **Proshkin, S., R. A. Rahmouni, A. Mironov, and E. Nudler.** 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science **328:**504–508.
60. **Quandt, J., and M. F. Hynes.** 1993. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. Gene **127:** 15–21.
61. **Quintero, M. J., A. M. Muro-Pastor, A. Herrero, and E. Flores.** 2000. Arginine catabolism in the cyanobacterium *Synechocystis* sp. Strain PCC 6803 involves the urea cycle and arginase pathway. J. Bacteriol. **182:**1008–1015.
62. **Quinto, C., H. De la Vega, M. Flores, I. Fernández, T. Ballado, G. Soberón, and R. Palacios.** 1982. Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. Nature **299:**724–726.
63. **Randhawa, G. S., and R. Hassani.** 2002. Role of rhizobial biosynthetic pathways of amino acids, nucleotide bases and vitamins in symbiosis. Indian J. Exp. Bot. **40:**755–764.
64. **Rocha, E. P.** 2004. Order an disorder in bacterial genomes. Curr. Opin. Microbiol. **7:**519–527.
65. **Sallstrom, B., and S. G. E. Andersson.** 2005. Genome reduction in alphaproteobacteria. Curr. Opin. Microbiol. **8:**579–585.
66. **Shevchuk, N. A., A. V. Bryksin, Y. A. Nusinovich, F. C. Cabello, M. Sutherland, and S. Ladisch.** 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res. **32:**e19.
67. **Svenningsen, S. L., K. C. Tu, and B. L. Bassler.** 2009. Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. EMBO J. **28:**429–439.
68. **Tabor, C. W., and H. Tabor.** 1985. Polyamines in microorganisms. Microbiol. Rev. **49:**81–99.
69. **Tamames, J.** 2001. Evolution of gene order conservation in prokaryotes. Genome Biol. **2:**R0020.
70. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
71. **Tuller, T., A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, and Y. Pilpel.** 2010. An evolutionary conserved mechanism for controlling the efficiency of protein translation. Cell **141:**344–354.
72. **Zhao, J., D. Ridgway, G. Broderick, A. Kovalenko, and M. Ellison.** 2008. Extraction of elementary rate constants from global network analysis of *Escherichia coli* central metabolism. BMC Syst. Biol. **2:**41.

Material suplementario.

Fig. S1

Supplementary **Fig. S1**. Scheme of the detected promoter for *argC* in *S. meliloti*. The putative sigma 70 promoter (box -10) was underlined. The transcription start site, a G denoted with +1 was determined with rapid amplification of cDNA 5´ends (RACE) using total RNA purified from wild-type strain cells. The end of *speB* coding sequence is denoted with a dot. The start of *argC* coding sequence also was remarked.

Table S2. Expression of *gus* gene in *S. meliloti* 1021 (wild-type) under the control of different promoters.

Strain	Fusion genotype	Mean β-glucuronidase specific activity (±SD)[a]
Sm1021/pFGP5	p*lac::gus*	551 ± 99
Sm1021/pFGP6	p*argC::gus*	169 ± 30
Sm1021/pFGP7	p*speB::gus*	21 ± 3

[a]Values are expressed as nmol of p-nitrophenol per minute per milligram of protein. Data represent the means ± SD of two replicates from three independent experiments.

Table S1. List of primers used in this work.

***argC* promoter costructions.**

S. meliloti argC promoter with overlapping end from *A. tumefaciens argC* gene. This fragment was obtained by PCR with the specific primers SmUpXba (GCT CTA GAC GAT GCA GCC ACG TCG AAT CGC C, underlined, built-in *Xba*I site) and proCSm+AtC_AR (**GAA GAT CTT CGC TGT CAT** GAT TTC CCT GCT CCT GCA GTC C, bold nucleotides corresponds to *A. tumefaciens argC* overlapping end).

S. meliloti argC promoter with *R.etli argC* overlapping end. This fragment was obtained by PCR with the specific primers SmUpXba (GCT CTA GAC GAT GCA GCC ACG TCG AAT CGC C, underlined, built-in *Xba*I site) and proCSm+ReC_AR **(GAT GAA GAT TTT CGG TGC CAT** GAT TTC CCT GCT CCT GCA GTC C, bold nucleotides corresponds to *R. etli argC* overlapping end).

S. meliloti argC promoter with *M. loti argC* overlapping end. This fragment was obtained by PCR with the specific primers SmUpXba (GCT CTA GAC GAT GCA GCC ACG TCG AAT CGC C, underlined, built-in *Xba*I site) and proCSm+MlC_AR **(CGA TGA AGA TTT TCG GTT TCA T**GA TTT CCC TGC TCC TGC AGT CC, bold nucleotides corresponds to *M. loti argC)*

A. tumefaciens argC gene, with *S. meliloti argC* promoter overlapping end. obtained using primers Up_H2_P*at* (**GCC AGA CTG ATT CCG GAA ACG CCG CTA ACG GAC TGC AGG AGC AGG GAA ATC** ATG ACA GCG AAG ATC TTC AT) and Lw_H2_P*at* (**CGA AAA CCC CGG CAG GAA CGC TGC CGG GGT TTT CGT CTC GAT TTG CTG TG**T CAT GCC GAG AGC ATC AAC T).

R. etli argC gene, with *S. meliloti argC* promoter overlapping end, obtained using primers Up_H4_P*re* (**GCC AGA CTG ATT CCG GAA ACG CCG CTA ACG GAC TGC AGG AGC AGG GAA ATC** ATG GCA CCG AAA ATC TTC AT) and Lw_H4_P*re* (**CGA AAA CCC CGG CAG GAA CGC TGC CGG GGT TTT CGT CTC GAT TTG CTG TG**T CAG GAG GCG AGC ATC AGG).

M. loti argC gene, with *S. meliloti argC* promoter overlapping end, obtained using primers Up_H3_P*ml* (**GCC AGA CTG ATT CCG GAA ACG CCG CTA ACG GAC TGC AGG AGC AGG GAA ATC** ATG AAA CCG AAA ATC TTC AT) and Lw_H3_P*ml* (**CGA AAA CCC CGG CAG GAA CGC TGC CGG GGT TTT CGT CTC GAT TTG CTG TG**T CAA AGC CCG AGC ATC AGG T). Bold letters in primers correspond to the *S. meliloti* homologous region.

Primers used to construct a plasmid carrying the promoter and coding region of the *argC* gene from *S. meliloti* for complementation experiments **(**p53gus::*argCSm*) SmUpXba (GCT CTA GAC GAT GCA GCC ACG TCG AAT CGC C, underlined, built-in *Xba*I site) and Lw_H1_PSm (CGA AAA CCC CGG CAG GAA CGC TGC CGG GGT TTT CGT CTC GAT TTG CTG TGT CAG GCG GAC AGC ATC AGG T).

To construct transcriptional fusions using the b-glucuronidase gene as reporter under the control of the *S. meliloti argC* promoter PCR products were obtained with primers SmUpXba (GCT CTA GAC GAT GCA GCC ACG TCG AAT CGC C, underlined, built-in *Xba*I site) and FusC_Gus_R (GGG GGC CCT TGC CGA GAT TGT CG, this primer contain an *Apa*I restriction site).

***speB* promoter constructions.**

S. meliloti speB promoter with overlapping end from *S. meliloti argC* gene. This fragment was obtained by PCR with the specific primers speBsm-F2 (TCTAGATCAA C GGCAAGCCTTTCTCGGAC) and Sm_argC_R2 (TTGATTCAATTCTATGCCAG ACTGATTCCG).

S. meliloti speB promoter with overlapping end from *A. tumefaciens argC* gene. This 500 bp fragment was obtained by PCR with the specific primers speBsm-F2 (TCTAGA TCAACGGCAAGCCTTTCTCGGAC) and SmPro+AtC (GAAGATCTTCGCTGTCA TGCTGTTCTCCCGTCGGATGG).

S. meliloti speB promoter with overlapping end from *M. loti argC* gene. This 500 bp fragment was obtained by PCR with the specific primers speBsm-F2 (TCTAGATCAA C G GCAAGCCTTTCTCGGAC) and SmPro+MlC (CGATGAAGATTTTCGGTTTC ATGCTGTTCTCCCGTCGGATGG).

S. meliloti speB promoter with overlapping end from *R. etli argC* gene. This 500 bp fragment was obtained by PCR with the specific primers speBsm-F2 (TCTAGATCAA CGGCAAGCCTTTCTCGGAC) and SmPro+ReC (GATGAAGATTTTCGGTGCCAT GCTGTTCTCCCGTCGGATGG)

S. meliloti argC gene, with *S. meliloti speB* promoter overlapping end, were obtained using primers pBBR1MCS3F (TAAGTTGGGTAACGCCAGGG) and Sm_argC_R1 (ATG AAACCGAAGATCTTTATCGATGGC).

A. tumefaciens argC gene, with *S. meliloti speB* promoter overlapping end, were obtained using primers pBBR1MCS3F (TAAGTTGGGTAACGCCAGGG) and **At_argC** (ATG ACAGCGAAGATCTTCATCGATGGCGAACACGG).

R. etli argC gene, with *S. meliloti speB* promoter overlapping end, were obtained using primers pBBR1MCS3F (TAAGTTGGGTAACGCCAGGG) and Re-argC (ATGGCA CCGAAAATCTTCATCGATGGCGAACACGG).

M. loti argC gene, with *S. meliloti speB* promoter overlapping end, were obtained using primers pBBR1MCS3F (TAAGTTGGGTAACGCCAGGG) and Ml_argC (ATGAAA CCGAAAATCTTCATCGACGGCGAGCACGG).

Primers used to construct a plasmid carrying the promoter and coding region of the *argC* gene from *S. meliloti* for complementation experiments, speBsm-F2 (TCTAGAT CAACGGCAAGCCTTTCTCGGAC) and pBBR1MCS3F (TAAGTTGGGTAACGC CAGGG). These PCR products were cloned into PCR 2.1 TOPO and subcloned into pBBMCS53 and pBBMCS-3 as SpeI-ApaI fragments.

***lac* promoter constructions.**

S.meliloti, primers **argCSmKpn5 (**GGGGTACC**AAGGAA**ACAGCT**ATG**AAACC GA AAATCTTTATCGATGGCG. Bold nucleotides correspond to RBS and start of translation of *argC.*) and argCSmXba3 (CGTCTAGAACTGAGCCAATGGATTAT CGCGAAGC).

A. tumefaciens, primers argCAtKpn5 **(**GGGGTACC**AAGGAA**AC AGCT**ATG**AC AGCGAAAATCTTCATTGATGGCGA. Bold nucleotides correspond to RBS and start of translation of *argC.*) and argCAtXba3 (CGTCTAGAGTATCGGCGATGCGT TTTGCGGCC).

R. etli, primers argCReKpn5 **(**GGGGTACC**AAGGAAACAGCTATG**GCACCGAA A ATCTTTATCGATGGCG, Bold nucleotides correspond to RBS and start of translation of *argC.*) and **a**rgCReXba3 (CGTCTAGAGCTGCGCCGAAACCGGAAA AGCC).

M loti, primers argCMlKpn5 **(**GGGGTACC**AAGGAA ACAGCTATG**AAACCGA AAATCTTCATCGACGGCG. Bold nucleotides correspond to RBS and start of translation of *argC.*) and argCMlXba3 (CGTCTAGACGAAAGGTTTGAGCGCCT TGACGACGG).

These PCR products were cloned into PCR 2.1 TOPO and subcloned into pBBMCS53 and pBBMCS-3 as *KpnI-XbaI* fragments.

Native p*argC-argC* constructions.

S.meliloti, primers SmUpXba **(**GCTCTAGACGATG CAGCCACGTCGAATCGCC) and SmLowKpn (CCCGGCACCTATCCGAGAGACACGGCG).

A. tumefaciens, primers AtUpXba **(**GCTCTAGACGACG GCGATCTGGGGTAT TCCG) and AtLowKpn (CGGGTACCTATCGGCGATGCGTTTTGCGGCC).

R. etli, primers ReUpXba **(**GCTCTAGAACATCGAACAGACCGGGCACGC) and ReLowKpn (CAGGTACCC TGCGCCGAAACCGGAAAAGCC).

M loti, primers MlUpXba **(**CTAGAGG GCCGCAAGCGATTCGCCGC) and MlLowKpn (CCGGTACCCGAAAGGTTTGAGCGCCTTGACGACGG).

These PCR products were cloned into PCR 2.1 TOPO and subcloned into pBBMCS53 and pBBMCS-3 as XbaI-KpnI.

Primers for transcriptional fusion

In all constructions we only used the reverse primer FusC_Gus_R **(**GGGGGCCCTTG CCGAGATTGTCG) and the corresponding forward primer.

Primers for Real Time-PCR

a) for cDNA quantification

A tumefaciens argC, for

At-argCfwd 5´-CCGCCTATCGCGTCCAT-3

At-argCrev 5´-GGTCCCGGTCCATTTCG-3

TagMan-AtargC- 6FAMAGGACTGGGCTTACGMGBNFQ

M. loti argC,

Ml-argCfwd 5´-GAGCTTCATGCGGCCATT-3

Ml-argCrev 5´-ACCACACCGCCCTTGATG-3

TagMan-MlargC- 6FAMCCGACCATTTTGCMGBNFQ

R. etli argC,

Re-argCfwd 5´-GCATTATGCCGGACAGGAAA-3

Re-argCrev 5´-CGCCTTGCTGTCGGAAAG-3

TagManReargC- 6FAMCGTCACCGTCGTGCCMGBNFQ

S. meliloti argC,

Sm-argCfwd 5´-CAACGCTCGAAACGATCCA-3

Sm-argCrev 5´-CGACTGTCCCGCATAATGG-3

TagMan-SmargC- 6FAMCGTGCCCTCGTCGMGBNFQ

S. meliloti rpoA,

rpoA-fwd 5´-TGAAGGCCAGGTTCTCGACTA-3

rpoA-rev 5´-CGGAGCCATCGGTTTC-3

TagMan-rpoA VICCAAGCTGACGATGTCMGBNFQ

gus,

gus-fwd 5´-CAAAGCGGCGATTTGGAA-3

gus-rev 5´-GCCAGGCCAGAAGTTCTTTTT-3

TagMan-gus 6FAMCGGCAGAGAAGGTACMGBNFQ

b) for plasmid copy number

rpoA_FW (TCAACCCCGCGCTTCTC)

rpoA_REV (TTCTGCTTCGGTCTTCTGAATG)

Tc_FWD (GCGGCGGCCAAAGCGG)

Tc_REV (TATGGCGTGCTGCTAGCG)

Resultados adicionales no publicados

Análisis funcional de ortólogos *argC* de Rhizobiales en el genoma de *Escherichia coli*

Planteamiento del trabajo:

El propósito de este proyecto fue evaluar funcionalmente la FDE de ArgC de Rhizobiales (Sm*, Re, At* y *Ml*) mediante su reemplazamiento en el genoma de *E. coli*.

En el articulo publicado (Díaz *et al.* 2011) se observó que los ortólogos de *argC* de especies cercanas (Rhizobiales) al ser sustituidos en el genoma de *S. meliloti*, no son completamente equivalentes y además se identificaron cambios a diferentes niveles (incluyendo el metabólico, transcripcional y traduccional). Al hacer los reemplazamientos en el genoma de *E. coli*, esperabamos que estas diferencias funcionales en la FDE fueran mas notorias debido a la lejania filogenética entre Rhizobales y *E.coli*, ya que ambas ocupan nichos diferentes, los rizobios son alfa proteobacterias y *E. coli* es una γ-proteobacteria, sus temperaturas óptimas de crecimiento son de 30 vs 37, los tiempos de duplicación son 2-3 vs 1-1.5 hrs y sus porcentages de GC son 60-65% vs 51% entre otras muchas diferencias metabólicas y genéticas.

El reemplazo en *E. coli por* "recombineering" de las regiones codificantes de *argC*, nos permitió dejar intacto su contexto genómico y por lo tanto sus mecanismos de regulación transcripcional, de tal forma que la unica región genética extraña a *E. coli* fueran unicamente las regiones codificantes de *argC*.

Antecedentes

Como se comentó anteriormente el análisis comparativo de ArgC mostró altos niveles de identidad y similitud entre las cuatro especies de Rhizobiales analizadas (**Fig 1, Fig.9**) (*S. meliloti*, *A. tumefaciens*, *M. loti* y *R. etli*), mientras que la comparación con ArgC de *E. coli* muestra bajos niveles de identidad (**Fig. 10**) (41 residuos idénticos), y a diferencia de los 310 aminoácidos presentes en la proteína ArgC de Rhizobiales, ArgC de *E. coli* esta formado por 334 aminoácidos.

Fig. 10. Alineamiento multiple de las proteínas ArgC de 5 especies de Rhizobiales y E. coli. En amarillo los residuos identicos. En rojo residuos diferentes de *E. coli.* 41 posiciones con residuos identicos*. E. coli*,199 residuos diferentes (64.1%).

Estrategia experimental

Para sustituir *in situ* , las regiones regiones codificantes de ArgC se utilizó la tecnología de "recombineering" que consiste en la recombinación mediada por proteínas del sistema lambda red (Datsenko & Wanner 2000).

Como estrategia de selección de recombinantes, se generó una mutante *E. coli* auxótrofa de arginina, mediante la inserción del gen de resistencia al antibiótico Kanamicina (Km) (Fig 11A). Dicha estrategia generó un reemplazo del gen *argC* nativo por el gen de resistencia a Kanamicina. Posteriormente esta construcción inicial se utilizó como sustrato de recombinación y selección para el reemplazo de los *argC* de Rhizobiales (Fig 11B). Las *E. coli* recombinantes conteniendo los ortólogos *argC* de Rhizobiales, fueron seleccionadas por su capacidad de crecer en medio mínimo M9 y pérdida de resistencia a Km.

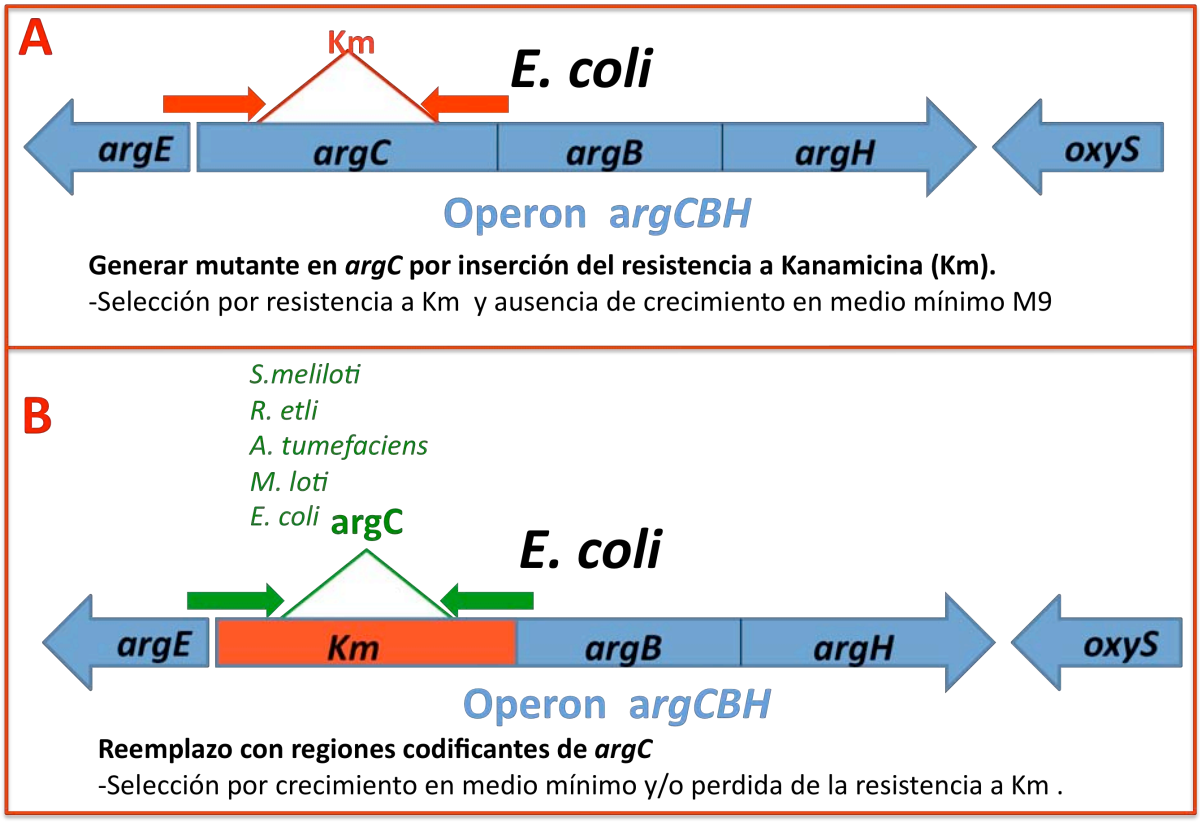


Fig. 11. Estrategia para obtener reemplazos de *argC* de Rhizobiales en *E. coli MG1653 mediante "recombineering"*. A) en rojo, se generó una amplificación que contenía la region codificante de resistencia a Km y 80 nucleotidos de las regiones genéticas adyacentes a *argC* de *E. coli*. Esta amplificación se recombinó mediante "recombineering" en la zona codificante de *argC* de *E. coli*. La selección de la

recombinante fue mediante auxotrofía y resistencia a Km. B) En verde, la recombinación generada se utilizó como sustrato de una segunda serie de recombinaciones con amplificaciones de los diferentes argC de *Sm, Re, At, Ml* y *Ec* conteniendo 80 nucleotidos de las regiones genéticas adyacentes a *argC* de *E. coli*. La selección de estas nuevas recombinantes fue por pérdida de auxotrofía y resistencia a Km.

Resultados

Siguiendo la estrategia antes mencionada de la generación de reemplazos sitio específicos de *argC* de Rhizobiales, se selecionaron las recombinantes *argC* de Rhizobiales (*Sm, Re, At, Ml*) en el genoma de *E. coli*. Estas, además de ser seleccionadas por fenotipo, fueron comprobados por PCR e hibridación tipo Southern. Respecto al fenotipo (**Fig. 12)**, se observó que después de 24 h de incubación, todos los reemplazos *in situ* de *argC* fueron capaces de restaurar el crecimiento de la mutante $argC^{-}$ Km^{R} cuando fueron crecidas en medio mínimo M9 sólido.

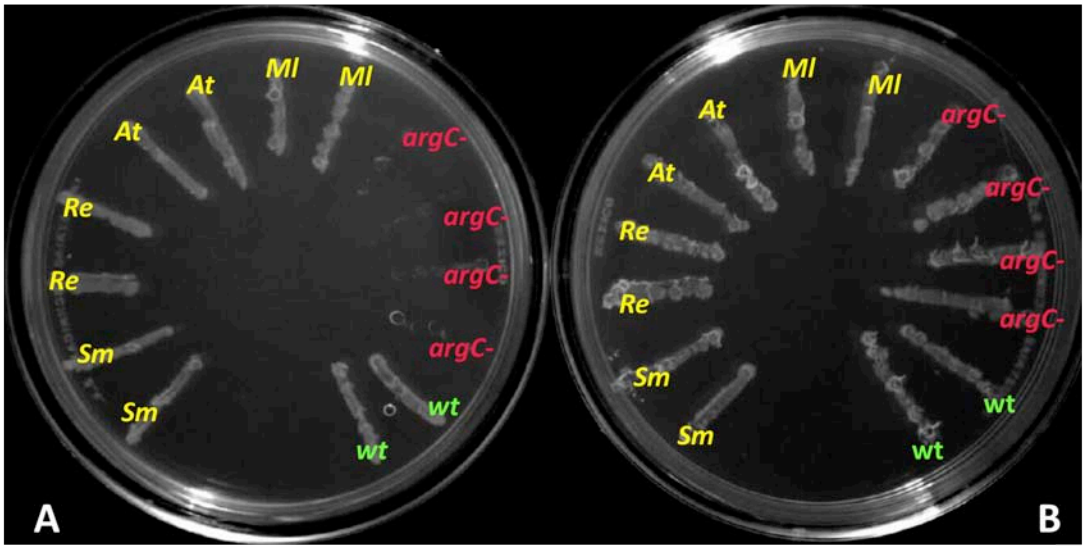


Fig. 12. E. coli con reemplazos *in situ* de ortólogos *argC* de Rhizobiales. Cepas recombinantes incubadas 24 horas em medio mínimo M9 sólido. Verde, cepa silvestre (wt); Rojo, mutante $argC^{-}$ Km^{R}; Amarillo, *E. coli* con reemplazos de *argC* (*Sm, Re, At, Ml*) **A)** Sin arginina. **B)** Con arginina1.14 mM.

Para analizar el efecto de los reemplazos sitio específicos *argC,* en el crecimento de una manera más fina, las cepas fueron crecidas en medio mínimo M9+Glucosa 10 mM

y Cloruro de amonio 10mM. Se cuantificó la producción de proteína con respecto al tiempo **(Fig. 13)**. Se observó que con respecto a la cepa silvestre, todas las recombinantes *argC* fueron afectadas en el crecimiento. La cepa silvestre y la mutante *argC*$^{-}$ Km^{R} suplementada con Arginina tuvieron ambas un tiempo de duplicación de 1.5 h, mientras que *Sm* 1:45; At 2:15; Re 2:20 y Ml 2:00.

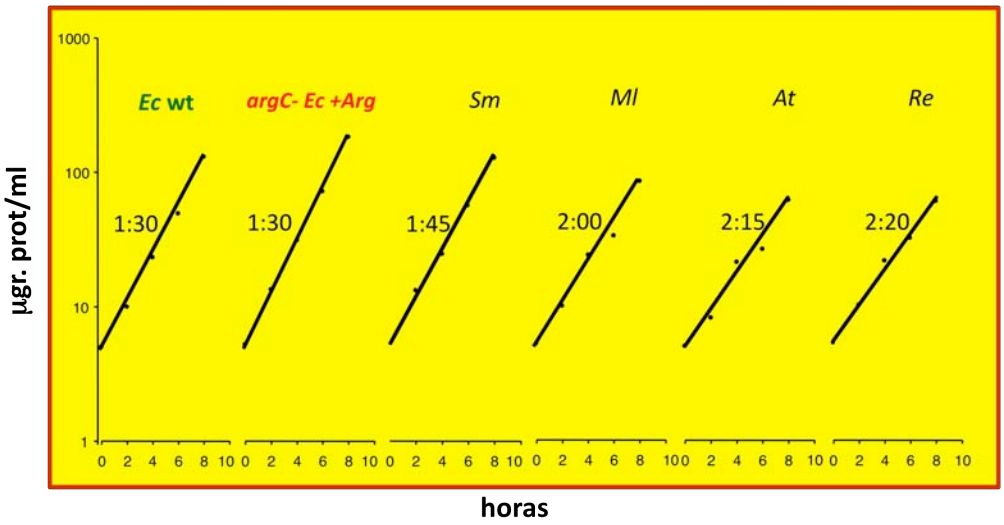


Fig. 13. Crecimiento de *E. coli* con reemplazos *argC* de Rhizobiales. Crecimientos en medio mínimo M9 líquido adicionado con Glucosa 10 mM y Cloruro de amonio 10 mM. Letras verdes, cepa parental de *E. coli*; Letras rojas, mutante argC^{-} KmR. Letras negras, *E. coli* con sustituciones: *Sm*; *Ml*; *At*; *Ml*.

Respecto al análisis de la cuantificación de mensajero presente en las recombinantes *argC* **(Fig. 14)**, se observó que todas tuvieron diferentes niveles de expresión transcripcional de mRNA *argC* con respecto a la *E. coli* no modificada. En la recombinante de *At* fueron observados los más bajos niveles de expresión, mientras que la recombinante de *Sm* exhibió los mas altos niveles, con valores cercanos al doble de la expresión en *E. coli* sin modificar.

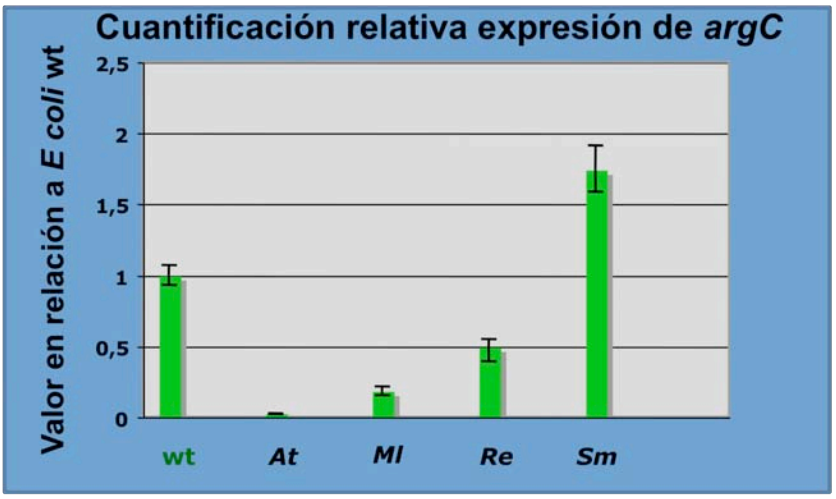


Fig. 14. Cuantificación relativa del mRNA de *argC* de Rhizobiales presentes en *E. coli*. Gráfica que muestra la cuantificación relativa del mRNA argC obtenida por PCR en Tiempo Real. Las cepas con reemplazos fueron crecidas en medio mínimo M9 líquido, adicionado con glucosa 10mM y Cloruro de amonio 10 mM, en ausencia de arginina. Se utilizó el control endógeno *gapA*. Letras verdes, la cepa silvestre de *E. coli*. Letras negras, *E. coli* con reemplazos *argC* de *At, Ml, Re, Sm.*

La Actividad enzimática de ArgC (**Fig. 15**) exhibió diferentes capacidades, dependiendo del *argC* (*Sm, Re, At, Ml*). Todas las actividades de las enzima ArgC sustitutas fueron menores a la actividad enzimática nativa.

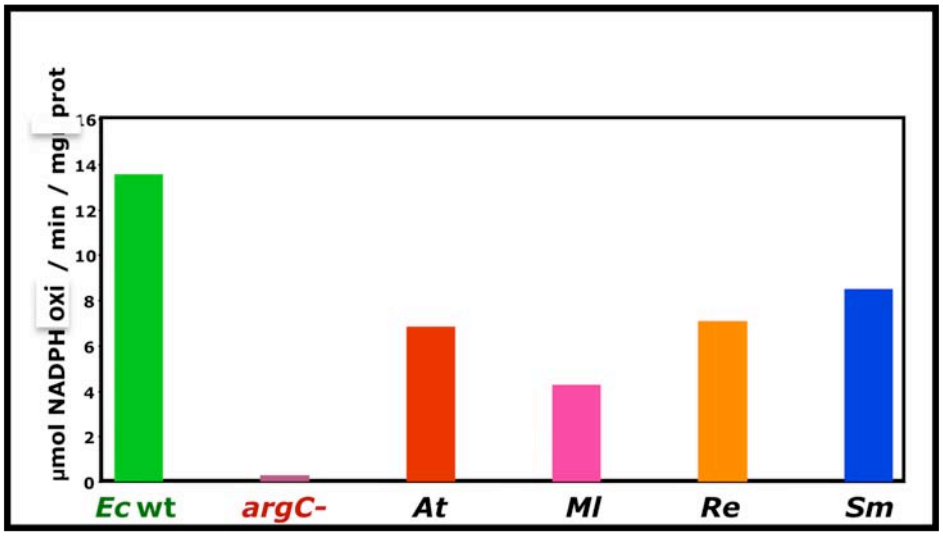


Fig. 15. Actividad enzimática de ArgC expresada como µmol de NADPH oxidados/min/mg de proteína. Letras verdes, cepa parental de *E. coli*; Letras rojas, cepa mutante $argC^{-}$ Km^{R}. de *E. coli.* Letras negras, reemplazos con *argC* de Rhizobiales *(sm, Re, At, Ml).*

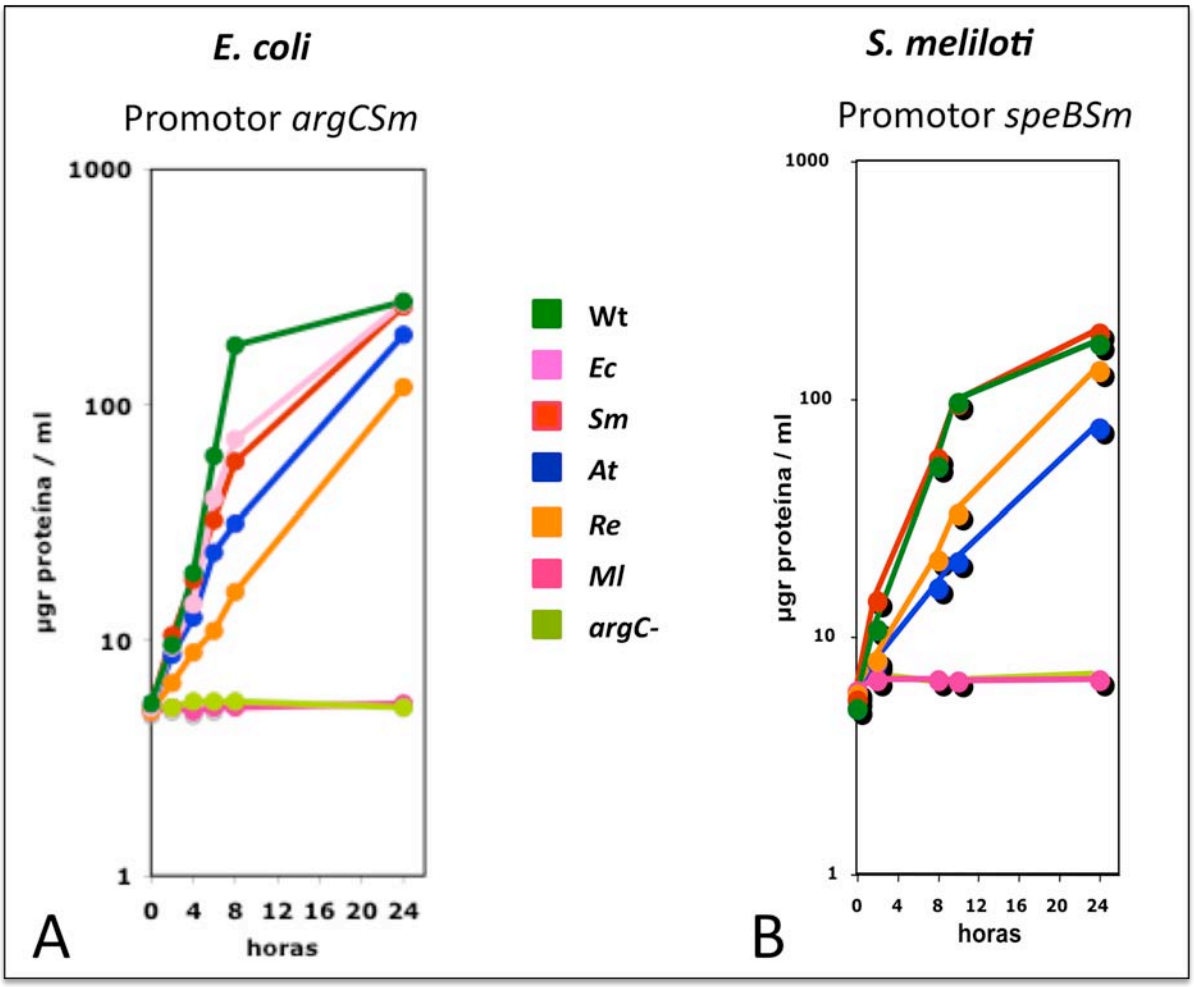


Fig. 16. Curvas de crecimiento de cepas *E. coli* y *Sm argC*$^{-}$ complementadas con clonaciones *argC*. **A**) Complementaciones de *E. coli* argC^{-} Km^{R} con *argCs* de Rhizobiales fusionados al promotor *argC* de *S. meliloti. Cepas crecidas en medio mínimo M9 líquido adicionado con Glucosa 10 mM y Cloruro de amonio 10 mM. Verde oscuro, Cepa silvestre. Verde claro, mutante argC-. Rosa claro, complementada con argC Ec. Magenta, Mutante complemaentada con argC Ml. Roja, complementada con argC Sm. Azul, complementada con argC At; Naranja, complementada con argC Re.* **B**) Complementaciones de *Sm* argC^{-} con *argCs* fusionados al promotor *speB* de S. meliloti. Crecimientos Medio mínimo adicionado con Succinato 10 mM ycloruro de amonio 10 mM (*Díaz 2011*, Figura 4E). Verde Claro, cepa mutante; Rosa, complentada con *Ml*; Azul, complementada con *At*; Naranja, complementada con *Re*; Verde, cepa silvestre; Rojo, complementada con *Sm*.

En el trabajo publicado (Díaz *et al.* 2011) las condiciones de complementación que reflejaron mejor las diferencias funcionales de las FDE, fueron las complementaciones de los *argC* fusionados al promotor *speB* de *S. meliloti.* Por lo que, con el propósito de analizar y comparar el efecto que pudiera tener la complementación con clonaciones de

argC sobre el crecimiento de *E. coli*. Fueron movilizadas a la mutante $argC^{-}$ Km^{R} de *E. coli*, las clonaciones de los *argC* de Rhizobiales y *E. coli* fusionadas al promotor *argC* de *S. meliloti.* En la Figura 16, se muestra la comparación de los crecimentos de las complementaciones de ambas mutantes *argC*$^{-}$*, E. coli* y *S. meliloti.* En ambas series de complementaciones *E. coli* con *argCs* fusionados al promotor de *Sm* y *S. meliloti* con *argCs* fusionados al promotor de p*speB*, se observa que ninguna complementación fue igual y que las complementaciones con *argC* de *M. loti* no restablecen la prototrofía de ambas mutantes.

Discusión de resultados.

En un alineamiento múltiple de proteínas ortólogas de especies cercanas se pueden identificar aquellas diferencias particulares en la secuencia de aminoácidos, estas diferencias particulares constituyen la FDE (Firma de la especie), (Guerrero *et al.* 2005). Debido a que estas diferencias radican en estructuras químicas de residuos de aminoácidos, la FDE, implica cambios en las propiedades fisicoquímicas de las proteínas ortólogas comparadas.

En este trabajo se analizó la FDE desde un punto de vista funcional, mediante la complementación de genes ortólogos *argC* en el mismo fondo genético (*S. meliloti argC*$^{-}$), con los mismos promotores y comparando con la cepa silvestre de *S. meliloti.*

Utilizamos ortólogos de *argC* porque este gen forma parte de una vía muy importante comprometida con la síntesis de aminoácidos y proteínas. ArgC es una enzima intermediaria en la síntesis de arginina. La arginina es un aminoácido esencial que demanda la mayor cantidad de nitrógeno con respecto a los otros aminoácidos, su síntesis requiere de una gran cantidad de energía, además de ser precursor en la síntesis de poliaminas (Cunin *et al.* 1986*,* Llácer *et al.* 2008*,* Randhawa & Hassani 2002*,* Tabor & Tabor 1985).

Para tener un organismo receptor de los ortólogos *argC*, se mutó la región codificante correspondiente a *argC* en el cromosoma de *S. meliloti*, y cada ortólogo cercano fue clonado en el mismo vector y fusionado a los mismos promotores para posteriormente ser introducidos en el fondo genético de la mutante *argC*$^{-}$. Con el propósito de tener diferentes niveles de expresión, se utilizaron tres promotores: de baja (*pspeB*), media (*pargC*), y alta (*lac*) expresión. Los promotores de *speB* y *argC* provenientes de *S. meliloti*, se expresan en las condiciones utilizadas y probablemente son susceptibles de ser regulados. Mientras que el promotor *lac* es un promotor de *E. coli* que se expresa de forma constitutiva.

Por otro lado, los ortólogos de *argC* fueron reemplazados *in situ* en el genoma de *E. coli*, para tener una condición de contexto genómico intacto a excepción de las regiones codificantes de los ortólogos *argC*.

El desarrollo del experimento se basó en la recuperación de la prototrofía de la mutante *argC*$^{-}$, mediante complementación de genes ortólogos cercanos, donde se analizó el rol funcional de la firma. Se observó que existen varios factores que pueden influenciar el desempeño de los ortólogos analizados. Estos factores son cambios en el número de copias, eficiencia de transcripción y traducción, actividad enzimática, cambios fisiológicos y crecimiento, todos relacionados a la diferente capacidad de complementación cada proteína ortóloga sustituta.

Las diferencias de expresión proporcionadas por cada promotor (*pspeB*, *pargC* , *plac*) se reflejó en diferentes grados de actividad enzimática y crecimiento con el mismo ortólogo. Las complementaciones de *argC* bajo el promotor p*speB*, resultaron en las más bajas actividades enzimáticas y las mayores diferencias respecto al crecimiento, donde las complementaciones con los ortólogos *R. etli* y *A. tumefaciens* mostraron crecimientos reducidos, mientras que la cepa complementada con *argC de M. loti* no creció (Fig. 4E y 4F).

De forma semejante la mutante *E. coli argC*$^{-}$ complementada con reemplazos *in situ,* también reflejó diferentes grados de actividad enzimática (Fig. 15), donde todos los reemplazos ortológos de *argC* de Rhizobiales exhibieron menores actividades enzimáticas, con respecto a la cepa silvestre de *E. coli*. De igual forma, respecto al crecimiento, todos los reemplazos ortológos de Rhizobiales resultaron afectados en el crecimiento (Fig. 13), pero fueron capaces de recuperar la prototrofía a diferencia de las complementaciones de *S. meliloti argC*$^{-}$ con los promotores p*speB* y p*argC* (Fig. 4E y 4F), donde el ortólogo *argC* de *Ml* no fue capaz de recuperar la prototrofía.

En las complementaciones de *Sm argC*$^{-}$ bajo promotor de *speB*, fueron encontradas de 2 a 3 copias del plásmido conteniendo los ortólogos de *S. meliloti*, *R. etli* y *A. tumefaciens*, mientras que se encontraron 9 copias con el ortologo de *M. loti*. La secuencia mas divergente a *S. melilltoti* es precisamente la de *M. loti*, y la cepa complementada con *M. loti* bajo el promotor de *speB*, no puede sintetizar arginina y restaurar el crecimiento de la mutante *argC*$^{-}$ (Fig. 4E, 4F). El incremento en el número de copias del plásmido está probablemente relacionado a la compensación de la dosis génica. Lind *et al.* (Lind *et al.* 2010), trabajando con reemplazos de ortólogos de proteínas ribosomales en *Salmonella enterica* , encontraron que la amplificación génica es un mecanismo compensatorio para restaurar el decaimiento de la adaptación. En nuestras condiciones, a pesar del incremento en el número de copias, el crecimiento no fue completamente recuperado.

Lind *et al.* 2010 mostraron que la amplificación genética es importante para la transferencia horizontal y la evolución a nuevas funciones mediante duplicación genética.

Respecto a la transcripción, teóricamente los genes localizados bajo el mismo promotor tendrán los mismo niveles de expresión transcripcional. Sin embargo, la secuencia codificante parece modular su propia expresión transcripcional (Tablas 2 a 5; Fig. 14). Este fenómeno puede ser atribuido a la diferencia y disponibilidad del uso de codones (Indice de adaptación de codones (CAI)), la distribución de codones (Cannarozzi *et al*. 2010), la interacción de factores de traducción con la RNA-polimerasa y Ribosomas (Burmann *et al.* 2010, Proshkin *et al.* 2010), y la formación de estructuras en el mRNA que pudieran afectar la velocidad del complejo de transcripción. Otros autores han hecho observaciones similares a nivel de transcripción y traducción (Kudla *et al.* 2009). Por otro lado, se reportó también que las primeras 100 pb de la región codificante tiene un papel importante en la modulación de la transcripción (Tuller *et al.* 2010). Un efecto similar se observó en las fusiones transcripcionales en diferentes segmentos de *argC* (Tabla 2), y también cuando fue analizada la expresión del mRNA por PCR en tiempo real (Tabla 5, figura 14), donde se

encontraron diferentes concentraciones de mensajero de los ortólogos *argC* bajo el mismo promotor.

Los factores que afectan la expresión, transcripción y traducción de *argC* debido a las diferencias particulares en la secuencia de nucleótidos y aminoácidos (la FDE), consecuentemente afectan la actividad enzimática de ArgC (Tabla 4) y también tienen repercusión en el crecimiento (Fig. 4) y excreción de metabolitos en el medio de cultivo (Tabla 3). La disminución en la tasa de crecimiento esta relacionada a los metabolitos intermediarios que no pueden ser asimilados completamente y son excretados en grandes cantidades (Encarnación *et al.* 1995).

Para análizar el índice de eficiencia traduccional de cada ortólogo de *argC* en el fondo genético de *S. meliloti argC*$^{-}$, se cuantificaron directamente los niveles de expresión de mRNA y proteína. Se utilizó como referencia a la cepa silvestre de *S. meliloti*, asumiendo que ésta posee la mejor eficiencia traduccional de ArgC. Los valores constantes y la reproducibilidad de la cepa de referencia nos permitió estimar la confiabilidad de nuestro análisis (Tabla 5). En la complementación de *S. meliloti* bajo el promotor de *speB*, se incrementó en 5 veces la cantidad de mRNA, con respecto a la cepa silvestre. Sin embargo este valor no correlaciona directamente con el número de copias encontrado (2.9), lo que se debe posiblemente a algún factor que pudiera estar influenciando la tasa de transcripción, tal como el superenrrollamiento del replicón (Figueroa & Bossi 1988).

En las complementaciones hechas bajo el promotor *lac*, se obtuvieron las más grandes cantidades de transcrito, pero no se obtuvieron los mas altos valores en los índices de eficiencia traduccional. Al parecer, aunque estén presentes grandes cantidades de mRNA, existe un límite traduccional en la célula. En estas condiciones la cepa complementada con *argC* de *R. etli* mostró los más bajos índices de eficiencia traduccional, pero fueron suficientes para dar una adecuada recuperación del crecimiento (Fig. 4C y D). Bajo el promotor de *speB*, aunque la cepa complementada mostró más copias de *argC* de *M. loti*, esta cantidad no fue suficiente para sostener su

crecimiento, debido posiblemente a su bajo índice de eficiencia traduccional (Tabla 5, Fig. 4E y F).

Es importante hacer notar que en estos ensayos se encontró poca correlación entre la síntesis de proteína y la actividad enzimática. Este fenómeno de poca correlación fue más notorio en las complementaciones con el promotor *lac*. Creemos que existen factores fisiológicos que limitan el alcanzar altos niveles de actividad enzimática y que esta limitación fue ejercida sobre los *argC* heterólogos utilizados. No así, en la complementación con *argC* de *S. meliloti* bajo el promotor *lac*, que mostró altos niveles de actividad enzimática ArgC. Varios factores enzimáticos y fisiológicos pueden participar y alterar la correlación entre transcripción y traducción (tales como la estequiometría, cofactores, reguladores, enzimas de la ruta, flujo de energía, etc.) (Gu *et al.* 2014, Gingold & Pilpel 2011). Otra consideración importante es que la actividad fue medida *in vitro* por lo que esta condición no puede reflejar completamente la alteración fisiológica de la expresión de los ortólogos *argC*. Consideramos de este modo, que el crecimiento es la medida más fehaciente de analizar lo que ocurre con las complementaciones ortólogas.

Por otro lado, los datos obtenidos en *E. coli* correlacionan con los datos observados en las complementaciones llevadas a cabo en *Rhizobium*. Donde en un contexto genómico intacto, pudimos observar afectaciones en el crecimiento, niveles de expresión transcripcional y actividad enzimática (Fig. 13, Fig. 14, Fig. 15). *E. coli* es un organismo con marcadas diferencias ecológicas, metabólicas y genéticas con respecto a *Rhizobium*, donde existen varios factores que limitan la expresión de los ortólogos *argC* de *Rhizobium*, tales como el uso de codones, sin embargo los cambios observados pueden ser atribuidos a las secuencias propias de cada especie, o sea la FDE.

Asimismo, resulta interesante sugerir que las proteínas de organismos muy cercanos (*Rhizobium*) deberían tener un comportamiento semejante al ser expresadas en un organismo distante (*E. coli*), sin embargo, como se observó (Fig. 13, Fig. 14 y Fig. 15),

cada uno de los reemplazos y complementaciones llevadas a cabo en *E. coli* exhiben características particulares que están directamente relacionadas a la secuencia de los ortólogos complementantes.

Las proteínas ortólogas de *S. meliloti*, *R. etli* y *A. tumefaciens* exhiben cantidades similares de diferencias en FDE (~10 %), lo cual indica su proximidad filogenética. Mientras que las secuencias de *M. loti* y *B. melitensis* están evolutivamente más distantes (Fig. 2), por lo que las diferencias en la FDE son más altas (36 %) (Fig. 1). Con el propósito de definir el cluster de especies y su conservación de la FDE, se secuenciaron los genes *argC* de las cepas de cada especie (a excepción de *B. melitensis*). El análisis de estas secuencias resultó en clusters muy definidos y con una alta conservación (Fig. 2). La organización genómica de *argC* de Rhizobiales fue muy inusual al estar contiguo al gen *speB* (de la ruta de biosíntesis de poliaminas).

Se caracterizó también la expresión transcripcional de cada uno de los *argC* con sus promotores nativos (sus propios promotores), en sus propios fondos genéticos nativos, y también en *S. meliloti* y *M. loti*. Esto, con el propósito de conocer los niveles de expresión transcripcional típicos de cada especie y además poder analizar su expresión en diferentes fondos genéticos como son *S. meliloti* y *M. loti*.
Se observó que cada *argC* de las especies analizadas exhibe diferentes niveles de expresión transcripcional en su propio fondo genético. Y al ser intercambiados en el fondo genético de *S. meliloti* muestran un patrón de expresión similar al propio, a excepción de *argC* de *M. loti,* donde se obtienen niveles de expresión basal. Por otro lado, la expresión de los *argC* en el fondo genético de *M. loti* fue alta, en todos los *argC*.

Los datos de expresión transcripcional indican que cada *argC* posee diferencias en sus niveles de expresión y regulación transcripcional. Estas diferencias fueron más notorias al intercambiar los *argC* al fondo genético de *M. loti*, donde se observaron altos niveles de expresión (Fig. 3). Por otro lado, los altos niveles de expresión en el fondo de *M. loti*

versus *S. meliloti* podrían sugerir la probable participación de un represor en este último organismo.

Los genes ortólogos por lo general codifican para la misma función pero también contienen características que reflejan adaptaciones específicas (Keynon &Sabree 2014, Kelkar & Ochman 2013, Duplantis *et al.* 2011). Por ejemplo, en el endosimbionte *Buchnera* han sido observado cambios en expresión de chaperonas como GroEL, donde la expresión constitutiva de GroEL podría compensar deficiencias en el plegado y estabilidad de las proteínas ortólogas (Fares *et al.* 2004).

Otro fenómeno observado en ortólogos de endosimbiontes bacterianos, es que la reducción de su genoma aumenta la complejidad funcional de sus proteínas ortólogas en respuesta adaptativa al nicho (Kelkar & Ochman 2013). Por otro lado, las bacterias que crecen a bajas temperaturas (bacterias psicrofílicas), tienen varias adaptaciones, las cuales incluyen insaturación de ácidos grasos y un mayor número de chaperonas para resolver la estructura secundaria de ácidos nucleicos (Deming 2002). Respecto a sus enzimas, es probable que casi todas (incluyendo aquellas con funciones esenciales), tengan adaptaciones al ambiente frío, reduciendo sus K_m (constante de Michaelis) o aumentando su K_{cat} (Número de recambio), o ambos parámetros (Feller & Gerday 1997, Feller & Gerday 2003, Georlette *et al.* 2004). Casi siempre estas enzimas de bacterias psicrofílicas son susceptibles a desnaturalización a bajas temperaturas, lo cual puede reflejar cambios estructurales necesarios para optimizar su función en ambientes frios, o podría ser el resultado de deriva genética en ausencia de selección a altas temperaturas.

Por otro lado, existen varios factores que pueden afectar el desempeño de cada ortólogo, tales como la cinética enzimática, su propia regulación transcripcional, la disponibilidad del sustrato, la interacción con otras proteínas y metabolitos, la concentración de iones y cofactores, así como sus tasas de replicación y traducción (Goryanin *et al.* 2006, Khersonsky & Tawfik 2010, Phair 1997, Zhao *et al.* 2008). En ese

sentido, durante el proceso evolutivo las células optimizan la eficiencia de traducción de sus diferentes genes a diferentes niveles. Algunos productos son necesarios en más cantidades que otros. Otros requieren de un recambio constante y deben ser traducidos a diferentes niveles y en diferentes condiciones (Takagi *et al.* 2005; Lu 2006, Ingolia *et al.* 2009, Gingold & Pilpel 2011). Por ello, la consiguiente expresión genética no debe ser concebida como una condición estable, sino más bien como un factor central celular que responde a cambios y perturbaciones intracelulares y del medio.

De este modo, la eficiencia y coordinación de la expresión genética requiere de balance entre costo y rendimiento (Dekel & Alon 2005). Los costos del proceso son numerosos e implican el uso de compuestos intermediarios, energía y asignación de recursos celulares tales como ribosomas y tRNAs (Stoebelet *et al*. 2008). La aparente redundancia del código genético en el que más aminoácidos pueden ser traducidos por mas de un codón, es otro factor que puede influenciar la expresión de una proteína, ya que permite a la célula sintonizar a varios niveles con eficiencia y precisión la producción de proteínas, mientras mantiene la misma secuencia de aminoácidos. Respecto a los codones que corresponden al mismo aminoácido (codones sinónimos), sus tRNAs correspondientes pueden diferir en cantidades en la célula y también la velocidad en la que son reconocidos por el ribosoma (Varenne *et al.* 1984, Sorensen *et al.* 1989, Sorensen *et al.* 1991). La estructura secundaria de la región codificante también puede influenciar la estabilidad y afectar la traducción de la proteína(Kudla *et al.* 2009) e incluso el plegado correcto de las proteínas (Komar *et al*. 1999, Kimchi-Sarfaty *et al*. 2007). Los datos experimentales a escala genómica, *in vitro* (Li *et al.* 2012, Kertesz *et al*. 2010, Wan *et al.* 2012, Zheng *et al.* 2010, Ding *et al.* 2014) e *in vivo* han confirmado el rol regulatorio de la estructura secundaria del mRNA.

En un estudio para dilucidar la relación entre el efecto de la conservación de la estructura del mRNA y la evolución de la secuencia codificante, se encontró que los genes más expresados presentan más restricciones en la conservación de su estructura secundaría del mRNA, y además se sugiere que el mRNA esta sujeto a las

fuerzas de selección purificadoras para conservar su estructura secundaria (Gu *et al.* 2014).

Los genes ortólogos por su compromiso con las funciones esenciales se encuentran bajo presión de selección, sin embargo al parecer éstos también reflejan el ambiente en el cual han sido seleccionados. El intercambio del ortólogo *argC* en este trabajo, resultó en una serie de fenómenos no esperados, todos atribuidos directamente a la secuencia de la proteína y la selección de la misma. Por lo consiguiente, aunque los ortólogos cercanos analizados cumplen la misma función, no son totalmente sustituibles.

En este trabajo se demostró que la complementación con genes ortólogos de especies cercanas puede exhibir cambios inusuales, tanto en complejidad como en reajustes celulares en respuesta a la secuencia de las proteínas ortólogas introducidas, y que varios factores participan en la eficiencia de complementación de cada ortólogo analizado. Además, se reportó un nuevo promotor para *argC* en Rhizobiales, cuya presencia no había sido predicha por métodos informáticos, además, se demostró que la región codificante participa en la regulación de la transcripción y la eficiencia de la traducción, en concordancia con trabajos recientes (Fredrick & Ibba 2010, Kudla *et al.* 2009).

Conclusiones.

•Los genes ortólogos por su compromiso con las funciones esenciales se encuentran bajo presión de selección, sin embargo, al parecer éstos también reflejan el ambiente en el cual han sido seleccionados. El intercambio de *argC* en este trabajo, resultó en una serie de fenómenos no esperados, todos atribuidos directamente a la FDE.

•La FDE refleja las particularidades evolutivas de cada especie en respuesta a un microambiente específico, demostrando que aunque los ortólogos cercanos analizados cumplen la misma función, no son totalmente sustituibles.

•Se observó un gradiente en la capacidad de complementación con el gen ortólogo *argC* en *S. meliloti*, desde la secuencia más cercana, perteneciente a *R. etli*, seguida por la de *A. tumefaciens* y finalmente la de *M. loti*, que no fue capaz de sostener el crecimiento de la cepa.

•Se encontró que se activaron una serie de mecanismos de expresión genética, metabólicos y fisiológicos en la bacteria para compensar los cambios de secuencia en el gen ortólogo.

•La sustitución de ortólogos de *argC* en *S. meliloti* tuvo repercusiones globales en el número de copias de los genes, tasa transcripcional, síntesis de RNA mensajero y traducción, actividad enzimática, fisiología, crecimiento y excreción de metabolitos.

•Datos semejantes fueron observados en *E. coli*, donde cada ortólogo introducido exhibió características particulares, todas atribuidas a la secuencia de los ortólogos sustitutos.

•Entre especies cercanas, los cambios metabólicos y de expresión pueden reflejar adaptación a nichos diferentes y por lo consiguiente, particularidades metabólicas y de expresión propias de cada especie. Concluimos que estas particularidades también se encuentran en la parte mas conservada del genoma, los genes ortólogos cromosomales esenciales.

Referencias:

Abdelal, A. T. (1979). Arginine catabolism by microorganisms. *Annual Reviews in Microbiology*, *33*(1), 139-168.

Adler, M., Anjum, M., Berg, O. G., Andersson, D. I., & Sandegren, L. (2014). High fitness costs an instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. *Molecular biology and evolution*, *31*(6), 1526-1535.

Altenhoff, A. M., Studer, R. A., Robinson-Rechavi, M., & Dessimoz, C. (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comput Biol*, *8*(5), e1002514-e1002514.

Amor, B. B., Shaw, S. L., Oldroyd, G. E., Maillet, F., Penmetsa, R. V., Cook, D., ... & Gough, C. (2003). The NFP locus of Medicago truncatula controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *The Plant Journal*, *34*(4), 495-506. Arrighi, J.-F. et al., 2006. The *Medicago truncatula* lysin [corrected] motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant physiology, 142(1), pp.265–79.

Ax, P. (1987). The phylogenetic system: the systematization of organisms on the basis of their phylogenesis.

Baltrus, D. A., Nishimura, M. T., Romanchuk, A., Chang, J. H., Mukhtar, M. S., Cherkis, K., ... & Dangl, J. L. (2011). Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 Pseudomonas syringae isolates. *PLoS pathog*, *7*(7), e1002132.

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in ecology & evolution*, *28*(8), 489-495.

Battisti, L., Lara, J. C., & Leigh, J. A. (1992). Specific oligosaccharide form of the Rhizobium meliloti exopolysaccharide promotes nodule invasion in alfalfa. *Proceedings of the National Academy of Sciences*, *89*(12), 5625-5629.

Beny, G., Cunin, R., Glansdorff, N., Boyen, A., Charlier, J., & Kelker, N. (1982). Transcription of Regions within the divergent argECBH operon of Escherichia coli: evidence for lack of an attenuation mechanism. *Journal of bacteriology*, *151*(1), 58-61.

Bergthorsson, U., Andersson, D. I., & Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences*, *104*(43), 17004-17009.

Burgetz, I. J., Shariff, S., Pang, A., & Tillier, E. R. (2006). Positional homology in bacterial genomes. *Evolutionary Bioinformatics Online*, *2*, 77.

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E., & Rösch, P. (2010). A NusE: NusG complex links transcription and translation. *Science*, *328*(5977), 501-504.

Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., ... & Barral, Y. (2010). A role for codon order in translation dynamics. *Cell*, *141*(2), 355-367.

Cárdenas, L., Dominguez, J., Santana, O., & Quinto, C. (1996). The role of the nodI and nodJ genes in the transport of Nod metabolites in Rhizobium etli. *Gene*, *173*(2), 183-187.

Catalano, C. M., Lane, W. S., & Sherrier, D. J. (2004). Biochemical characterization of symbiosome membrane proteins from Medicago truncatula root nodules. *Electrophoresis*, *25*(3), 519-531.

Catoira, R., Galera, C., de Billy, F., Penmetsa, R. V., Journet, E. P., Maillet, F., ... & Dénarié, J. (2000). Four genes of Medicago truncatula controlling components of a Nod factor transduction pathway. *The Plant Cell*, *12*(9), 1647-1665.

Chang, A. Y. F., & Liao, B. Y. (2011). DNA methylation rebalances gene dosage after mammalian gene duplications. *Molecular biology and evolution*, msr174.

Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., & Parniske, M. (2008). Lotus japonicus CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *The Plant Cell*, *20*(12), 3467-3479.

Chen, F. C., Chen, C. J., Li, W. H., & Chuang, T. J. (2010). Gene family size conservation is a good indicator of evolutionary rates. *Molecular biology and evolution*, *27*(8), 1750-1758.

Cheng, F., Wu, J., Fang, L., & Wang, X. (2012). Syntenic gene analysis between Brassica rapa and other Brassicaceae species. *Frontiers in plant science*, *3*.

Citerne, H. L., Le Guilloux, M., Sannier, J., Nadot, S., & Damerval, C. (2013). Combining phylogenetic and syntenic analyses for understanding the evolution of TCP ECE genes in eudicots.

Conant, G. C., & Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics*, *9*(12), 938-950.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., ... & Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, *464*(7289), 704-712.

Crow, K. D., & Wagner, G. P. (2006). What is the role of genome duplication in the evolution of complexity and diversity?. *Molecular biology and evolution*, *23*(5), 887-892.

Cullimore, J. V., Ranjeva, R., & Bono, J. J. (2001). Perception of lipo-chitooligosaccharidic Nod factors in legumes. *Trends in plant science*, *6*(1), 24-30.

Cunin, R., Glansdorff, N., Pierard, A., & Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiological reviews*, *50*(3), 314.

Cusack, B. P., & Wolfe, K. H. (2007). Not born equal: increased rate asymmetry in relocated and retrotransposed rodent gene duplicates. *Molecular biology and evolution*, *24*(3), 679-686.

Dagan, T., Artzy-Randrup, Y., & Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences*, *105*(29), 10039-10044.

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, *97*(12), 6640-6645.

De Jong, A. J., Heidstra, R., Spaink, H. P., Hartog, M. V., Meijer, E. A., Hendriks, T., ... & De Vries, S. C. (1993). Rhizobium lipooligosaccharides rescue a carrot somatic embryo mutant. *The Plant Cell*, *5*(6), 615-620.

Debellé, F., Plazanet, C., Roche, P., Pujol, C., Savagnac, A., Rosenberg, C., ... & Dénarié, J. (1996). The NodA proteins of Rhizobium meliloti and Rhizobium tropici specify the N-acylation of Nod factors by different fatty acids. *Molecular microbiology*, *22*(2), 303-314.

Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, *436*(7050), 588-592.

Deming, J. W. (2002). Psychrophiles and polar regions. *Current opinion in microbiology*, *5*(3), 301-309.

Dénarié, J., & Cullimore, J. (1993). Lipo-oligosaccharide nodulation factors: a new class of signaling molecules mediating recognition and morphogenesis. *Cell*, *74*(6), 951-954.

Dénarié, J., Debellé, F., & Promé, J. C. (1996). Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annual review of biochemistry*, *65*(1), 503-535.

Dewey, C. N. (2011). Positional orthology: putting genomic evolutionary relationships into context. *Briefings in bioinformatics*, *12*(5), 401-412.

Di Martino, M. L., Campilongo, R., Casalino, M., Micheli, G., Colonna, B., & Prosseda, G. (2013). Polyamines: emerging players in bacteria–host interactions. *International Journal of Medical Microbiology*, *303*(8), 484-491.

Dickstein, R., Bisseling, T., Reinhold, V. N., & Ausubel, F. M. (1988). Expression of nodule-specific genes in alfalfa root nodules blocked at an early stage of development. *Genes & development*, *2*(6), 677-687.

Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, *505*(7485), 696-700.

Dittmar, K., & Liberles, D. (2011). *Evolution after gene duplication*. John Wiley & Sons.

Dolinski, K., & Botstein, D. (2007). Orthology and functional conservation in eukaryotes. *Annu. Rev. Genet.*, *41*, 465-507.

Duplantis, B. N., Bosio, C. M., & Nano, F. E. (2011). Temperature-sensitive bacterial pathogens generated by the substitution of essential genes from cold-loving bacteria: potential use as live vaccines. *Journal of molecular medicine*, *89*(5), 437-444.

Dylan, T. Y. L. E. R., Nagpal, P. U. N. I. T. A., Helinski, D. R., & Ditta, G. S. (1990). Symbiotic pseudorevertants of Rhizobium meliloti ndv mutants. *Journal of bacteriology*, *172*(3), 1409-1417.

Elhaik, E., Landan, G., & Graur, D. (2009). Can GC content at third-codon positions be used as a proxy for isochore composition?. *Molecular biology and evolution*, *26*(8), 1829-1833.

Encarnacion, S., Dunn, M., Willms, K., & Mora, J. (1995). Fermentative and aerobic metabolism in Rhizobium etli. *Journal of bacteriology*, *177*(11), 3058-3066.

Epstein, B., Sadowsky, M. J., & Tiffin, P. (2014). Selection on horizontally transferred and duplicated genes in Sinorhizobium (Ensifer), the root-nodule symbionts of Medicago. *Genome biology and evolution*, *6*(5), 1199-1209.

Fares, M. A., Moya, A., & Barrio, E. (2004). GroEL and the maintenance of bacterial endosymbiosis. *TRENDS in Genetics*, *20*(9), 413-416.

Feller, G., & Gerday, C. (1997). Psychrophilic enzymes: molecular basis of cold adaptation. *Cellular and Molecular Life Sciences CMLS*, *53*(10), 830-841.

Feller, G., & Gerday, C. (2003). Psychrophilic enzymes: hot topics in cold adaptation. *Nature Reviews Microbiology*, *1*(3), 200-208.

Fernández-López, M., D'Haeze, W., Mergaert, P., Verplancke, C., Prome, J. C., Montagu, M., & Holsters, M. (1996). Role of nodl and nodj in lipo-chitooligosaccharide secretion in Azorhizobium caulinodans and Escherichia coli. *Molecular microbiology*, *20*(5), 993-1000.

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic Biology*, *19*(2), 99-113.

Fitch, W. M. (2000). Homology: a personal view on some of the problems. *Trends in genetics*, *16*(5), 227-231.

Figueroa, N., & Bossi, L. (1988). Transcription induces gyration of the DNA template in Escherichia coli. *Proceedings of the National Academy of Sciences*, *85*(24), 9416-9420.

Flavell, R. B. (1994). Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proceedings of the National Academy of Sciences*, *91*(9), 3490-3496.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, *269*(5223), 496-512.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., ... & Venter, J. C. (1995). The minimal gene complement of Mycoplasma genitalium. *Science*, *270*(5235), 397-404.

Fredrick, K., & Ibba, M. (2010). How the sequence of a gene can tune its translation. *Cell*, *141*(2), 227-229.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, *151*(4), 1531-1545.

Gabaldón, T., & Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, *14*(5), 360-366.

Gage, D. J. (2002). Analysis of infection thread development using Gfp-and DsRed-expressing Sinorhizobium meliloti. *Journal of Bacteriology*, *184*(24), 7042-7046.

Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and Molecular Biology Reviews*, *68*(2), 280-300.

Geremia, R. A., Mergaert, P., Geelen, D., Van Montagu, M., & Holsters, M. (1994). The NodC protein of Azorhizobium caulinodans is an N-acetylglucosaminyltransferase. *Proceedings of the National Academy of Sciences*, *91*(7), 2669-2673.

Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., ... & Gerday, C. (2004). Some like it cold: biocatalysis at low temperatures. *FEMS microbiology reviews*, *28*(1), 25-42.

Geurts, R., Fedorova, E., & Bisseling, T. (2005). Nod factor signaling genes and their function in the early stages of Rhizobium infection. *Current opinion in plant biology*, *8*(4), 346-352.

Gingold, H., & Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. *Molecular systems biology*, *7*(1), 481.

Glazebrook, J., & Walker, G. C. (1989). A novel exopolysaccharide can function in place of the calcofluor-binding exopolysaccharide in nodulation of alfalfa by Rhizobium meliloti. *Cell*, *56*(4), 661-672.

González, J. E., Reuhs, B. L., & Walker, G. C. (1996). Low molecular weight EPS II of Rhizobium meliloti allows nodule invasion in Medicago sativa. *Proceedings of the National Academy of Sciences*, *93*(16), 8636-8641.

Goryanin, I. I., Lebedeva, G. V., Mogilevskaya, E. A., Metelkin, E. A., & Demin, O. V. (2006). Cellular kinetic modeling of the microbial metabolism. *Microbial Proteomics: Functional Biology of Whole Organisms*, 437-488.

Gu, W., Zhou, T., & Wilke, C. O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol*, *6*(2), e1000664-e1000664.

Gu, W., Li, M., Xu, Y., Wang, T., Ko, J. H., & Zhou, T. (2014). The impact of RNA structure on coding sequence evolution in both bacteria and eukaryotes. *BMC evolutionary biology*, *14*(1), 87.

Gu, Z., Nicolae, D., Lu, H. H., & Li, W. H. (2002). Rapid divergence in expression between duplicate genes inferred from microarray data. *TRENDS in Genetics*, *18*(12), 609-613.

Guerrero, G., Peralta, H., Aguilar, A., Díaz, R., Villalobos, M. A., Medrano-Soto, A., & Mora, J. (2005). Evolutionary, structural and functional relationships revealed by comparative analysis of syntenic genes in Rhizobiales. *BMC evolutionary biology*, *5*(1), 55.

Hassan, S., & Mathesius, U. (2012). The role of flavonoids in root–rhizosphere signalling: opportunities and challenges for improving plant–microbe interactions. *Journal of experimental botany*, *63*(9), 3429-3444.

Herridge, D. F., Peoples, M. B., & Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*, *311*(1-2), 1-18.

Hirsch, A. M. (1992). Developmental biology of legume nodulation. *New Phytologist*, *122*(2), 211-237.

Hovel-Miner, G., Faucher, S. P., Charpentier, X., & Shuman, H. A. (2010). ArgR-regulated genes are derepressed in the Legionella-containing vacuole. *Journal of bacteriology*, *192*(17), 4504-4516.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London B: Biological Sciences*, *256*(1346), 119-124.

Huynen, M., Snel, B., Lathe, W., & Bork, P. (2000). Exploitation of gene context. *Current opinion in structural biology*, *10*(3), 366-370.

Igarashi, K., & Kashiwagi, K. (2010). Modulation of cellular function by polyamines. *The international journal of biochemistry & cell biology*, *42*(1), 39-51.

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, *11*(2), 97-108.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *science*, *324*(5924), 218-223.

Itoh, Y., & Nakada, Y. (2004). Arginine and polyamine metabolism. In *Pseudomonas* (pp. 243-272). Springer US.

Iwabe, N., Kuma, K. I., Hasegawa, M., Osawa, S., & Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences*, *86*(23), 9355-9359.

Jofré, E., & Becker, A. (2009). Production of succinoglycan polymer in Sinorhizobium meliloti is affected by SMb21506 and requires the N-terminal domain of ExoP. *Molecular plant-microbe interactions*, *22*(12), 1656-1668.

Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., & Walker, G. C. (2007). How rhizobial symbionts invade plants: the Sinorhizobium–Medicago model. *Nature Reviews Microbiology*, *5*(8), 619-633.

Jordan, D. C. (1984). Family III. Rhizobiaceae. *Bergey's manual of systematic bacteriology*, *1*, 234-242.

Jun, J., Ryvkin, P., Hemphill, E., & Nelson, C. (2009). Duplication mechanism and disruptions in flanking regions determine the fate of Mammalian gene duplicates. *Journal of Computational Biology*, *16*(9), 1253-1266.

Jun, J., Mandoiu, I. I., & Nelson, C. E. (2009). Identification of mammalian orthologs using local synteny. *BMC genomics*, *10*(1), 630.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome research*, *20*(10), 1313-1326.

Kamst, E., Bakkers, J., Quaedvlieg, N. E., Pilling, J., Kijne, J. W., Lugtenberg, B. J., & Spaink, H. P. (1999). Chitin oligosaccharide synthesis by rhizobia and zebrafish embryos starts by glycosyl transfer to O4 of the reducing-terminal residue. *Biochemistry*, *38*(13), 4045-4052.

Kamst, E., Pilling, J., Raamsdonk, L. M., Lugtenberg, B. J., & Spaink, H. P. (1997). Rhizobium nodulation protein NodC is an important determinant of chitin oligosaccharide chain length in Nod factor biosynthesis. *Journal of bacteriology*, *179*(7), 2103-2108.

Kelkar, Y. D., & Ochman, H. (2013). Genome reduction promotes increase in protein functional complexity in bacteria. *Genetics*, *193*(1), 303-307.

Keller, T. E., & Soojin, V. Y. (2014). DNA methylation and evolution of duplicate genes. *Proceedings of the National Academy of Sciences*, *111*(16), 5932-5937.

Kemkemer, C., Kohn, M., Cooper, D. N., Froenicke, L., Högel, J., Hameister, H., & Kehrer-Sawatzki, H. (2009). Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC evolutionary biology*, *9*(1), 84.

Kenyon, L. J., & Sabree, Z. L. (2014). Obligate insect endosymbionts exhibit increased ortholog length variation and loss of large accessory proteins concurrent with genome shrinkage. *Genome biology and evolution*, *6*(4), 763-775.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., & Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature*, *467*(7311), 103-107.

Khersonsky, O., and D. S. Tawfik. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. 79:11.1–11.35.

Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., & Gottesman, M. M. (2007). A" silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, *315*(5811), 525-528.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624-626.

Kimura, M. (1984). *The neutral theory of molecular evolution*. Cambridge University Press.

Klarsfeld, A. D., Goossens, P. L., & Cossart, P. (1994). Five Listeria monocytogenes genes preferentially expressed in infected mammalian cells: plcA, purH, purD, pyrE and an arginine ABC transporter gene, arpJ. *Molecular microbiology*, *13*(4), 585-597.

Klassen, J. L. (2009). Pathway evolution by horizontal transfer and positive selection is accommodated by relaxed negative selection upon upstream pathway genes in purple bacterial carotenoid biosynthesis. *Journal of bacteriology*, *191*(24), 7500-7508.

Kleinjan, D. J., & Van Heyningen, V. (1998). Position effect in human genetic disease. *Human molecular genetics*, *7*(10), 1611-1618.

Kloesges, T., Popa, O., Martin, W., & Dagan, T. (2011). Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Molecular biology and evolution*, *28*(2), 1057-1074.

Kloosterman, T. G., & Kuipers, O. P. (2011). Regulation of arginine acquisition and virulence gene expression in the human pathogen Streptococcus pneumoniae by transcription regulators ArgR1 and AhrC. *Journal of Biological Chemistry*, *286*(52), 44594-44605.

Komar, A. A., Lesnik, T., & Reiss, C. (1999). Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS letters*, *462*(3), 387-391.

Koonin, E. V., Makarova, K. S., & Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification 1. *Annual Reviews in Microbiology*, *55*(1), 709-742.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics 1. *Annu. Rev. Genet.*, *39*, 309-338.

Koski, L. B., Morton, R. A., & Golding, G. B. (2001). Codon bias and base composition are poor indicators of horizontally transferred genes. *Molecular Biology and Evolution*, *18*(3), 404-412.

Król, J. E., Mazur, A., Marczak, M., & Skorupska, A. (2007). Syntenic arrangements of the surface polysaccharide biosynthesis genes in Rhizobium leguminosarum. *Genomics*, *89*(2), 237-247.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *science*, *324*(5924), 255-258.

Kumada, Y., Benson, D. R., Hillemann, D., Hosted, T. J., Rochefort, D. A., Thompson, C. J., ... & Tateno, Y. (1993). Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proceedings of the National Academy of Sciences*, *90*(7), 3009-3013.

Larimer, F. W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., ... & Harwood, C. S. (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nature biotechnology*, *22*(1), 55-61.

Laus, M. C., Logman, T. J., Lamers, G. E., Van Brussel, A. A., Carlson, R. W., & Kijne, J. W. (2006). A novel polar surface polysaccharide from Rhizobium leguminosarum binds host plant lectin. *Molecular microbiology*, *59*(6), 1704-1713.

Le, P. T., Pontarotti, P., & Raoult, D. (2014). Alphaproteobacteria species as a source and target of lateral sequence transfers. *Trends in microbiology*, *22*(3), 147-156.

Lemoine, F., Lespinet, O., & Labedan, B. (2007). Assessing the evolutionary rate of positional orthologous genes in prokaryotes using synteny data. *BMC evolutionary biology*, *7*(1), 237.

Levin, B. R., & Bergstrom, C. T. (2000). Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proceedings of the National Academy of Sciences*, *97*(13), 6981-6985.

Li, F., Zheng, Q., Vandivier, L. E., Willmann, M. R., Chen, Y., & Gregory, B. D. (2012). Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. *The Plant Cell*, *24*(11), 4346-4359.

Li, H., Li, M., Huang, Y., Rensing, C., & Wang, G. (2013). In silico analysis of bacterial arsenic islands reveals remarkable synteny and functional relatedness between arsenate and phosphate. *Frontiers in microbiology*, *4*.

Lian, B., Prithiviraj, B., Souleimanov, A., & Smith, D. L. (2001). Evidence for the production of chemical compounds analogous to nod factor by the silicate bacterium Bacillus circulans GY92. *Microbiological research*, *156*(3), 289-292.

Liberles, D. A., Kolesov, G., & Dittmar, K. (2010). Understanding gene duplication through biochemistry and population genetics. *Evolution after gene duplication. Hoboken (NJ): John Wiley & Sons*, 10-11.

Lima, W. C., & Menck, C. F. (2008). Replacement of the arginine biosynthesis operon in Xanthomonadales by lateral gene transfer. *Journal of molecular evolution*, *66*(3), 266-275.

Lind, P. A., Tobin, C., Berg, O. G., Kurland, C. G., & Andersson, D. I. (2010). Compensatory gene amplification restores fitness after inter-species gene replacements. *Molecular microbiology*, *75*(5), 1078-1089.

Llácer, J. L., Fita, I., & Rubio, V. (2008). Arginine and nitrogen storage. *Current opinion in structural biology*, *18*(6), 673-681.

Lloret, L., & Martínez-Romero, E. (2005). Evolución y filogenia de Rhizobium. *Rev. Latinoam. Microbiol*, *47*(1-2), 43-60.

Lu, C. D. (2006). Pathways and regulation of bacterial arginine metabolism and perspectives for obtaining arginine overproducing strains. *Applied microbiology and biotechnology*, *70*(3), 261-272.

Lu, C. D., Yang, Z., & Li, W. (2004). Transcriptome analysis of the ArgR regulon in Pseudomonas aeruginosa. *Journal of bacteriology*, *186*(12), 3855-3861.

Lüthi, E., Baur, H., Gamper, M., Brunner, F., Villeval, D., Mercemier, A., & Haas, D. (1990). The arc operon for anaerobic arginine catabolism in Pseudomonas aeruginosa contains an additional gene, arcD, encoding a membrane protein. *Gene*, *87*(1), 37-43.

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, *290*(5494), 1151-1155.

Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics*, *154*(1), 459-473.

Maas, W. K. (1994). The arginine repressor of Escherichia coli. *Microbiological reviews*, *58*(4), 631-640.

Marc, F., Weigel, P., Legrain, C., Almeras, Y., Santrot, M., Glansdorff, N., & Sakanyan, V. (2000). Characterization and kinetic mechanism of mono-and bifunctional ornithine acetyltransferases from thermophilic microorganisms. *European Journal of Biochemistry*, *267*(16), 5217-5226.

Martínez-Romero, E. (2009). Coevolution in Rhizobium-legume symbiosis?. *DNA and cell biology*, *28*(8), 361-370.

McCune, A. R., & Schimenti, J. C. (2012). Using genetic networks and homology to understand the evolution of phenotypic traits. *Current genomics*, *13*(1), 74.

Mendrygal, K. E., & González, J. E. (2000). Environmental regulation of exopolysaccharide production in Sinorhizobium meliloti. *Journal of bacteriology*, *182*(3), 599-606.

.

Morizono, H., Cabrera-Luque, J., Shi, D., Gallegos, R., Yamaguchi, S., Yu, X., ... & Tuchman, M. (2006). Acetylornithine transcarbamylase: a novel enzyme in arginine biosynthesis. *Journal of bacteriology*, *188*(8), 2974-2982.

Näsvall, J., Sun, L., Roth, J. R., & Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. *Science*, *338*(6105), 384-387.

Nehrt, N. L., Clark, W. T., Radivojac, P., & Hahn, M. W. (2011). Testing the ortholog conjecture with comparative functional genomic data from mammals. *PLoS Comput Biol*, *7*(6), e1002073-e1002073.

Nei, M., & Roychoudhury, A. K. (1973). Probability of fixation of nonfunctional genes at duplicate loci. *American Naturalist*, 362-372.

Nester, E. (2000). DNA and protein transfer from bacteria to eukaryotes—the agrobacterium story. *Molecular plant pathology*, *1*(1), 87-90.

Noda-García, L., & Barona-Gómez, F. (2013). Enzyme evolution beyond gene duplication: a model for incorporating horizontal gene transfer. *Mobile genetic elements*, *3*(5), 2024-34.

Noda-García, L., Camacho-Zarco, A. R., Medina-Ruíz, S., Gaytán, P., Carrillo-Tripp, M., Fülöp, V., & Barona-Gómez, F. (2013). Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Molecular biology and evolution*, mst115.

Notebaart, R. A., Huynen, M. A., Teusink, B., Siezen, R. J., & Snel, B. (2005). Correlation between sequence conservation and the genomic context after gene duplication. *Nucleic acids research*, *33*(19), 6164-6171.

Nowell, R. W., Green, S., Laue, B. E., & Sharp, P. M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome biology and evolution*, *6*(6), 1514-1529.

Ohno, S. (1970). Evolution by Gene Duplication, Springer, Verlag. Berlin,Germany.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, 263-286.

Oldroyd, G. E., & Downie, J. A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.*, *59*, 519-546.

Oldroyd, G. E., & Downie, J. A. (2006). Nuclear calcium changes at the core of symbiosis signalling. *Current opinion in plant biology*, *9*(4), 351-357.

Owen, R. (1848). *On the archetype and homologies of the vertebrate skeleton*. van Voorst.

Pál, C., Papp, B., & Lercher, M. J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature genetics*, *37*(12), 1372-1375.

Pavlinov, I. Y. (2012). The contemporary concepts of homology in biology: A theoretical review. *Biology Bulletin Reviews*, *2*(1), 36-54.

Peck, M. C., Fisher, R. F., & Long, S. R. (2006). Diverse flavonoids stimulate NodD1 binding to nod gene promoters in Sinorhizobium meliloti. *Journal of bacteriology*, *188*(15), 5417-5427.

Pellock, B. J., Teplitski, M., Boinay, R. P., Bauer, W. D., & Walker, G. C. (2002). A LuxR homolog controls production of symbiotically active extracellular polysaccharide II by Sinorhizobium meliloti. *Journal of bacteriology*, *184*(18), 5067-5076.

Pellock, B. J., Cheng, H. P., & Walker, G. C. (2000). Alfalfa Root Nodule Invasion Efficiency Is Dependent on Sinorhizobium melilotiPolysaccharides. *Journal of bacteriology*, *182*(15), 4310-4318.

Peralta, H., Guerrero, G., Aguilar, A., & Mora, J. (2011). Sequence variability of Rhizobiales orthologs and relationship with physico-chemical characteristics of proteins. *Biol Direct*, *6*, 48.

Perret, X., Staehelin, C., & Broughton, W. J. (2000). Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews*, *64*(1), 180-201.

Phair, R. D. (1997). Development of kinetic models in the nonlinear world of molecular cell biology. *Metabolism*, *46*(12), 1489-1495.

Poupot, R., Martinez-Romero, E., Gautier, N., & Promé, J. C. (1995). Wild type Rhizobium etli, a bean symbiont, produces acetyl-fucosylated, N-methylated, and carbamoylated nodulation factors. *Journal of Biological Chemistry*, *270*(11), 6050-6055.

Prell, J., & Poole, P. (2006). Metabolic changes of rhizobia in legume nodules. *Trends in microbiology*, *14*(4), 161-168.

Proshkin, S., Rahmouni, A. R., Mironov, A., & Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, *328*(5977), 504-508.

Qu, Q., Morizono, H., Shi, D., Tuchman, M., & Caldovic, L. (2007). A novel bifunctional N-acetylglutamate synthase-kinase from Xanthomonas campestris that is closely related to mammalian N-acetylglutamate synthase. *BMC biochemistry*, *8*(1), 4.

Randhawa, G. S., & Hassani, R. (2002). Role of rhizobial biosynthetic pathways of amino acids, nucleotide bases and vitamins in symbiosis. *Indian journal of experimental biology*, *40*(7), 755-764.

Rauf, S., & Mir, A. (2013). Exploring evolution of brain genes involved in microcephaly through phylogeny and synteny analysis. *Theoretical Biology and Medical Modelling*, *10*(1), 61.

Reinhold, B. B., Chan, S. Y., Reuber, T. L., Marra, A., Walker, G. C., & Reinhold, V. N. (1994). Detailed structural characterization of succinoglycan, the major exopolysaccharide of Rhizobium meliloti Rm1021. *Journal of bacteriology*, *176*(7), 1997-2002.

Reuber, T. L., & Walker, G. C. (1993). Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of Rhizobium meliloti. *Cell*, *74*(2), 269-280.

Rodin, S. N., & Riggs, A. D. (2003). Epigenetic silencing may aid evolution by gene duplication. *Journal of molecular evolution*, *56*(6), 718-729.

Röhrig, H., Schmidt, J., Wieneke, U., Kondorosi, E., Barlier, I., Schell, J., & John, M. (1994). Biosynthesis of lipooligosaccharide nodulation factors: Rhizobium NodA protein is involved in N-acylation of the chitooligosaccharide backbone. *Proceedings of the National Academy of Sciences*, *91*(8), 3122-3126.

Ryan, S., Begley, M., Gahan, C. G., & Hill, C. (2009). Molecular characterization of the arginine deiminase system in Listeria monocytogenes: regulation and role in acid tolerance. *Environmental microbiology*, *11*(2), 432-445.

Sadowsky, M. J., Keyser, H. H., & Bohlool, B. B. (1983). Biochemical characterization of fast- and slow-growing rhizobia that nodulate soybeans. *International journal of systematic bacteriology*, *33*(4), 716-722

Sakanyan, V., Petrosyan, P., Lecocq, M., Boyen, A., Legrain, C., Demarez, M., ... & Glansdorff, N. (1996). Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway. *Microbiology*, *142*(1), 99-108.

Sällström, B., & Andersson, S. G. (2005). Genome reduction in the α-Proteobacteria. *Current opinion in microbiology*, *8*(5), 579-585.

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology*, *48*(1), 77-84.

Shi, D., Morizono, H., Yu, X., Roth, L., Caldovic, L., Allewell, N. M., ... & Tuchman, M. (2005). Crystal Structure of N-Acetylornithine Transcarbamylase from Xanthomonas campestris

A NOVEL ENZYME IN A NEW ARGININE BIOSYNTHETIC PATHWAY FOUND IN SEVERAL EUBACTERIA. *Journal of Biological Chemistry*, *280*(15), 14366-14369.

Shi, D., Yu, X., Roth, L., Morizono, H., Hathout, Y., Allewell, N. M., & Tuchman, M. (2005). Expression, purification, crystallization and preliminary X-ray crystallographic studies of a novel acetylcitrulline deacetylase from Xanthomonas campestris. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *61*(7), 676-679.

Shi, D., Caldovic, L., Jin, Z., Yu, X., Qu, Q., Roth, L., ... & Tuchman, M. (2006). Expression, crystallization and preliminary crystallographic studies of a novel bifunctional N-acetylglutamate synthase/kinase from Xanthomonas campestris homologous to vertebrate N-acetylglutamate synthase. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *62*(12), 1218-1222.

Shiu, S. H., Byrnes, J. K., Pan, R., Zhang, P., & Li, W. H. (2006). Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2232-2236.

Sørensen, M. A., Kurland, C. G., & Pedersen, S. (1989). Codon usage determines translation rate in Escherichia coli. *Journal of molecular biology*, *207*(2), 365-377.

Sørensen, M. A., & Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in Escherichia coli: The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *Journal of molecular biology*, *222*(2), 265-280.

Sorroche, F. G., Rinaudi, L. V., Zorreguieta, Á., & Giordano, W. (2010). EPS II-dependent autoaggregation of Sinorhizobium meliloti planktonic cells. *Current microbiology*, *61*(5), 465-470.

Sprent, J. I., & Platzmann, J. (2001). *Nodulation in legumes* (pp. 1-13). Kew: Royal Botanic Gardens.

Sprent, J. I., & James, E. K. (2007). Legume evolution: where do nodules and mycorrhizas fit in?. *Plant Physiology*, *144*(2), 575-581.

Staehelin, C., Granado, J., Müller, J., Wiemken, A., Mellor, R. B., Felix, G., ... & Boller, T. (1994). Perception of Rhizobium nodulation factors by tomato cells and inactivation by root chitinases. *Proceedings of the National Academy of Sciences*, *91*(6), 2196-2200.

Stoebel, D. M., Dean, A. M., & Dykhuizen, D. E. (2008). The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. *Genetics*, *178*(3), 1653-1660.

Stokkermans, T. J., Orlando, R., Kolli, V. K., Carlson, R. W., & Peters, N. K. (1996). Biological activities and structures of Bradyrhizobium elkanii low abundance lipo chitin-oligosaccharides. *MPMI-Molecular Plant Microbe Interactions*, *9*(4), 298-304.

Storm, C. E., & Sonnhammer, E. L. (2002). Automated ortholog inference from phylogenetic trees and calculation of orthology reliability. *Bioinformatics*, *18*(1), 92-99.

Studer, R. A., & Robinson-Rechavi, M. (2009). How confident can we be that orthologs are similar, but paralogs differ?. *Trends in Genetics*, *25*(5), 210-216.

Tabor, C. W., & Tabor, H. (1985). Polyamines in microorganisms. *Microbiological Reviews*, *49*(1), 81.

Takagi, M., Absalon, M. J., McLure, K. G., & Kastan, M. B. (2005). Regulation of p53 translation and induction after DNA damage by ribosomal protein L26 and nucleolin. *Cell*, *123*(1), 49-63.

Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M., & Rayment, I. (1997). Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product. *Biochemistry*, *36*(21), 6305-6316.

Treangen, T. J., & Rocha, E. P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet*, *7*(1), e1001284.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., ... & Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, *141*(2), 344-354.

Turner, S. L., & Young, J. P. W. (2000). The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Molecular Biology and Evolution*, *17*(2), 309-319.

Udvardi, M. K., & Day, D. A. (1997). Metabolite transport across symbiotic membranes of legume nodules. *Annual review of plant biology*, *48*(1), 493-523.

Varenne, S., Buc, J., Lloubes, R., & Lazdunski, C. (1984). Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of molecular biology*, *180*(3), 549-576.

Wagner, G. P., Amemiya, C., & Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proceedings of the National Academy of Sciences*, *100*(25), 14603-14606.

Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., ... & Chang, H. Y. (2012). Genome-wide measurement of RNA folding energies. *Molecular cell*, *48*(2), 169-181.

Wang, Z., Dong, X., Ding, G., & Li, Y. (2010). Research Comparing the retention mechanisms of tandem duplicates and retrogenes in human and mouse genomes.

Willems, A. (2006). The taxonomy of rhizobia: an overview. *Plant and Soil*, *287*(1-2), 3-14.

Yang, G. P., Debellé, F., Savagnac, A., Ferro, M., Schiltz, O., Maillet, F., ... & Promé, J. C. (1999). Structure of the Mesorhizobium huakuii and Rhizobium galegae Nod factors: a cluster of phylogenetically related legumes are nodulated by rhizobia producing Nod

factors with α, β-unsaturated N-acyl substitutions. *Molecular microbiology*, *34*(2), 227-237.

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in ecology & evolution*, *15*(12), 496-503.

Yao, S. Y., Luo, L., Har, K. J., Becker, A., Rüberg, S., Yu, G. Q., ... & Cheng, H. P. (2004). Sinorhizobium meliloti ExoR and ExoS proteins regulate both succinoglycan and flagellum production. *Journal of bacteriology*, *186*(18), 6042-6049.

Yi, S. V., & Goodisman, M. A. (2009). Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics*, *4*(8), 551-556.

Young, J. P. W., & Johnston, A. W. B. (1989). The evolution of specificity in the legume-Rhizobium symbiosis. *Trends in ecology & evolution*, *4*(11), 341-349.

Young, N. D., Debellé, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., ... & Gish, J. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*, *480*(7378), 520-524.

Zhan, H. J., Levery, S. B., Lee, C. C., & Leigh, J. A. (1989). A second exopolysaccharide of Rhizobium meliloti strain SU47 that can function in root nodule invasion. *Proceedings of the National Academy of Sciences*, *86*(9), 3055-3059.

Zhan, H. J., Lee, C. C., & Leigh, J. A. (1991). Induction of the second exopolysaccharide (EPSb) in Rhizobium meliloti SU47 by low phosphate concentrations. *Journal of bacteriology*, *173*(22), 7391-7394.

Zhang, X. S., & Cheng, H. P. (2006). Identification of Sinorhizobium meliloti early symbiotic genes by use of a positive functional screen. *Applied and environmental microbiology*, *72*(4), 2738-2748.

Zhao, J., Ridgway, D., Broderick, G., Kovalenko, A., & Ellison, M. (2008). Extraction of elementary rate constants from global network analysis of E. coli central metabolism. *BMC systems biology*, *2*(1), 41.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., ... & Gregory, B. D. (2010). Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genet*, *6*(9), e1001141.