



**CSIC**

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

UNIVERSIDAD DE GRANADA

DEPARTAMENTO DE MICROBIOLOGÍA

CONSEJO SUPERIOR DE INVESTIGACIONES  
CIENTÍFICAS

Caracterización del RNoma del endosimbionte diazotrófico  
*Sinorhizobium meliloti*: aproximación a la función de la  
chaperona de RNA Hfq y de los ribo-reguladores Smr7C y  
Smr15C1/C2

**OMAR TORRES QUESADA**

**TESIS DOCTORAL**

**2012**

Editor: Editorial de la Universidad de Granada  
Autor: Omar Torres Quesada  
D.L.: GR 1835-2012  
ISBN: 978-84-9028-095-9



Caracterización del RNoma del endosimbionte  
diazotrófico *Sinorhizobium meliloti*: aproximación a la  
función de la chaperona de RNA Hfq y de los ribo-  
reguladores Smr7C y Smr15C1/C2

Memoria que presenta el licenciado en Biología  
D. Omar Torres Quesada  
como aspirante al grado de Doctor

Fdo: **Omar Torres Quesada**

VºBº de los Directores de la Tesis Doctoral

Fdo: Dr. José Ignacio  
Jiménez Zurdo  
Doctor en Farmacia  
Científico Titular del CSIC

Fdo: Dr. Nicolás Toro García  
Doctor en Ciencias Biológicas  
Profesor de Investigación del CSIC

Universidad de Granada  
2012





Este trabajo de Tesis Doctoral ha sido realizado en el Departamento de Microbiología del Suelo y Sistemas Simbióticos (Grupo de Ecología Genética de la Rizosfera) de la Estación Experimental del Zaidín (CSIC-Granada).

Para la realización del siguiente trabajo, el Ldo. Omar Torres Quesada fue financiado por las siguientes fuentes:

- Beca predoctoral FPI adscrita al proyecto AGL2006-12466 del Ministerio de Ciencia e Innovación, comprendida entre septiembre de 2007 y septiembre de 2011.
- Beca de movilidad FPI para estancias breves del Ministerio de Ciencia e Innovación, disfrutada en Instituto de Ciencias Vegetales (Centre National de la Recherche, Gif-sur-Yvette, Francia) bajo la dirección del Doctor Martín Crespi. Periodo: Septiembre-Diciembre 2009.

Parte de los resultados presentados en esta Tesis Doctoral han sido publicados en revistas internacionales o están en preparación:

**del Val, C., Rivas, E., Torres-Quesada, O., Toro, N. and Jimenez-Zurdo, J. I.** (2007). Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. *Mol Microbiol.* **66**, (5), 1080-91.

**Torres-Quesada, O., Oruezabal, R. I., Peregrina, A., Jofre, E., Lloret, J., Rivilla, R., Toro, N. and Jimenez-Zurdo, J. I.** (2010). The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. *BMC Microbiol.* **10**, 71.

**del Val, C., Romero-Zaliz, R., Torres-Quesada, O., Peregrina, A., Toro, N. and Jiménez-Zurdo, J. I.** (2011). A survey of sRNA families in  $\alpha$ -proteobacteria. *RNA Biology.*

**Torres-Quesada, O., Laporte, P., Crespi, M., Toro, N., Jiménez-Zurdo, J.I.** (2012). The two seemingly homologs Smr15C1/C2 control transporter genes in *Sinorhizobium meliloti*. (en preparación).

**Torres-Quesada, O., Laporte, P., Crespi, M., Toro, N., Jiménez-Zurdo, J.I.** (2012). The Hfq-independent non-coding RNA Smr7C regulates iron metabolism in *Sinorhizobium meliloti*. (en preparación).

**Torres-Quesada, O., Reinkesmeier, J., Schlüter, J.P., Toro, N., Becker, A., Jiménez-Zurdo, J.I.** (2012). Deep sequencing analysis of Hfq-dependent small non-coding RNAs in *Sinorhizobium meliloti* reveals an unexpected abundance of *trans*-sRNAs in pSymA. (en preparación).

**Torres-Quesada, O., Bardou, F., Sobrero, P., Crespi, M., Toro, N., Jiménez-Zurdo, J.I., Valverde, C.** (2012). The RNA chaperone Hfq shows polar localization in free-living bacteria and bacteroids in *Sinorhizobium meliloti*. (en preparación).

**A mis padres, mi hermana y a Carolina**





## SUMMARY

---

Post-genomic research has revealed that the functional untranslated RNA species (collectively referred to as the RNome) are more abundant and diverse than expected in organisms of all kingdoms of life. In bacteria, most of these newly identified non-coding RNAs are small transcripts (50-400 nt; sRNAs) which generally act as post-transcriptional regulators of gene expression in response to environmental stimuli. Riboregulation involves either the interaction of the sRNA with specific proteins or, most often, base pairing with complementary sequence stretches within 5'-UTR regions of *trans*-encoded target mRNAs. These antisense interactions ultimately influence the translation and/or stability of the message and require the binding of the riboregulator to the widespread RNA chaperone Hfq.

*S. meliloti* is a soil bacterium that belongs to the  $\alpha$ -proteobacterial species of the order of the Rhizobiales (collectively called rhizobia) known for their ability to establish symbiotic relationships with legume plants. The outcome of these interactions is the formation of new specialized organs within the host, the root nodules, where bacteria undergo a process of profound morphological differentiation into bacteroids that are capable of nitrogen-fixation to the benefit of the host plant. As a recognized model microorganism the genome of *S. meliloti* was published in 2001. It consists of one chromosome (3.65 Mb) and two megaplasmids encoding most of the symbiotic functions, pSymA (1.35 Mb) and pSymB (1.68 Mb), with a total of 6.225 predicted protein-encoding genes. Besides the tRNA and rRNA *loci* (54 genes and 3 operons, respectively) only the tmRNA had been annotated after the primary analysis of the *S. meliloti* genome as non-coding RNA. Therefore, despite the emerging role of regulatory sRNAs in the establishment of bacteria-eukaryote interactions, the *S. meliloti* RNome had remained largely unexplored.

The objectives of this research work were:

1. Identification, molecular characterization and classification of the components of the *S. meliloti* RNome.
2. Genetic and genomic approach to the function of the RNA chaperone Hfq to assess the impact of riboregulation on *S. meliloti* biology.
3. Development of genetic tools for the investigation of the *trans*-acting riboregulators in this symbiotic bacterium. Application of these tools to the functional characterization of the Smr7C, Smr15C1 and Smr15C2 sRNAs.

## *Identification of non-coding RNAs in the genome of S. meliloti*

Comparative sequence data from eight related  $\alpha$ -proteobacteria were obtained and the interspecies pairwise alignments were scored with the programs eQRNA and RNAz as complementary predictive tools to identify conserved and stable secondary structures corresponding to putative non-coding RNAs. Northern experiments confirmed that ten of the predicted *loci*, selected among the original 32 candidates as most probable sRNA genes, expressed small transcripts. This result supports the combined use of eQRNA and RNAz as a robust strategy to identify novel sRNAs in bacteria.

Three of these sRNAs are structural homologs of the housekeeping *E.coli* RNase P, SRP and 6S RNAs. The remaining seven transcripts accumulated differentially in free-living and symbiotic conditions. Experimental mapping of their 5'-ends revealed that their encoding genes are organized in autonomous transcription units with recognizable promoter and termination signatures. These seven transcripts represent novel bacterial sRNAs that we termed Smr, which are conserved in phylogenetically related bacteria of the order of the Rhizobiales within the  $\alpha$ -subgroup of proteobacteria.

Co-immunoprecipitation (CoIP) with a chromosomally-encoded FLAG epitope-tagged Hfq protein specifically recognized by monoclonal anti-FLAG antibodies in cell extracts of a *S. meliloti* *hfq*<sup>FLAG</sup> strain was used to isolate RNA species associated to this protein. CoIP-RNA obtained from lysates of bacteria grown to log and stationary phases in TY broth was subjected to deep-sequencing in a Illumina/Solexa platform.

Analysis of sequencing reads revealed that this protein mostly binds mRNAs. Nonetheless, this strategy resulted in the identification of 226 *trans*-sRNAs and 234 *cis*-encoded antisense sRNAs, 69% of which are novel, unknown transcripts in *S. meliloti*. However, this approach did not saturate the search for *trans*-RNAs expressed by this bacterium, suggesting that in *S. meliloti* co-exist large populations of Hfq-dependent and independent *trans*-acting riboregulators.

### *Biology of the RNA chaperone Hfq*

Two independent *S. meliloti* mutants, 2011-3.4 and 1021 $\Delta$ *hfq*, were obtained by disruption and deletion of the *hfq* gene in the wild-type strains 2011 and 1021, respectively, both exhibiting similar growth defects as free-living bacteria. Transcriptomic profiling of 1021 $\Delta$ *hfq* revealed a general down-regulation of genes of sugar transporters and some enzymes of the central carbon metabolism, whereas transcripts specifying the uptake and metabolism of nitrogen sources (mainly amino

acids) were more abundant than in the wild-type strain. Proteomic analysis of the 2011-3.4 mutant independently confirmed these observations. Symbiotic tests showed that lack of Hfq led to a delayed nodulation, severely compromised bacterial competitiveness on alfalfa roots and impaired normal plant growth. Furthermore, a large proportion of nodules (55%-64%) elicited by the 1021 $\Delta$ *hfq* mutant were non-fixing, with scarce content in bacteroids and signs of premature senescence of endosymbiotic bacteria.

RNA-Seq of Hfq-bound mRNAs revealed that 40% of transcripts differentially regulated in the *hfq* mutant depend directly on the chaperone, including transport and *fix/nif* genes which further explains support the observed phenotype in the mutant.

Our results support that the *S. meliloti* RNA chaperone Hfq acts as a global post-transcriptional regulator of gene expression underlying the control of central metabolic pathways in free-living bacteria and influencing rhizospheric competence, survival of the microsymbiont within the nodule cells and nitrogen fixation during the symbiotic interaction with its legume host alfalfa.

#### ***The ar15 and ar7 sRNA families: characterization of Smr15C1, Smr15C2 and Smr7C***

Smr15C1 and Smr15C2 sRNAs are members of the  $\alpha$ 15 family of  $\alpha$ -proteobacterial sRNAs which occur with high levels of paralogy in individual genomes. Their expression profiles suggest that these sRNA genes have divergently regulated promoters in *S. meliloti* 1021 that drive differential expression of both transcripts. These sRNAs act either independently or additively rather than hierarchically for the post-transcriptional regulation of different or partially overlapping sets of mRNAs encoding a wide range of ABC transporters. Their regulatory mechanism is Hfq-dependent and mainly involves interaction of unpaired nucleotides within the 5' hairpin loop of each sRNA with complementary nucleotides within or in the vicinity of the RBS of the target mRNAs.

The Smr7C sRNA is encoded by a chromosomal gene occurring in single copy in the *S. meliloti* and related  $\alpha$ -proteobacterial genomes. The Smr7C transcript accumulates under iron depletion in free-living and undifferentiated endosymbiotic bacteria to activate translation of the transporter of the siderophore rhizobactin, RhtA. This positive effect on RhtA translation is likely due to the melting of an inhibitory hairpin at the 5'-UTR of the target mRNA caused by an Hfq-independent

base-pairing interaction that is predicted to be favored by the post-transcriptional processing of the primary Smr7C transcript at its 42 nucleotide position.

Smr15C1/C2 and Smr7C sRNAs are dispensable for free-living growth of bacteria in complete TY medium and for the establishment of an effective symbiosis with alfalfa under laboratory conditions.

Despite their differential biological traits, many  $\alpha$ -proteobacteria share the capacity to establish a variety of long-term interactions with higher eukaryotes. Our results predict that the functional characterization of the *S. meliloti* sRNAs and their homologs will contribute to the unraveling of common strategies used by  $\alpha$ -proteobacteria to infect and survive in eukaryotic cells.

# **AGRADECIMIENTOS**



## AGRADECIMIENTOS

---

El trabajo de una Tesis Doctoral no es mérito de una persona, sino de muchas, siendo cada una de ellas una parte fundamental del engranaje que permite que este barco llegue a buen puerto.

En primer lugar quiero agradecer a mis directores José Ignacio Jiménez Zurdo y Nicolás Toro su contribución inestimable a este trabajo. A Nicolás por haberme permitido formar parte de la gran familia de “Ecología genética de la rizosfera”. A José Ignacio, maestro de los RNAs no codificantes, por ser tutor y capitán en este largo viaje y por su contribución a la labor de formación de este nuevo científico.

A el resto de “jefes”, Paco por ser la mirada crítica de la Ciencia y por estas siempre dispuesto a discutir sobre todos los aspectos (científicos y no científicos). A Manolo por ser la “entidad financiadora y firmante”. A Fernando por su disponibilidad en todo momento.

A Martín Crespi por haberme permitido formar parte durante tres meses de su laboratorio. A todos los miembros de mi Tribunal de Tesis, por aceptar formar parte de esta gran familia. Especialmente quiero agradecer a Maria José Soto el estar en el principio y el final de esta etapa tan decisiva en la vida de un científico.

A las Dras. Elena Rivas y Coral del Val por su activa participación y asistencia en el análisis bioinformático correspondiente al análisis de genómica comparada. A la Dra. Anke Becker por su colaboración en las técnicas de secuenciación masiva y proporción de información relevante en esta tesis doctoral. A J.P. Schlüter y J. Reinkesmeier por su asistencia y proporción de scripts bioinformáticos.

Dado que durante una tesis doctoral el laboratorio se convierte en una segunda casa, tengo que agradecer a mis compañeros su apoyo durante todo este tiempo. Al grupo de los “bolletes”, Virgi, Helena, Carol, Ismael (er Gañán), Juanito y Rebeca por evitar que eche a correr en los principios y por proporcionarme las primeras herramientas en mi vida científica. A todos los “ecólogos de la rizosfera” personal técnico y becarios, Chema, Vane (Fosf), Tere, Alicia, Álex, Sensi, especialmente a los que han vivido desde cerca estos años. A Isa por ser mi compañera de “rincón” los primeros años, por sus anécdotas inspiradas en *Mr.Bean*, por sus horas siempre gratificantes en ARN y radioactividad. A Jose por ser compañero fiel los últimos años, en las cuervas de crecimiento nocturnas y los cartones de leche, en las torres de placas, en la poyata y puesto de estudio en donde siempre hay trasvase aleatorio de objetos, en todo. A la “Rubia”, que se fue para la próspera Alemania pero que ha sido



compañera de risas y vivencias en todo momento. A Pepe, Hurry y Pablo, el “nabolabo” por proporcionar risas y testosterona al grupo. A Lola, por ser la “wiki” con más conocimiento que un “Maniatics”. A Rafa por ser voz serena en todo momento. A Merche, por ser la otra compañera del mundo proteico y ser la voz contradictoria de Jose. A Laura, por su exactitud y cordura en esta “casa de locos”. Por supuesto no me olvido de Tita, cuyas manos están presentes en este trabajo y cuya labor ha sido determinante en todos los aspectos.

A mis compañeros “franceses” de Gif, sobre todo a Florian B. que tendió la mano desde el principio y me ayudó en todo momento, pese a las dificultades de mi francés. A Philippe Laporte por su inestimable ayuda en las “*in situs*” y microscopía.

Por supuesto, no todo en la vida es Ciencia, y también quiero agradecer a mis amigos, compañeros y artistas de Cuestabajo, Kike, Javi y Jose que han dado música a esta tesis que por supuesto suena con canciones de CUESTABAJO. Y junto con Cuestabajo están siempre nuestros fieles seguidores Jorge, Mari Carmen y Lidia que son y serán siempre nuestros mejores fans.

Y como un buen edificio se tiene que sustentar con una buena estructura, tengo que agradecer de corazón a mis cuatro pilares fundamentales su labor en este trabajo. A mis padres por haberme proporcionado todas las herramientas y recursos para hacer frente a la vida, a preguntarme tantas cosas que al final la Ciencia era la única opción de futuro. A mi hermana, por ser compañera de vivencias y de “Ciencia” durante toda mi vida, sabiendo que seguirá los mismos pasos. Y no puedo olvidar a mi gran pilar, Carolina, que con su sonrisa, su mirada y su compañía han dado luz a mi oscuridad, optimismo a mi pesimismo, alegría a mi tristeza. Sin ellos este trabajo nunca hubiera existido.

A TODOS...

GRACIAS

# ÍNDICE



# ÍNDICE

---

<b><u>INTRODUCCIÓN</u></b> .....	5
<b><u>1. EVOLUCIÓN SOBRE EL CONOCIMIENTO DE LOS RNAs NO CODIFICANTES EN BACTERIAS</u></b> .....	6
<b><u>2. CLASIFICACIÓN DE LOS RNAs NO CODIFICANTES BACTERIANOS</u></b> .....	7
<b><u>3. RNAs ESENCIALES O DE MANTENIMIENTO CELULAR</u></b> .....	9
3.1. RNAsA P.....	9
3.2. tmRNA.....	10
3.3. RNA 4.5S.....	11
<b><u>4. RIBO-REGULADORES (sRNAs)</u></b> .....	12
4.1. sRNAs QUE ACTÚAN EN CIS ( <i>cis</i> -sRNAs).....	12
4.2. sRNAs QUE ACTÚAN EN TRANS ( <i>trans</i> -sRNAs).....	14
4.2.1. <i>Trans-sRNAs que modifican la actividad de proteínas</i> .....	14
4.2.2. <i>Trans-sRNAs que actúan por apareamiento de bases</i> .....	16
<b><u>5. FISIOLÓGÍA DE LA RIBO-REGULACIÓN</u></b> .....	20
i) <u>Efecto sobre proteínas de membrana</u> .....	20
ii) <u>Efecto sobre la reprogramación del metabolismo</u> .....	22
iii) <u>Efecto sobre la síntesis de factores de transcripción</u> .....	22
5.1. sRNAs Y PATOGÉNESIS.....	23
5.1.1. <i>Regulación del quorum sensing</i> .....	24
5.1.2. <i>Regulación de la homeostasis de hierro</i> .....	24
5.1.3. <i>Regulación de múltiples factores de virulencia</i> .....	25
<b><u>6. SIMBIOSIS RHIZOBIUM-LEGUMINOSA</u></b> .....	27
6.1. EL PROCESO SIMBIÓTICO.....	28
6.1.1. <i>Colonización rizosférica e infección</i> .....	30
6.1.2. <i>Diferenciación morfológica del bacteroide y fijación de nitrógeno</i> .....	31
6.2. <i>SINORHIZOBIUM MELILOTI</i> .....	33
6.2.1. <i>El genoma de S. meliloti 1021</i> .....	34
6.2.2. <i>Las cepas S. meliloti 1021 y 2011</i> .....	35

<b><u>OBJETIVOS</u></b> .....	39
<b><u>MATERIALES Y MÉTODOS</u></b> .....	43
<b><u>M-1. ESPECIES Y CEPAS BACTERIANAS</u></b> .....	43
<b><u>M-2. PLÁSMIDOS Y VECTORES DE CLONAJE</u></b> .....	44
<b><u>M-3. OLIGONUCLEÓTIDOS EMPLEADOS EN ESTE ESTUDIO</u></b> .....	45
<b><u>M-4. TÉCNICAS MICROBIOLÓGICAS</u></b> .....	47
<b><u>M-4.1. MEDIOS DE CULTIVO Y ANTIBIÓTICOS.</u></b> .....	47
<u>M-4.1.1. MEDIOS DE CULTIVO.</u> .....	47
<u>M-4.1.2 ANTIBIÓTICOS.</u> .....	48
<b><u>M-4.2. CONDICIONES DE CULTIVO.</u></b> .....	48
<b><u>M-4.3. CONSERVACIÓN DE CULTIVOS BACTERIANOS.</u></b> .....	48
<b><u>M-4.4. MÉTODOS DE MOVILIZACIÓN DE DNA EXÓGENO EN BACTERIAS.</u></b> .....	49
<u>M-4.4.1. TRANSFORMACIÓN EN <i>E. COLI</i>.</u> .....	49
<i>M-4.4.1.1. Preparación de células competentes.</i> .....	49
M-4.4.1.1.1. Métodos químicos: RbCl.....	49
M-4.4.1.1.2. Células electrocompetentes.....	49
<i>M-4.4.1.2. Transformación de células competentes.</i> .....	50
M-4.4.1.2.1. Transformación por choque térmico.....	50
M-4.4.1.2.1. Electrotransformación.....	50
<u>M-4.4.2. CONJUGACIONES EN <i>S. MELILOTI</i>.</u> .....	50
<u>M-4.4.3. SELECCIÓN DE INTERCAMBIO ALÉLICO Y CO-INTEGRACIÓN.</u> ...	51
<b><u>M-5. TÉCNICAS DE BIOLOGÍA MOLECULAR</u></b> .....	52
<b><u>M-5.1. MANIPULACIÓN DE ÁCIDOS DESOXIRRIBONUCLEICOS (DNA)</u></b> .....	52
<u>M-5.1.1. AISLAMIENTO DE DNA.</u> .....	52
<i>M-5.1.1.1. Extracción de DNA total mediante kit comercial.</i> .....	52
<i>M-5.1.1.2. Lisados de colonia.</i> .....	52
<i>M-5.1.1.3. Extracción de DNA plasmídico mediante lisis alcalina.</i> .....	53

<i>M-5.1.1.4. Minipreparación de DNA plasmídico por precipitación con sales de magnesio.</i> .....	53
<u>M-5.1.2. EVALUACIÓN DE LA CANTIDAD Y CALIDAD DEL DNA EXTRAÍDO.</u> .....	54
<u>M-5.1.3. DIGESTIÓN DE DNA MEDIANTE ENDONUCLEASAS DE RESTRICCIÓN.</u> .....	54
<u>M-5.1.4. REACCIÓN DE DEFOSFORILACIÓN.</u> .....	55
<u>M-5.1.5. PCR (POLIMERASE-CHAIN-REACTION).</u> .....	55
<u>M-5.1.6. CLONAJE DE FRAGMENTOS.</u> .....	55
<i>M-5.1.6.1. Aislamiento de fragmentos de DNA de gel de agarosa.</i> .....	55
<i>M-5.1.6.2. Ligación de fragmentos de restricción.</i> .....	56
<i>M-5.1.6.3. Adenilación de amplificadores de PCR.</i> .....	56
<u>M-5.1.7. SECUENCIACIÓN Y ANÁLISIS DE SECUENCIAS.</u> .....	56
<u>M-5.1.8. ELECTROFORESIS DE DNA.</u> .....	56
<i>M-5.1.8.1. Electroforesis en geles de agarosa no desnaturalizantes.</i> .....	57
<i>M-5.1.8.2. Electroforesis en geles de poliacrilamida no desnaturalizantes.</i> .....	57
<u>M-5.1.9. MARCADORES DE PESO MOLECULAR DE DNA.</u> .....	58
<u>M-5.1.10. HIBRIDACIÓN DNA-DNA (SOUTHERN BLOT).</u> .....	58
<i>M-5.1.10.1. Marcaje radiactivo de la sonda.</i> .....	58
<i>M-5.1.10.2. Hibridación, lavados y revelado.</i> .....	59
<u>M-5.2. MANIPULACIÓN DEL RNA.</u> .....	59
<u>M-5.2.1. EXTRACCIÓN DE RNA TOTAL DE <i>S. MELILOTI</i>.</u> .....	59
<i>M-5.2.1.1. Extracción a partir de cultivos bacterianos.</i> .....	59
<i>M-5.2.1.2. Extracción a partir de nódulos.</i> .....	60
<i>M-5.2.1.3. Tratamiento con DNasa I.</i> .....	60
<i>M-5.2.1.4. Extracción de RNA total de <i>S. meliloti</i> mediante kit comercial.</i> .....	61
<u>M-5.2.2. ELECTROFORESIS DE RNA.</u> .....	62
<i>M-5.2.2.1. Electroforesis en geles de agarosa desnaturalizantes.</i> .....	62
<u>M-5.2.3. HIBRIDACIÓN DNA-RNA (NORTHERN-BLOT).</u> .....	62
<i>M-5.2.3.1. Electroforesis y transferencia del RNA.</i> .....	62

<i>M-5.2.3.2. Hibridación</i> .....	63
<u>M-5.2.4. HIBRIDACIÓN DE MICROARRAYS</u> .....	63
<i>M-5.2.4.1. Los microarrays Sm14KOLI</i> .....	63
<i>M-5.2.4.2. Marcaje de las muestras</i> .....	64
<i>M-5.2.4.3. Pretratamiento de los microarrays</i> .....	65
<i>M-5.2.4.4. Hibridación y lavados</i> .....	66
<i>M-5.2.4.5. Revelado y procesamiento de los datos</i> .....	66
<u>M-5.2.5. REVERSO-TRANSCRIPCIÓN (REACCIÓN RT)</u> .....	67
<b><u>M-5.3. MANIPULACIÓN DE PROTEÍNAS</u></b> .....	67
<u>M-5.3.1. EXTRACCIÓN DEL EXTRACTO PROTEICO TOTAL DE <i>S. MELILOTI</i></u> .....	67
<u>M-5.3.2. FRACCIONAMIENTO CELULAR DE <i>S. MELILOTI</i></u> .....	67
<u>M-5.3.3. ELECTROFORESIS DE PROTEÍNAS EN GELES SDS-PAGE</u> .....	68
<i>M-5.3.3.1. Preparación de los geles</i> .....	68
<i>M-5.3.3.2. Electroforesis</i> .....	69
<i>M-5.3.3.3. Detección de proteínas en geles de poliacrilamida</i> .....	70
Tinción con Azul de Coomassie .....	70
Tinción con nitrato de plata .....	70
<i>M-5.3.3.4. Conservación de geles</i> .....	71
<u>M-5.3.4. DETECCIÓN INMUNOLÓGICA DE PROTEÍNAS (WESTERN-BLOT)</u> .....	71
<i>M-5.3.4.1. Transferencia de proteínas a membranas de PVDF</i> .....	71
<i>M-5.3.4.3. Revelado inmunológico</i> .....	72
<u>M-5.3.5. PROTEÓMICA</u> .....	72
<i>M-5.3.5.1. Isoelectroenfoque</i> .....	73
<i>M-5.3.5.2. Segunda dimensión (SDS-PAGE)</i> .....	73
<b><u>M-6. ENSAYOS CON PLANTAS</u></b> .....	74
<u>M-6.1. ESTERILIZACIÓN Y GERMINACIÓN DE SEMILLAS</u> .....	74
<u>M-6.2. SOLUCIÓN NUTRITIVA PARA PLANTAS</u> .....	74
<u>M-6.3. CULTIVO DE PLANTAS</u> .....	75
<u>M-6.3.1. CULTIVO AXÉNICO DE TUBOS</u> .....	75

<u>M-6.3.2. CULTIVO AXÉNICO EN PLACAS DE AGAR.</u> .....	76
<u>M-6.3.3. CULTIVO EN JARRAS LEONARD.</u> .....	76
<b><u>M-6.4. ANÁLISIS DE LAS CINÉTICAS DE NODULACIÓN.</u></b> .....	76
<b><u>M-6.5. MEDIDA DEL GRADO DE COMPETITIVIDAD.</u></b> .....	77
<b><u>M-6.6. MEDIDA DEL PESO SECO.</u></b> .....	77
<b><u>M-6.7. REAISLAMIENTO DE BACTERIAS DE NÓDULOS DE LEGUMINOSAS.</u></b> .....	77
<b><u>M-6.8. PREPARACIÓN DE MUESTRAS DE SECCIONES DE NÓDULOS.</u></b> .....	78
<u>M-6.8.1. ESTUDIO DE MUESTRAS INCLUIDAS EN AGAROSA.</u> .....	78
<u>M-6.8.2. ESTUDIO DE MUESTRAS INCLUIDAS EN RESINAS DE ALTA RESOLUCIÓN (TECHNOVIT).</u> .....	78
<i>M-6.8.2.1. Primer día: recolección de los nódulos y fijación.</i> .....	78
<i>M-6.8.2.2. Segundo día: deshidratación de los nódulos.</i> .....	79
<i>M-6.8.2.3. Tercer día: deshidratación de los nódulos (continuación).</i> .....	79
<i>M-6.8.2.4. Cuarto día: polimerización de la resina.</i> .....	79
<i>M-6.8.2.5. Quinto día: polimerización de la resina (continuación).</i> .....	79
<i>M-6.8.2.6. Sexto día: realización de los cortes, tinción y visualización.</i> .....	80
<b><u>M-6.9. HIBRIDACIÓN IN SITU SOBRE NÓDULOS MADUROS.</u></b> .....	80
<u>M-6.9.1. SÍNTESIS DE LA RIBOSONDA.</u> .....	80
<u>M-6.9.2. PREPARACIÓN Y FIJACIÓN DE LOS NÓDULOS.</u> .....	81
<i>M-6.9.2.1. Fijación.</i> .....	81
<i>M-6.9.2.2. Deshidratación.</i> .....	81
<i>M-6.9.2.3. Inclusión en parafina, y generación de los cortes longitudinales.</i> .....	81
<u>M-6.9.3. HIBRIDACIÓN IN SITU (ISH).</u> .....	82
<i>M-6.9.3.1. Desparafinado y rehidratación de las secciones de nódulos.</i> .....	82
<i>M-6.9.3.2. Tratamiento de pre-hibridación e hibridación.</i> .....	82
<i>M-6.9.3.3. Lavados de hibridación.</i> .....	83
<i>M-6.9.3.4. Inmunomarcaje y revelado de la reacción.</i> .....	83
<b><u>M-7. TÉCNICAS ANALÍTICAS</u></b> .....	84



<u>M-7.1. DETERMINACIÓN DE LA ACTIVIDAD <math>\beta</math>-GALACTOSIDASA EN CULTIVOS CELULARES</u> .....	84
<u>RESULTADOS Y DISCUSIÓN</u> .....	85
<u>CAPÍTULO 1</u> .....	87
<u>I-1. INTRODUCCIÓN</u> .....	89
<u>I-1.1. METODOLOGÍAS PARA LA IDENTIFICACIÓN DE sRNAs EN GENOMAS BACTERIANOS</u> .....	89
<u>I-1.1.1. ABORDAJES BIOINFORMÁTICOS Y TÉCNICAS MOLECULARES CLÁSICAS</u> .....	89
<u>I-1.1.1. TECNOLOGÍAS DE ALTO RENDIMIENTO: RNÓMICA</u> .....	90
<u>I-2. DISEÑO EXPERIMENTAL</u> .....	94
<u>I-2.1. ANÁLISIS BIOINFORMÁTICO</u> .....	94
<u>I-2.1.1. SECUENCIAS GENÓMICAS, EXTRACCIÓN DE RIGS</u> .....	94
<u>I-2.1.2. ANÁLISIS CON LOS PROGRAMAS EQRNA Y RNAZ</u> .....	94
<u>I-2.1.2. MODELOS DE COVARIANZA</u> .....	96
<u>I-2.2. MAPEO DE EXTREMOS 5' Y 3' DE TRANSCRITOS MEDIANTE AMPLIFICACIÓN (RACE)</u> .....	96
<u>I-2.2.1. RACE 5'</u> .....	96
<u>I-2.2.2. RACE 3'</u> .....	97
<u>I-2.3. CONSTRUCCIÓN DE LA CEPA <i>S. MELILOTI</i> 2011 B23001<math>hfq^{FLAG}</math></u> .....	98
<u>I-2.4. SECUENCIACIÓN MASIVA (RNA-Seq) DE LA SUBPOBLACIÓN DE TRANSCRITOS ASOCIADOS A Hfq<math>^{FLAG}</math> (coIP-RNA)</u> .....	100
<u>I-2.4.1. PREPARACIÓN DEL coIP-RNA</u> .....	100
<u>I-2.4.2. RNA-Seq Y ANÁLISIS INFORMÁTICO</u> .....	101
<i>II.2.4.2.1. Preparación de muestras y secuenciación mediante Illumina/Solexa</i> ..	101
<i>I-2.4.2.2. Análisis bioinformático</i> .....	103
<u>I-3. RESULTADOS</u> .....	104
<u>I-3.1. IDENTIFICACIÓN DE NUEVOS sRNAs EN <i>S. MELILOTI</i> MEDIANTE GENÓMICA COMPARADA</u> .....	104
<u>I-3.1.1. ANÁLISIS BIOINFORMÁTICO</u> .....	104

<b><u>I-3.2. sRNAs DE FUNCIÓN CONOCIDA</u></b> .....	109
<u>I-3.2.1. RNASA P</u> .....	109
<u>I-3.2.2. RNA SRP O 4.5S</u> .....	111
<u>I-3.2.3. RNA 6S</u> .....	112
<b><u>I-3.3. IDENTIFICACIÓN DE NUEVOS sRNAs</u></b> .....	115
<u>I-3.3.1 POLARIDAD Y EXPRESIÓN DIFERENCIAL DE LOS GENES <i>smr</i></u> .....	115
<u>I-3.3.2. CARACTERIZACIÓN DE LAS UNIDADES TRANSCRIPCIONALES DE LOS GENES <i>smr</i></u> .....	118
<u><i>I-3.3.2.1. Identificación de los sitios de inicio de transcripción</i></u> .....	118
<u>I-3.3.3. PREDICCIÓN DE PROMOTORES Y TERMINADORES DE TRANSCRIPCIÓN</u> .....	121
<u>I-3.3.4. CONSERVACIÓN DE LOS sRNAs <i>Smr</i> EN BACTERIAS</u> .....	124
<u><i>I-3.3.4.1. Las familias <math>\alpha</math> de sRNAs: modelos de co-varianza y estructuras secundarias consenso</i></u> .....	124
<u><i>I-3.3.4.2. Distribución de las familias <math>\alpha</math> de sRNAs en el orden Rhizobiales</i></u> .....	126
<b><u>I-3.4. IDENTIFICACIÓN DE NUEVOS sRNAs EN <i>S. MELILOTI</i>, MEDIANTE SECUENCIACIÓN MASIVA</u></b> .....	129
<u>I-3.4.1. GENERACIÓN DE LA CEPA <i>S.MELILOTI</i> 2011 B23001<math>hfq^{FLAG}</math></u> .....	129
<u>I-3.4.2. ANÁLISIS FENOTÍPICO DE LA CEPA <i>S. MELILOTI</i> 2011<math>hfq^{FLAG}</math></u> .....	130
<u>I-3.4.3. Co-IP DE LA FRACCIÓN DE RNAs ASOCIADOS A <math>Hfq^{FLAG}</math></u> .....	132
<u>I-3.4.4. ANÁLISIS DE LOS DATOS DE SECUENCIACIÓN</u> .....	134
<b><u>I-4. DISCUSIÓN</u></b> .....	141
<b><u>ANEXO CAPÍTULO 1</u></b> .....	151
<b><u>CAPÍTULO 2</u></b> .....	169
<b><u>II-1. INTRODUCCIÓN</u></b> .....	171
<u>II-1.1. ESTRUCTURA DE LA PROTEÍNA <i>Hfq</i></u> .....	171
<u>II-1.2. MECANISMOS MOLECULARES Y RIBO-REGULACIÓN</u> .....	173
<u>II-1.3. LOCALIZACIÓN SUBCELULAR</u> .....	174
<u>II-1.4. DISTRIBUCIÓN Y BIOLOGÍA EN BACTERIAS</u> .....	175

<b>II-2. DISEÑO EXPERIMENTAL</b> .....	176
<b>II-2.1. CONSTRUCCIÓN DEL MUTANTE <i>S. MELILOTI</i> <math>\Delta hfq</math></b> .....	176
<b>II-2.2. ANÁLISIS DEL PERFIL TRANSCRIPTÓMICO DE MUTANTE <i>S. MELILOTI</i> 1021<math>\Delta hfq</math></b> .....	177
<b>II-2.3. CONSTRUCCIÓN DE LA FUSIÓN Hfq:GFP Y ESTUDIO DE LA LOCALIZACIÓN DE LA PROTEÍNA EN <i>S. MELILOTI</i></b> .....	178
<b>II-3. RESULTADOS</b> .....	179
<b>II-3.1. LA PROTEÍNA Hfq DE <i>S. MELILOTI</i></b> .....	179
<b>II-3.2. CARACTERIZACIÓN FUNCIONAL DEL GEN HFQ EN <i>S. MELILOTI</i></b> ..	181
<b>II-3.2.1. GENERACIÓN DE UN MUTANTE <i>hfq</i> EN <i>S. MELILOTI</i></b> .....	181
<b>II-3.2.2. ANÁLISIS FENOTÍPICO DE LOS MUTANTES <i>hfq</i></b> .....	182
<b>II-3.2.2.1. Fenotipo en vida libre</b> .....	182
II-3.2.2.1.1. Curvas de crecimiento .....	182
II-3.2.2.1.2. Array fenotipico Biolog .....	183
<b>II-3.2.2.2. Fenotipo durante la interacción simbiótica con <i>M. sativa</i></b> .....	185
II-3.2.2.2.1. Nodulación, competitividad y eficiencia de la simbiosis.....	185
II-3.2.2.2.2. Histología y organogénesis nodular .....	188
<b>II-3.2.3. ANÁLISIS TRANSCRIPTÓMICO/PROTEÓMICO DE LOS MUTANTES <i>hfq</i></b> .....	192
<b>II-3.2.3.1. Efecto de Hfq en el transporte y metabolismo en <i>S. meliloti</i></b> .....	194
<b>II-3.2.3.2. Hfq contribuye a la regulación de <i>fixK1/fixK2</i></b> .....	196
<b>II-3.2.3.3. Redes génicas afectadas por la mutación en <i>hfq</i></b> .....	197
<b>II-3.2.3.4. Análisis proteómico del mutante <i>S. meliloti</i> 2011-3.4</b> .....	198
<b>II-3.3. CARACTERIZACIÓN DE LA POBLACIÓN DE MRNAS QUE SE UNEN A LA PROTEÍNA Hfq EN <i>S. MELILOTI</i></b> .....	199
<b>II-3.4. LOCALIZACIÓN DE LA PROTEÍNA Hfq EN <i>S. MELILOTI</i></b> .....	201
II-3.5.1. Hfq Y RNASA E.....	203
<b>II-4. DISCUSIÓN</b> .....	204
<b>ANEXO CAPÍTULO 2</b> .....	215

<b><u>CAPÍTULO 3</u></b> .....	227
<b><u>III-1. INTRODUCCIÓN</u></b> .....	229
<b><u>III-1.1. REDUNDANCIA GENÓMICA Y FUNCIONAL DE LOS TRANS-sRNAs</u></b> .....	229
<b><u>III-1.2. MECANISMOS DE INTERACCIÓN TRANS-sRNA-mRNA: ELEMENTOS IMPLICADOS</u></b> .....	230
<b><u>III-1.2.1. ELEMENTOS DEL RNA IMPLICADOS EN EL APAREAMIENTO DE BASES</u></b> .....	230
<b><u>III-1.2.2. PAPEL DE LAS CHAPERONAS DE RNA</u></b> .....	232
<b><u>III-2. DISEÑO EXPERIMENTAL</u></b> .....	235
<b><u>III-2.1. ESTRUCTURA Y DISTRIBUCIÓN DE LOS SRNAS SMR7C, SMR15C1 Y SMR15C2</u></b> .....	235
III-2.1.1. mapeo de extremos 3' basado en la circularización de moléculas de RNA.....	235
III-2.1.2. predicción y representación de estructuras secundarias de RNAs.....	236
III-2.1.3. conservación en bacterias: modelos de co-varianza.....	236
<b><u>III-2.2. ESTUDIO DE LA EXPRESIÓN DIFERENCIAL DE LOS sRNAs Smr</u></b> .....	236
<b><u>III-2.2.1. ANÁLISIS DE REGIONES PROMOTORAS</u></b> .....	236
<b><u>III-2.2.2. CONDICIONES DE CULTIVO</u></b> .....	237
<b><u>III-2.3. GENERACIÓN DE MUTANTES EN LOS GENES <i>smr</i></u></b> .....	240
<b><u>III-2.3.1. MUTANTES POR DELECCIÓN</u></b> .....	240
<b><u>III-2.3.2. MUTANTES POR SOBREEXPRESIÓN</u></b> .....	242
<b><u>III-2.4. IDENTIFICACIÓN DE mRNAs DIANA</u></b> .....	244
<b><u>III-2.4.1. PREDICCIÓN BIOINFORMÁTICA: INTARNA</u></b> .....	244
<b><u>III-2.4.2. ENSAYO DE DOBLE PLÁSMIDO <i>IN VIVO</i></u></b> .....	244
<b><u>III-3. RESULTADOS</u></b> .....	248
<b><u>III-3.1. LA FAMILIA DE sRNAs <math>\alpha</math>r15: CARACTERIZACIÓN DE <i>Smr15C1</i> Y <i>Smr15C2</i></u></b> .....	249
<b><u>III-3.1.1. MAPEO DE LOS EXTREMOS 3' Y CARACTERIZACIÓN <i>IN SILICO</i> DE LOS sRNAs <i>Smr15C1</i> Y <i>Smr15C2</i></u></b> .....	249

<b><i>III-3.1.1.1. Conservación y distribución de Smr15C1 y Smr15C2 en el orden Rhizobiales</i></b> .....	251
<b><u>III-3.1.2. EXPRESIÓN DE smr15C1 Y smr15C2.</u></b> .....	253
<b><i>III-3.1.2.1. Cinética de expresión en medio TY</i></b> .....	253
<b><i>III-3.1.2.2. Expresión en condiciones de estrés</i></b> .....	255
<b><i>III-3.1.2.3 Expresión en nódulos maduros de alfalfa</i></b> .....	256
<b><i>III-3.1.2.4. Análisis de las regiones promotoras</i></b> .....	258
<b><u>III-3.1.3. APROXIMACIÓN GENÉTICA A LA FUNCIÓN DE Smr15C1 Y Smr15C2.</u></b> .....	260
<b><i>III-3.1.3.1. Generación y verificación experimental de los mutantes smr15C</i></b> .....	260
III-3.1.3.1.1. Mutantes por delección.....	260
III-3.1.3.1.2. Mutantes por sobreexpresión.....	262
<b><i>III-3.1.3.2. Análisis fenotípico de los mutantes smr15C</i></b> .....	263
III-3.1.3.2.1. Fenotipo en vida libre.....	263
III-3.1.3.2.2. Fenotipo simbiótico.....	264
<b><u>III-3.1.4. INTERACCIÓN DE Smr15C1 Y Smr15C2 CON LA CHAPERONA Hfq.</u></b> .....	268
<b><u>III-3.1.5. APROXIMACIÓN A LA IDENTIFICACIÓN DE LOS MRNA DIANA DE Smr15C1 Y Smr15C2.</u></b> .....	271
<b><i>III-3.1.5.1. Predicciones con IntaRNA</i></b> .....	271
<b><i>III-3.1.5.2. Evidencias experimentales</i></b> .....	275
III-3.1.5.2.1. Ensayo de doble plásmido.....	276
III-3.1.5.2.2 Evidencias basadas en la dependencia de Hfq.....	278
III-3.1.5.2.3 Alteración de las proteínas periplásmicas por expresión de Smr15C1/C2.....	279
<b><u>III-3.2. LA FAMILIA DE SRNAS <math>\alpha</math>7: CARACTERIZACIÓN DE Smr7C</u></b> .....	283
<b><u>III-3.2.1. MAPEO DEL EXTREMO 3' Y ANÁLISIS <i>IN SILICO</i>.</u></b> .....	283
<b><i>III-3.2.1.1. Conservación y distribución de Smr7C en el orden Rhizobiales</i></b> .....	285
<b><u>III-3.2.2. EXPRESIÓN DE Smr7C.</u></b> .....	285
<b><i>III-3.2.2.1. Cinética de expresión en medio TY</i></b> .....	285
<b><i>III-3.2.2.2. Expresión en condiciones de estrés</i></b> .....	288

<i>III-3.2.2.3. Expresión en nódulos maduros de alfalfa</i> .....	278
<i>III-3.2.2.4. Análisis de las regiones promotoras</i> .....	278
<u>III-3.2.3. APROXIMACIÓN GENÉTICA A LA FUNCIÓN DE Smr7C</u> .....	279
<i>III-3.2.3.1. Generación y verificación experimental de los mutantes Smr7C</i> .....	279
III-3.2.3.1.1. Deleción del gen <i>smr7C</i> .....	279
III-3.1.3.1.2. Sobreexpresión constitutiva de Smr7C.....	292
<i>III-3.2.3.2. Análisis fenotípico de los mutantes smr7C: vida libre y simbiosis</i> .....	293
<u>III-3.2.4. INTERACCIÓN DE SMR7C CON LA CHAPERONA Hfq</u> .....	296
<u>III-3.2.5. APROXIMACIÓN A LA IDENTIFICACIÓN DEL mRNA DIANA DE Smr7C</u> .....	299
<i>III-3.2.5.1. Predicciones con IntaRNA</i> .....	299
<i>III-3.2.5.2. Evidencias experimentales: ensayo de doble plásmido</i> .....	300
<b><u>III-4. DISCUSIÓN</u></b> .....	302
<b>ANEXO CAPÍTULO 3</b> .....	317
<b><u>DISCUSIÓN GENERAL</u></b> .....	321
<b><u>CONCLUSIONS</u></b> .....	327
<b><u>BIBLIOGRAFÍA</u></b> .....	331



# ÍNDICE DE TABLAS Y FIGURAS

---

## INDICE DE TABLAS

Tabla 1. Procesos afectados por RNAs no codificantes.	6
Tabla 2. Características del genoma de <i>S. meliloti</i> 1021	35
Tabla M-1. Especies y cepas bacterianas utilizadas en este trabajo	43
Tabla M-2. Principales plásmidos utilizados en este trabajo	44
Tabla M-3. Oligonucleótidos universales utilizados en este trabajo	45
Tabla M-4. Antibióticos utilizados en este trabajo y su concentración empleada en <i>E. coli</i> y <i>S. meliloti</i>	48
Tablas M-5 y M-6. Cantidades recomendadas para geles SDS-PAGE	69
Tabla M-7. Cantidades recomendadas para geles SDS-PAGE destinados a proteómica	75
Tabla I-1. RNomas bacterianos y tecnología utilizada para su caracterización	93
Tabla I-2. Sensibilidad y especificidad de eQRNA y RNAz	106
Tabla I-3. Predicciones de sRNAs en el genoma de <i>S. meliloti</i> 1021 correspondientes a eQRNA	107
Tabla I-4. Predicciones de sRNAs en el genoma de <i>S. meliloti</i> 1021 correspondientes a RNAz	107
Tabla I-5. Predicciones de sRNAs en el genoma de <i>S. meliloti</i> 1021 correspondientes a los programas eQRNA y RNAz	108
Tabla I-6. CoIP RNAs obtenidos	134
Tabla I-7. Candidatos correspondientes a las categorías de <i>trans</i> -sRNAs y <i>cis</i> -sRNAs antisentido identificados en nuestro estudio	151
Tabla I-8. Oligonucleótidos utilizados en este trabajo	166
Tabla I-9. Construcciones utilizadas en este trabajo	167
Tabla II-1. Transcritos diferencialmente expresados en el mutante 1021 $\Delta$ hfq	217
Tabla II-2. Proteínas diferencialmente acumuladas en <i>S. meliloti</i> 2011 y 2011-3.4 revelado por análisis de geles 2D	223
Tabla II-3. Genes compartidos por los análisis transcriptómicos, proteómicos y RNA-Seq	224
Tabla II-4. Oligonucleótidos utilizados en este trabajo	225
Tabla II-5. Construcciones utilizadas en este trabajo	226
Tabla III-1. Predicciones de IntaRNA en los tres replicones de <i>S. meliloti</i> 1021	272
Tabla III-2. Transportadores potencialmente regulados por Smr15C1 y Smr15C2 según las predicciones de IntaRNA	274
Tabla III-4. Evidencias experimentales para la verificación de las predicciones de IntaRNA	276



Tabla III-5. Características relevantes de los plásmidos diana para Smr15C1/C2 utilizados en el ensayo <i>in vivo</i>	278
Tabla III-6. Interacciones más probables para Smr7C	300
Tabla III-7. Características relevantes de los plásmidos diana para Smr7C utilizados en el ensayo <i>in vivo</i>	301
Tabla III-8. Oligonucleótidos utilizados en este trabajo	319
Tabla III-9. Construcciones utilizadas en este trabajo.	321

## INDICE DE FIGURAS

Fig. 1. Clasificación de los RNAs no codificantes bacterianos	8
Fig. 2. Estructura global del complejo enzima-sustrato de la RNasa P	10
Fig. 3. Estructura del tmRNA y mecanismo de acción.	11
Fig. 4. La partícula de reconocimiento de señal (SRP) de <i>E. coli</i> .	12
Fig. 5. Mecanismos de acción de los <i>cis</i> -sRNAs	14
Fig. 6. <i>Trans</i> -sRNAs que modifican la actividad de proteínas	16
Fig. 7. Efectos ejercidos por los <i>trans</i> -sRNAs que funcionan por apareamiento de bases sobre sus mRNAs diana.	17
Fig. 8. Ejemplos de regulación de <i>trans</i> -sRNAs que actúan mediante apareamiento de bases.	19
Fig. 9. Ribo-regulación de proteínas de membrana y transportadores ABC en <i>Salmonella</i>	21
Fig. 10. Ribo-regulación de la patogénesis	26
Fig. 11. Árbol filogenético 16S del grupo de las $\alpha$ -proteobacterias	28
Fig. 12. Etapas de la simbiosis <i>Rhizobium</i> -leguminosa	29
Fig. M-1. Principales plásmidos de destino utilizados en este trabajo	46
Fig. I.1. Evolución de la identificación de sRNAs	92
Fig. I-2. Modelos probabilísticos en los que se basa eQRNA	95
Fig. I-3. Pasos fundamentales en la técnica de mapeo de transcritos mediante RACE 5' y RACE 3'	98
Fig. I-4. Construcciones más relevantes en el proceso de creación del mutante <i>S. meliloti hfq<sup>FLAG</sup></i>	99
Fig. I-5. Diagrama del esquema de trabajo de obtención y secuenciación del CoIP-RNA	102
Fig. I-6. Estrategia bioinformática utilizada para la predicción de sRNAs en el genoma de <i>S. meliloti</i> 1021	104
Fig. I-7. Distribución de los alineamientos	105
Fig. I-8. Anotación del gen <i>rpnB</i> (RNasa P) en <i>S. meliloti</i> 1021.	110
Fig. I-9. Anotación del RNA 4.5S o SRP de <i>S. meliloti</i> 1021	112
Fig. I-10. Caracterización del gen <i>smr22C</i> (RNA 6S) de <i>S. meliloti</i> 1021.	114
Fig. I-11. Asignación de la polaridad de los genes <i>smr</i>	116

Fig. I-12. Patrones de expresión de los genes <i>smr</i>	117
Fig. I-13. RACE 5' de los transcritos <i>Smr</i>	119
Fig. I-14. Regiones genómicas de los genes sRNA identificados en <i>S. meliloti</i> 1021	120
Fig. I-15. Secuencias de los genes <i>smr</i>	123
Fig. I-16. Estructura consenso de las familias de sRNA $\alpha$ r	125
Fig. I-17. Conservación de los sRNAs <i>Smr</i> de <i>S. meliloti</i> 1021 en el orden Rhizobiales	127
Fig. I-18. Verificación de la modificación introducida en el gen cromosómico <i>hfq</i> en la cepa <i>S. meliloti</i> 2011 <i>hfq</i> <sup>FLAG</sup>	130
Fig. I-19. Caracterización fenotípica de <i>S. meliloti</i> 2011 <i>hfq</i> <sup>FLAG</sup>	131
Fig. I-20. Monitorización de la IP de <i>Hfq</i> <sup>FLAG</sup>	133
Fig. I-21. Diagrama resumen de las diferentes etapas del procesamiento de los datos de secuenciación del CoIP-RNA	135
Fig. I-22. Análisis estadístico de las lecturas obtenidas mediante Illumina/Solexa	136
Fig. I-23. Clasificación de las categorías correspondientes a las distintas regiones identificadas mediante RNA-Seq	137
Fig. I-24. Distribución de las categorías de RNAs en las librerías correspondientes a fase logarítmica y estacionaria	138
Fig. I-25. Análisis conjunto de las librerías correspondientes a fase logarítmica y estacionaria	139
Fig. I-26. Ejemplos de RNAs no anotados en el genoma de <i>S. meliloti</i> 1021 identificados en nuestro estudio	140
Fig. I-27. Evolución de la identificación de sRNAs en el genoma de <i>S. meliloti</i>	147
Fig. II-1. Estructura de <i>Hfq</i>	172
Fig. II-2. Modelos de actividad aceptados para <i>Hfq</i>	174
Fig. II-3. Construcciones plasmídicas para la generación del mutante	177
Fig. II-4. Etapas en la creación de la fusión traduccional <i>Hfq</i> :eGFP	179
Fig. II-5. Caracterización <i>in silico</i> del gen y la proteína <i>Hfq</i> de <i>S. meliloti</i>	180
Fig. II-6. Caracterización genética del mutante <i>S. meliloti</i> 1021 $\Delta$ <i>hfq</i>	182
Fig. II-7. Fenotipo de crecimiento del mutante <i>S. meliloti</i> 1021 $\Delta$ <i>hfq</i>	183
Fig. II-8. Análisis fenotípico del mutante 1021 $\Delta$ <i>hfq</i> mediante Biolog®.	184
Fig. II-9. Fenotipo simbiótico del mutante 1021 $\Delta$ <i>hfq</i> en <i>S. meliloti</i>	186
Fig. II-10. Fenotipo endosimbiótico del mutante <i>S. meliloti</i> 1021 $\Delta$ <i>hfq</i>	189
Fig. II-11. Organogénesis inducida por <i>S. meliloti</i> 1021 y 1021 $\Delta$ <i>hfq</i>	191
Fig. II-12. Análisis transcriptómico de <i>S. meliloti</i> 1021 $\Delta$ <i>hfq</i> sobre <i>microarrays</i> Sm14kOLI	193
Fig. II-13. Categorías funcionales de los genes diferencialmente expresados en el mutante <i>S. meliloti</i> $\Delta$ <i>hfq</i>	194
Fig. II-14. Influencia de <i>Hfq</i> en el transporte y metabolismo de pequeñas moléculas en <i>S. meliloti</i>	195
Fig. II-15. Actividad $\beta$ -galactosidasa en fusiones <i>fixK</i> :: <i>lacZ</i>	197

Fig. II-16. Agrupamiento en redes génicas de los transcritos diferencialmente expresados en el mutante <i>hfq</i>	198
Fig. II-17. Patrones de distribución en la cobertura del enriquecimiento de mRNAs asociados con Hfq	200
Fig. II-18. Localización sub-celular de Hfq en <i>S. meliloti</i> 1021	202
Fig. II-19. El gen <i>SMc01336</i> de <i>S. meliloti</i>	204
Fig. II-20. Resumen de las rutas y fenotipos asociados a la mutación <i>hfq</i> en <i>S. meliloti</i>	208
Figura III-1. Propiedades generales de los <i>trans</i> -sRNAs que actúan mediante apareamiento de bases	232
Fig. III-2. Monitorización del crecimiento de <i>S. meliloti</i> 1021	238
Fig. III-3. Adaptación al estrés oxidativo en <i>S. meliloti</i> 1021	239
Fig. III-4. Actividad de los promotores de <i>nifA</i> y <i>fixK</i> en <i>microaerobiosis</i> .	240
Fig. III-5. Estrategia seguida para la mutación de los genes <i>smr</i>	242
Fig. III-6. Promotores utilizados en la sobreexpresión de genes <i>smr</i>	243
Fig. III-7. Construcciones genéticas para la verificación de las interacciones sRNA-mRNA <i>in vivo</i> en <i>S. meliloti</i>	246
Fig. III-8. Validación del ensayo de doble plásmido <i>in vivo</i>	249
Fig. III-9. Los sRNAs <i>Smr15C1/C2</i> de <i>S. meliloti</i>	251
Fig. III-10. Estructura secundaria en la familia de sRNAs $\alpha$ 15	253
Fig. III-11. Relaciones filogenéticas entre los sRNAs de la familia $\alpha$ 15	255
Fig. III-12. Expresión de los genes <i>smr15C1/C2</i>	258
Fig. III-13. Alineamiento de las regiones promotoras de los genes $\alpha$ 15	260
Fig. III-14. Verificación experimental de las mutaciones $\Delta$ <i>smr15C</i>	262
Fig. III-15. Sobreexpresión constitutiva de <i>Smr15C1</i> y <i>Smr15C2</i>	264
Fig. III-16. Fenotipo de crecimiento de los mutantes <i>smr15C</i>	265
Fig. III-17. Cinéticas de nodulación y competitividad de los mutantes en los genes <i>smr15C</i>	266
Fig. III-18. Fenotipos endosimbióticos de los mutantes <i>smr15C</i>	268
Fig. III-19. Asociación de los sRNAs <i>Smr15C1/C2</i> a la chaperona Hfq	270
Fig. III-20. Estabilidad de los transcritos <i>Smr15C1/C2</i> en el fondo genético $\Delta$ <i>hfq</i>	271
Fig. III-21. Interacción de <i>Smr15C1/C2</i> con <i>SMA01650</i>	280
Fig. III-22. Enriquecimiento en librerías de CoIP-RNA de mRNAs diana predichos para <i>Smr15C1/C2</i>	281
Fig. III-23. Evidencias de la regulación del mRNA de <i>livK</i> por <i>Smr15C2</i> .	283
Fig. III-24. El sRNA <i>Smr7C</i> de <i>S. meliloti</i>	285
Fig. III-25. Estructura secundaria de los sRNAs de la familia $\alpha$ 7	287
Fig. III-26. Relaciones filogenéticas entre los sRNAs de la familia $\alpha$ 7	288
Fig. III-27. Expresión del gen <i>smr7C</i>	290
Fig. III-28. Alineamiento de las regiones promotoras de los genes $\alpha$ 7	291
Fig. III-29. Verificación experimental de la mutación $\Delta$ <i>smr7C</i>	292

Fig. III-30. Sobreexpresión constitutiva de Smr7C	293
Fig. III-31. Fenotipo de crecimiento de los mutantes <i>smr7C</i>	295
Fig. III-32. Cinéticas de nodulación y competitividad de los mutantes en el gen <i>smr7C</i>	296
Fig. III-33. Fenotipos endosimbióticos de los mutantes <i>smr7C</i>	297
Fig. III-34. Asociación del sRNA Smr7C a la chaperona Hfq	298
Fig. III-35. Estabilidad de los transcritos Smr7C en el fondo genético $\Delta hfq$	299
Fig. III-36. Interacción de Smr7C con RhtA	302
Fig. III-34. Modelo de regulación propuesta para los sRNAs Smr15C1/C2 y Smr7C en <i>S. meliloti</i>	315



# **INTRODUCCIÓN**



*“The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid”*

**F. Crick**





## INTRODUCCIÓN

“Un gen, una proteína”. Así comienza una de las asunciones que más revolucionaron el mundo de la biología molecular, allá por los años 60. El dogma fundamental de la biología, elaborado por Francis Crick, postula que el flujo de información genética es unidireccional y culmina con la interpretación de una secuencia de DNA como el producto final de la expresión de todo gen, la proteína. En este flujo se introduce una molécula que actúa únicamente como intermediario entre DNA y proteína, el que poco después se conocería como RNA mensajero (Crick, 1970).

Durante muchos años, el RNA fue considerado, por tanto, un mero eslabón en la cadena de traducción del mensaje genético. Incluso con el descubrimiento de los otros RNAs que intervienen en ésta, los RNAs transferentes (tRNAs) y los ribosómicos (rRNAs), su papel en la fisiología celular siguió considerándose secundario. Sin embargo, la era post-genómica ha puesto en entredicho este dogma fundamental y la asunción de que el único producto de la expresión génica es la proteína ha sido rebatido con nuevas evidencias experimentales. Necesariamente el concepto de gen como fragmento de DNA que codifica una única proteína y la idea de la unidireccionalidad del flujo de información génica han tenido que ser reconsiderados. Actualmente se sabe que, además de los “clásicos” RNAs con función reconocida, tanto los organismos eucariotas como las bacterias expresan una diversidad de transcritos que no se traducen (no codificantes) y que conforman lo que actualmente se conoce como **RNoma**.

Ya hace 10 años comenzaba a ser evidente la participación de los **RNAs no codificantes** de proteínas en el control de la expresión génica y de la actividad de los productos de ésta. Los encontramos en procesos que abarcan desde la regulación de la replicación y la transcripción del DNA hasta el control de la actividad proteica, pasando por el procesamiento y maduración de RNAs y la asistencia en la traducción de RNAs mensajeros. Esta diversidad de funciones de los RNAs no codificantes en los organismos vivos aparecen recogidas en la Tabla 1.

Tabla 1. Procesos afectados por RNAs no codificantes <sup>1</sup>

Proceso	Ejemplo	Función	Referencia <sup>2</sup>
Transcripción	184-nt 6S <i>E. coli</i>	Modula la utilización de distintos promotores	(a,b)
	331-nt 7SK humano	Inhibe al factor de elongación de la transcripción P-TEFb	(c, d, e)
Silenciamiento génico	875-nt SRA humano	Co-activador del receptor de esteroides	(f, g)
	16.500-nt Xist humano	Requerido para la inactivación del cromosoma X	(f, g)
Replicación	100.000-nt Air humano	Requerido para la impronta de genes autosómicos	(h)
	451-nt telomerasa RNA humana	Núcleo y molde de la telomerasa	(e, j)
Procesamiento RNA	377-nt RNasa P <i>E. coli</i>	Núcleo catalítico de la RNasa P	(a, k)
	186-nt snRNA U2 humano	Núcleo del espliceosoma	(e)
Modificación RNA	102-nt snoRNA U18 C/D <i>S. cerevisiae</i>	Dirige la metilación del O-29 de la ribosa del rRNA diana	(l, m)
	189-nt snoRNA snR8 H/ACA <i>S. cerevisiae</i>	Dirige la pseudouridilación del rRNA diana	(l, m)
Estabilidad RNA	68-nt gRNA gCYb <i>T. brucei</i>	Dirige la inserción y escisión de uridinas	(n, o)
	80-nt sRNA RyhB <i>E. coli</i>	Etiqueta mRNAs diana para la degradación	(p)
Traducción mRNA	miRNA Eucariótico	Etiqueta mRNAs diana para la degradación	(r, s)
	109-nt OxyS <i>E. coli</i>	Reprime la traducción por oclusión del sitio de unión a ribosoma	(a, t)
	87-nt sRNA DsrA <i>E. coli</i>	Activa la traducción por prevención de la formación de la estructura inhibitoria en el mRNA	(a, t)
Estabilidad proteínas	22-nt miRNA lin-4 <i>C. elegans</i>	Reprime la traducción por apareamiento el extremo 3' del mRNA diana	(r, s)
	363-nt tmRNA <i>E. coli</i>	Dirige la adición de etiquetas a péptidos en ribosomas atascados	(a, u)
Translocación proteínas	114-nt RNA 4,5S <i>E. coli</i>	Componente integral de la <i>signal recognition particle</i> , que centraliza el transporte de proteínas a la membrana	(a, v)

<sup>1</sup> Tomado de Storz, 2002

<sup>2</sup> a. Wassarman *et al.*, 1999; b. Wassarman and Storz, 2000; c. Yang *et al.*, 2001; d. Nguyen *et al.*, 2001; e. Gu *et al.*, 1998; f. Erdmann *et al.*, 2000; g. Avner and Heard, 2001; h. Sleutels *et al.*, 2002; i. Chen *et al.*, 2000; j. Chen *et al.*, 2000; k. Frank and Pace, 1998; l. Kiss, 2001; m. Samarsky *et al.*, 1999; n. Simpson *et al.*, 2000; o. Stuart *et al.*, 1998; p. Masse and Gottesman, 2002; r. Ruvkun, 2001; s. Grosshans and Slack, 2002 t. Altuvia and Wagner, 2000; u. Gillet and Felden, 2001; v. Keenan *et al.*, 2001.

## 1. EVOLUCIÓN SOBRE EL CONOCIMIENTO DE LOS RNAs NO CODIFICANTES EN BACTERIAS

Hasta hace poco, el potencial regulador de los RNAs se consideraba exclusivamente limitado en bacterias, a transcritos que controlaban funciones accesorias en bacterias como la replicación y compatibilidad entre plásmidos (Wagner and Simons, 1994; Wagner *et al.*, 2002). Ya en los años 70, se observó que las dos hebras de algunos genes localizados en plásmidos, transposones y fagos daban lugar a sendos transcritos, uno que codificaba una proteína y otro complementario a éste o antisentido que no es traducido. Es el caso del plásmido *colE1*, que expresa dos RNAs cuando se forma la horquilla de replicación; RNA I y su antisentido RNAII. Cuando RNA II no se une a RNA I, éste es capaz de formar un dúplex con la hebra de DNA y servir como cebador en la replicación. Sin embargo, cuando hay interacción entre RNA II y RNA I, se forma un híbrido que no permite el establecimiento de ese

dúplex y por tanto la replicación del plásmido no tiene lugar (Tomizawa, 1984,1985,1986). RNA II, junto con otros muchos RNAs identificados en otras entidades genéticas extracromosómicas actúan por apareamiento de bases a lo largo del mensajero diana, por lo que se denominaron genéricamente RNAs antisentido.

La versatilidad, eficiencia y especificidad de este mecanismo regulador anticipaban que los genes codificados en los cromosomas bacterianos podían regularse también por interacciones entre transcritos de secuencias complementarias. Algunas bases de datos de hace poco más de 10 años, recogen sólo en torno a diez RNAs no codificantes (excluyendo a los tRNAs y rRNAs) expresados por genes cromosómicos en la bacteria modelo *Escherichia coli* (Wassarman *et al.*, 1999). Algunos de ellos tienen funciones esenciales o de mantenimiento en la célula (M1 RNA o RNasa P, tmRNA, SRP RNA) y al resto se les asignó, como a los referidos RNAs antisentido, una función reguladora de la expresión génica en esta bacteria (e.g. 6S RNA, Spot42, MicF, OxyS, DsrA, DicF, CsrB). Incluso algunos, como RNasa P, tmRNA, SRP RNA o 6S, fueron identificados posteriormente en otras bacterias por homología. Durante la última década el análisis funcional de un número exponencialmente creciente de genomas secuenciados está generando extensos catálogos de RNAs no codificantes que se recogen en la base de datos Rfam (Gardner *et al.*, 2009) y aguardan su caracterización funcional.

## 2. CLASIFICACIÓN DE LOS RNAs NO CODIFICANTES BACTERIANOS

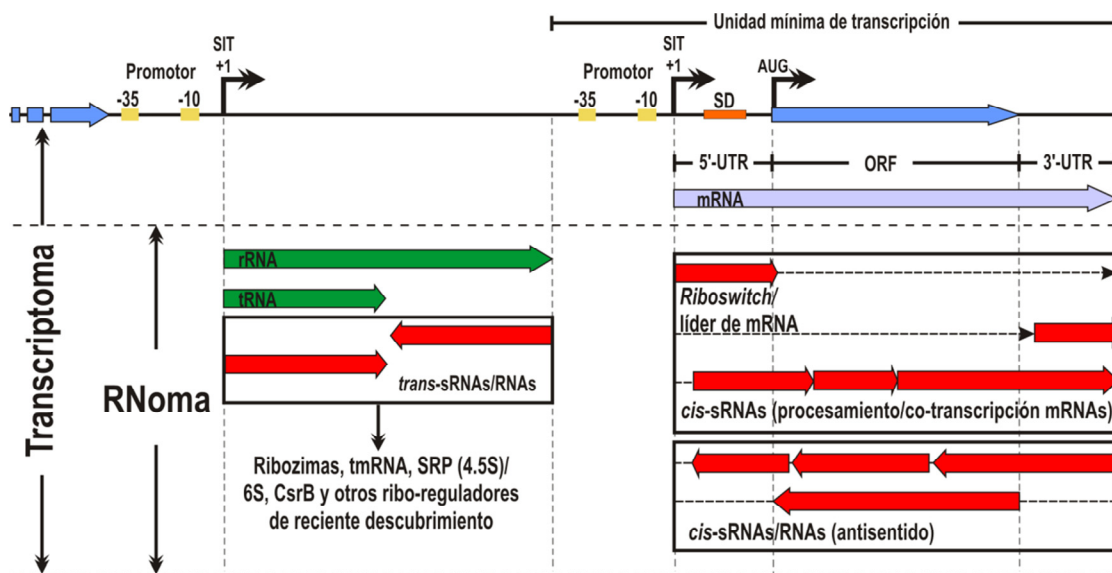
Por lo que se conoce hasta ahora sobre su función en bacterias, los RNAs no codificantes se pueden asignar a dos grandes grupos generales:

**1. Transcritos con funciones de mantenimiento (*housekeeping*) celular.** Agrupan a la mayoría de los primeros RNAs no codificantes identificados inicialmente. Al mutarse producen fenotipos evidentes o comprometen la viabilidad celular en algún caso. Aquí se encuentran los RNAs que controlan la síntesis y secreción de proteínas; tRNAs, rRNAs, tmRNA, RNA SRP y los RNAs catalíticos o ribozimas como la RNasa P. Entre estos últimos también se incluyen los intrones de los grupos I y II, que funcionan como elementos genéticos móviles.

**2. Transcritos que actúan como reguladores, generalmente post-transcripcionales, de la expresión génica (Ribo-reguladores):** es el grupo más numeroso de RNAs no codificantes de reciente descubrimiento en bacterias (Gottesman, 2004; Storz, 2002). Está formado por un conjunto heterogéneo de transcritos que se encuentran en un rango de tamaño de 50-400 nucleótidos, por lo que se les denomina de forma genérica **sRNAs** (*small* RNAs). Estos transcritos utilizan diversos mecanismos para

regular la expresión de mRNAs o interferir en la actividad de determinadas proteínas. Teniendo en cuenta la posición de sus respectivos *loci* genómicos respecto a los ORFs que codifican sus mRNAs/proteínas diana, los ribo-reguladores se clasifican en dos grupos (Schluter *et al.*, 2010) (Fig. 1):

- **Cis-sRNAs.** Son RNAs transcritos desde *loci* adyacentes a los ORFs sobre los que, generalmente ejercen su función reguladora, pudiendo solapar con la secuencia de los mRNAs codificados por éstos. Dentro de este grupo se pueden distinguir varios subtipos: *i*) RNAs localizados en las regiones 5' o 3' no traducidas (UTRs) del mensajero, *ii*) RNAs antisentido, codificados en la hebra complementaria del mRNA y, *iii*) RNAs sentido que se originan por procesamiento de determinados mRNAs o por co-transcripción con éstos desde promotores alternativos.



**Fig. 1. Clasificación de los RNAs no codificantes bacterianos.** En la parte superior se muestran los componentes de la unidad transcripcional mínima: secuencia codificante, inicio de traducción, RBS, sitio de unión a ribosoma, inicio de transcripción (SIT), cajas promotoras -10 y -35. El espacio entre unidades transcripcionales corresponde a la región intergénica (RIG). Los diferentes transcritos que se muestran son: mRNA, RNA mensajero; tRNA y rRNA, RNAs transferentes y ribosómicos; *trans*-sRNAs, localizados en RIGs; *cis*-sRNAs antisentido, localizados en la cadena complementaria de mensajeros y transcritos en dirección opuesta, abarcando parte o la totalidad del mRNA; *cis*-sRNAs sentido, localizados en las regiones 5' ó 3' no traducidas (UTR) del mensajero; sRNAs sentido, procesados o co-transcritos con el mRNA, que podían también actuar en *trans*.

- **Trans-sRNAs.** Son transcritos generalmente codificados entre ORFs (regiones intergénicas; RIGs), en unidades transcripcionales independientes que tienen sus propios promotores y terminadores de transcripción y se localizan en posiciones genómicas distantes a los genes que codifican sus mRNAs/proteínas diana. En este

grupo se encuentran la mayoría de los sRNAs con función conocida. Algunos sRNAs originados en mRNAs pueden tener una doble función como *cis*- (actuando sobre el mRNA que los codifica) o *trans*-sRNAs.

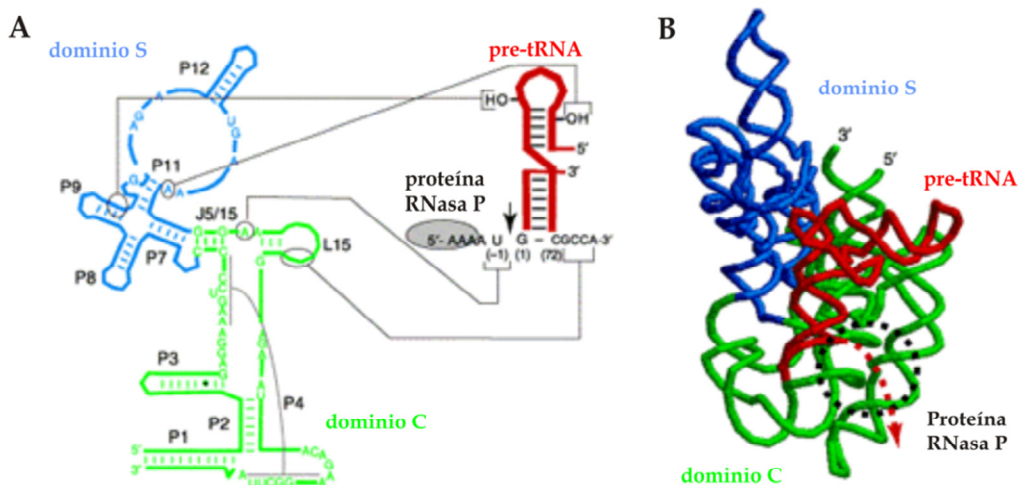
### 3. RNAs ESENCIALES O DE MANTENIMIENTO CELULAR

La mayoría de los RNAs de este grupo se descubrieron inicialmente en *E. coli* y posteriormente se identificaron homólogos en el resto de bacterias, por lo que su funcionamiento se conoce con más detalle en esta bacteria modelo. Destacan tres: la ribozima **RNasa P**, el **tmRNA** y el **RNA 4.5S**.

#### 3.1. RNASA P

Todos los tRNAs bacterianos son transcritos como precursores y requieren de un procesamiento posterior para generar extremos 5' y 3' maduros. La endorribonucleasa que se encarga del procesamiento de los extremos 5' de los tRNAs se conoce como RNasa P y es indispensable para la viabilidad celular (Altman *et al.*, 1993; Frank and Pace, 1998). En eubacterias, la holoenzima RNasa P está integrada por dos subunidades: el componente RNA M1, que posee un tamaño que oscila entre los 350 y 410 nucleótidos y el componente proteico C5, con un tamaño medio de 14 kDa (Altman, 1989). Todas las bacterias y el resto de organismos codifican homólogos de este RNA esencial, siendo entre las bacterianas mejor conocidas la de *E. coli*, *Bacillus subtilis*. (Gardiner and Pace, 1980; Vioque *et al.*, 1988) y *Thermotoga maritima*, la cual ha sido cristalizada en su totalidad junto con su tRNA diana (Reiter *et al.*, 2010).

Estudios *in vitro* han demostrado que el componente M1 RNA por sí solo es capaz de catalizar un correcto procesamiento de los precursores tRNA, aunque la proteína C5 se requiere en el proceso *in vivo* (Harris and Christian, 2003). Consecuentemente, M1 RNA es considerado como una ribozima clásica (RNA catalítico). Este RNA posee una estructura secundaria compleja formada por varias horquillas (identificadas con la letra P en la Fig. 2) que a su vez pueden conectar entre sí mediante interacciones terciarias, dando lugar a los dos dominios de la molécula, el dominio catalítico o dominio-C y el dominio específico o dominio-S (Fig. 2). La zona catalítica (horquilla P4) es la que establece el contacto con el tRNA diana (bucle TΨC) y va a ser la encargada de producir el corte del extremo 5' del tRNA, en un proceso en el que la presencia de iones metálicos, posiblemente magnesio, es crítica (Harris and Christian, 2003; Reiter *et al.*, 2010).

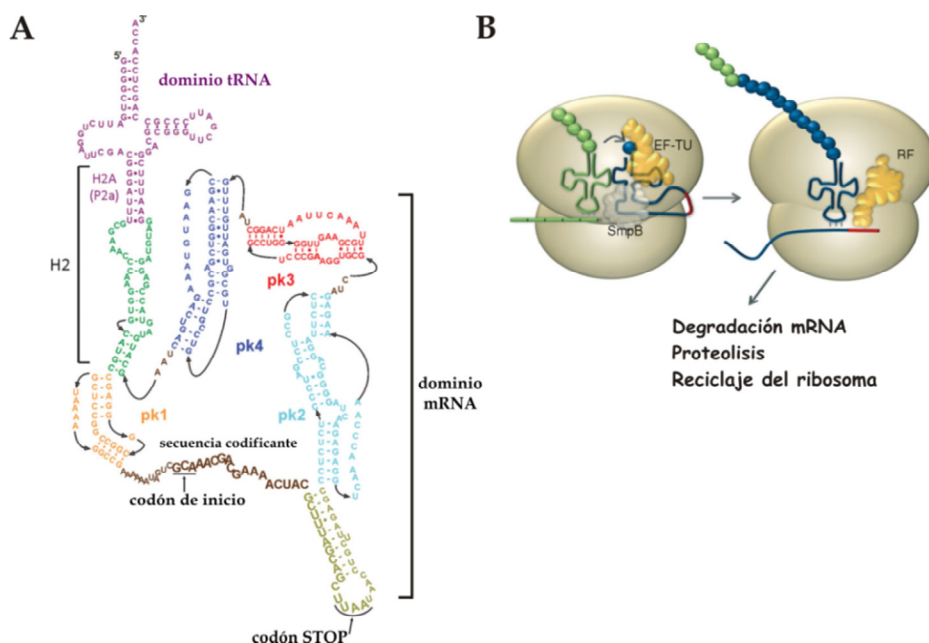


### 3.2. tmRNA

Ocasionalmente, los ribosomas, durante la traducción de los mensajeros, pueden quedar bloqueados y, por lo tanto, se interrumpe el proceso de traducción. Esto supone una disminución de la tasa de traducción y genera polipéptidos incompletos que deben ser eliminados. El tmRNA (*transfer-messenger RNA*), un transcrito de aproximadamente 363 nt, es el encargado de solventar este problema (Keiler *et al.*, 1996; Gillet and Felden, 2001).

Este RNA está altamente estructurado (Felden *et al.*, 1997; Lindell *et al.*, 2002); contiene un dominio que imita la estructura de un tRNA (dominio tRNA) y un segundo dominio que codifica una secuencia aminoacídica corta (dominio mRNA) (ver Fig. 3A). El dominio tRNA es reconocido por la aminoacil-tRNA sintetasa, una el factor de elongación *Tu* y entra en el sitio A libre del ribosoma que se encuentra bloqueado en el mRNA problemático. A través de la transpeptidación, el péptido naciente se transfiere al tmRNA. Después de la translocación y liberación del mRNA problemático, la traducción continúa en *trans* en el segmento mensajero del tmRNA, liberándose así un péptido que contiene la secuencia señal para ser degradado por las proteasas. Al mismo tiempo, el ribosoma es liberado y reciclado (Fig. 3B). En este

proceso es esencial una proteína que acomoda el tmRNA en el sitio A del ribosoma, la proteína SmpB, codificada por todos los genomas bacterianos conocidos (Barends *et al.*, 2011).



**Fig. 3. Estructura del tmRNA y mecanismo de acción.** A) Representación de la estructura secundaria del tmRNA. En colores se muestran las hélices correspondientes al dominio mRNA, denotadas por pk. Se señala el dominio tRNA (violeta) y la hélice que lo conecta con el dominio mRNA (H2). En el dominio mRNA se muestra la secuencia codificante correspondiente al péptido que se sintetiza desde el tmRNA. B) Esquema del mecanismo de acción del tmRNA. Se muestra el tmRNA (azul) en el ribosoma atascado y el tRNA que contiene el péptido naciente (verde). La liberación del mRNA problemático produce su degradación, la proteólisis del péptido naciente y el reciclaje del ribosoma. Modificado de Barends *et al.*, 2011.

### 3.3. RNA 4.5S

La partícula de reconocimiento de señal (*signal recognition particle*, SRP) constituye una maquinaria molecular que coordina el etiquetado de proteínas de membrana y de las que son secretadas durante el proceso de traducción. Las bacterias codifican una SRP formada por una proteína, Ffh (homóloga a las proteínas SRP eucarióticas), y un RNA de 114 nt, el 4.5S RNA, que usualmente se estructura formando una única horquilla (Fig. 4A, B). No es así por ejemplo en el caso de *B. subtilis*, cuyo RNA SRP (llamado 6S y no 4.5S por su mayor tamaño) posee un dominio Alu, típico de las SRP eucarióticas (Regalia *et al.*, 2002).

El ciclo SRP comienza con el reconocimiento de la secuencia señal del péptido naciente que emerge del ribosoma por parte del complejo 4.5S-Ffh. Una vez que ha ocurrido esta asociación, todo el complejo SRP junto con el ribosoma se traslada a la



membrana, en donde se encuentra el receptor de la SRP (proteína FtsY), y donde interacciona para culminar con el proceso de secreción, liberándose la partícula SRP del ribosoma de nuevo al citoplasma (Fig. 4C). En *E. coli*, la liberación de la proteína Ffh de su receptor FtsY es un proceso dependiente de hidrólisis de GTP y se ha demostrado que el tetra-bucle formado en el RNA 4.5S tiene un papel fundamental en esta liberación. (Siu *et al.*, 2007).

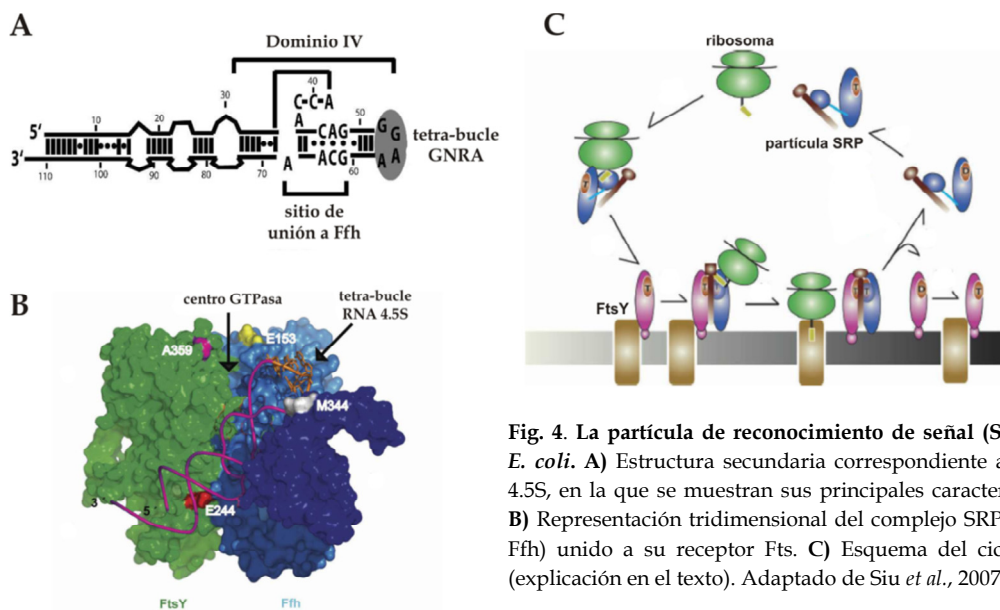


Fig. 4. La partícula de reconocimiento de señal (SRP) de *E. coli*. **A)** Estructura secundaria correspondiente al RNA 4.5S, en la que se muestran sus principales características. **B)** Representación tridimensional del complejo SRP (4.5S - Ffh) unido a su receptor Fts. **C)** Esquema del ciclo SRP (explicación en el texto). Adaptado de Siu *et al.*, 2007

## 4. RIBO-REGULADORES (sRNAs)

### 4.1. sRNAs QUE ACTÚAN EN CIS (CIS-sRNAs)

Además de estar relacionados con la regulación de funciones propias de elementos extracromosómicos como la replicación plasmídica ya descrita (e.g. plásmido ColE1), los *cis*-RNAs antisentido pueden intervenir en otros procesos. Es el caso de los **genes asesinos** (*killer genes*), que se encuentran muy conservados en bacterias. Un ejemplo sería el gen *gef*, que codifica una toxina que rápidamente mata a las células. Un RNA corto, Sof es codificado en la hebra complementaria del mismo *locus*, pero solapando con la región 5' del mensajero *Gef* (Poulsen *et al.*, 1991). La traducción de *Gef* es a su vez dependiente de la traducción de una secuencia que solapa con el marco de lectura del gen *ORF69*. Cuando Sof interacciona con el inicio de *ORF69* inhibe la traducción de éste y a su vez provoca indirectamente la inhibición de la expresión del gen letal *gef* (Fig. 5A).

El efecto positivo sobre la expresión génica de la ribo-regulación antisentido en *cis* es excepcional y sólo se ha descrito para **GadY** en *E. coli*. Este RNA está codificado por un gen que se localiza entre *gadX* y *gadW*, ambos integrantes del operón *gadAXW*. La transcripción de GadY es antisentido respecto a *gadX* y *gadW* y depende de los productos de estos dos genes y del factor  $\sigma^5$  ó  $\sigma^{38}$ , que se expresan en la célula en respuesta al estrés ácido (Tramonti *et al.*, 2008). Tanto el transcrito primario de GadY de 105 nt, como sus dos formas procesadas de 90 nt y 59 nt aparean con la región 3' no traducida del mensajero de *gadX* al menos en 30 nt. Este apareamiento se traduce en un aumento de la estabilidad de GadX, posiblemente porque el híbrido GadY-GadX constituya una forma resistente al ataque por nucleasas, como RNasa E (Opdyke *et al.*, 2004; Takada *et al.*, 2007) (Fig. 5B).

El grupo más numeroso de los *cis*-RNAs son los *riboswitches*. Éstos son regiones 5' no traducidas (5'-UTRs) de algunos mensajeros que pueden unir pequeños ligandos o moléculas (e.g. iones o vitaminas) lo que modifica su estructura secundaria y condiciona la traducción del mRNA. Estos cambios estructurales se relacionan con variaciones de las condiciones ambientales, y por tanto con las concentraciones de metabolitos, que son detectadas por estos elementos (Serganov, 2009). Un ejemplo, entre los muchos bien caracterizados, es el del *riboswitch* FMN. Este RNA se localiza en la región 5'-UTR de RNAs mensajeros implicados en la biosíntesis y transporte de flavín mononucleótidos (FMN) (Vitreschak *et al.*, 2002; Serganov *et al.*, 2009). En ausencia de FMN en el medio el *riboswitch* forma una estructura anti-terminadora de la transcripción, con lo que la transcripción y traducción del mRNA llegan a término y se expresan los genes de la biosíntesis de FMN. Sin embargo, cuando existen FMN en el citoplasma, éstos se unen al *riboswitch* y se produce un cambio conformacional que libera al terminador de la transcripción, no produciéndose expresión del ORF (Serganov, 2009). (Fig. 5C).

Además de éste, existen otros muchos *riboswitches* bien documentados con funciones diversas como el sensado de la temperatura (termosensores), el control de la respuesta frente a iones (sensores de magnesio) o el metabolismo de aminoácidos (sensores SAM y LYS que intervienen en respuesta frente a S-adenosilmetionina y lisina) (Serganov *et al.*, 2009).

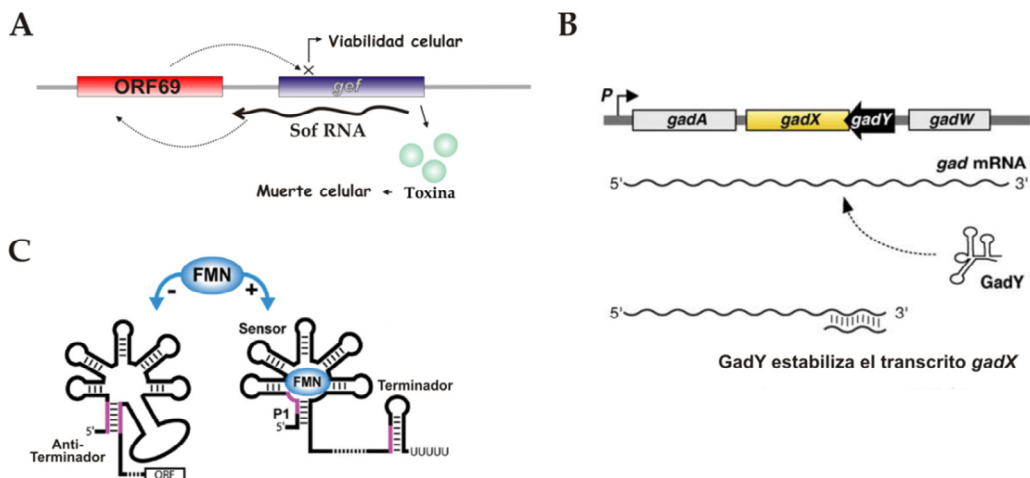


Fig. 5. Mecanismos de acción de los *cis*-sRNAs. A) Regulación del gen *gef* por parte del RNA antisentido Sof. B) Mecanismo de estabilización del mRNA *gadX* por parte de GadY, mediante unión a su región 3' no traducida. Tomado de Frohlich and Vogel, 2009. C) Funcionamiento del riboswitch FMN. En ausencia del ligando, el mRNA forma una hélice que actúa como anti-terminador y la transcripción del ORF puede llevarse a cabo. En presencia del ligando, el dominio sensor une el metabolito en cuestión produciendo la formación de una hélice terminadora en la región codificante, que inhibe la transcripción. Tomado de Seganov, 2009.

#### 4.2. sRNAs QUE ACTÚAN EN TRANS (TRANS-sRNAs)

Muchos de los sRNAs de reciente descubrimiento son codificados en *loci* alejados del lugar donde se transcribe el mensajero diana aunque también el apareamiento sRNA-mRNA es el mecanismo mayoritario de acción de estos *trans*-sRNAs. Sin embargo, una excepción destacable y muy bien caracterizada de este mecanismo general de ribo-regulación es el que presentan los sRNAs 6S y CsrB, también catalogables como *trans*-sRNAs. Estos sRNAs ejercen su actividad reguladora por secuestro y modificación de la actividad de proteínas específicas, pues presentan secuencias o motivos que éstas reconocen en los componentes celulares con los que normalmente se asocian, compitiendo así con ellos. A continuación se describen con detalle estos dos mecanismos de *trans* ribo-regulación con ejemplos representativos.

##### 4.2.1. *Trans*-sRNAs que modifican la actividad de proteínas

El RNA 6S fue uno de los primeros RNAs no codificantes identificados a finales de los años sesenta, aunque su función ha sido descubierta recientemente (Hindley, 1967; Brownlee, 1971). Este sRNA, codificado por el gen *ssrS*, está conservado en todos los grupos bacterianos y se expresa durante todas las fases de crecimiento de las bacterias, aunque presenta una acumulación preferente en fase estacionaria (unas 10.000 copias por célula). En estas condiciones, el 90% de las moléculas de 6S se

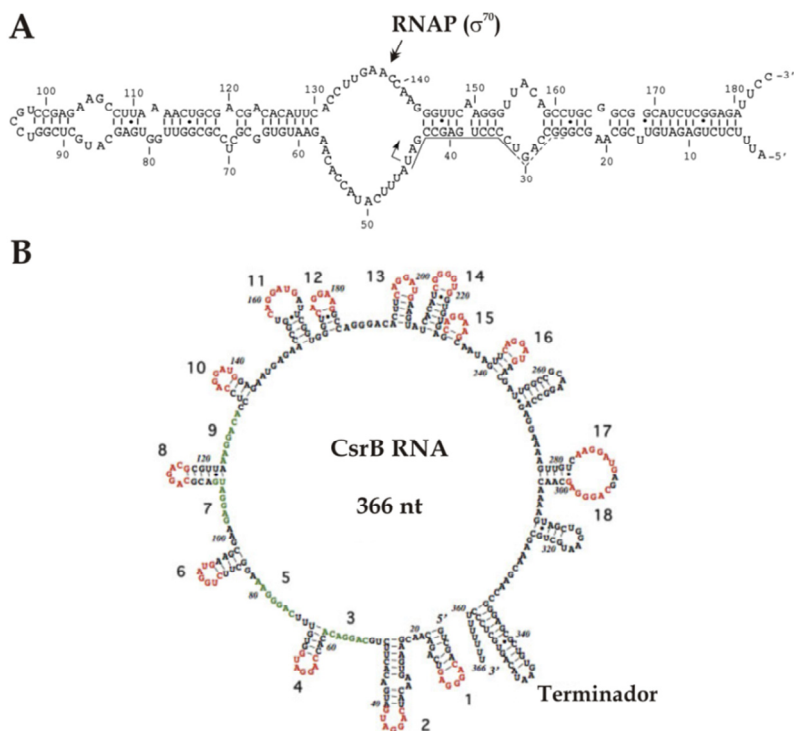
encuentran unidas a la subunidad  $\sigma^{70}$  de la RNA polimerasa impidiendo la actividad de ésta (Wassarman and Storz, 2000).

El RNA 6S de *E. coli* tiene un tamaño de 184 nt y forma una única horquilla con varios bucles que exponen segmentos de cadena sencilla (Fig. 6A), que son indispensables para su asociación con  $\sigma^{70}$  (Barrick *et al.*, 2005). La unión de 6S se localiza en el sitio activo de la polimerasa, es decir, donde ocurre la interacción de ésta con secuencias promotoras en el DNA, lo que sugiere que este sRNA actuaría como un competidor con promotores abiertos dependientes de  $\sigma^{70}$ , impidiendo la transcripción desde los mismos. (Cavanagh *et al.*, 2008).

En consecuencia, durante la fase estacionaria, en *E. coli* y otras enterobacterias relacionadas, la transcripción pasa a ser mayoritariamente dependiente de  $\sigma^{38}$  (RpoS), reconocido como uno de los factores principales de respuesta a estrés en estos grupos bacterianos.

Por su parte, el sRNA **CsrB**, también identificado y caracterizado inicialmente en *E. coli*, posee motivos específicos para la interacción con CsrA, una proteína de unión a RNA que interviene en la regulación de la traducción y la estabilidad de mRNAs (Liu *et al.*, 1997). Recientemente se han descubierto más sRNAs capaces de secuestrar CsrA, de forma que CsrB y sus ortólogos conforman lo que se conoce como familia de reguladores tipo CsrB (Timmermans and Van Melderen, 2010).

La rápida degradación de los RNAs mensajeros es en parte un efecto indirecto de la inhibición de la traducción y la familia de proteínas tipo CsrA se caracteriza por tener funciones relacionadas con este proceso. En *E. coli*, CsrA es una proteína soluble de aproximadamente 18 kDa que suele asociarse formando dímeros (Gutierrez *et al.*, 2005). El dímero de CsrA o el de sus ortólogos exponen dos superficies susceptibles de unirse a RNAs, concretamente a secuencias que contengan el motivo GGA (presente en secuencias Shine-Dalgarno). La estructura del sRNA CsrB está formada por una sucesión de horquillas en las que se encuentran los 18 motivos de unión a CsrA (Fig. 6B). El efecto principal de este sRNA es el reclutamiento de proteínas tipo CsrA en cada motivo de unión produciendo una disminución de la presencia efectiva de la proteína en la célula. Esta interacción es competitiva, siendo más efectiva cuando más moléculas de CsrA estén unidas al sRNA lo que resulta en una expresión de los mRNAs dependientes de CsrA dado el cese del proceso inhibitorio de su traducción (Babitzke and Romeo, 2007). Dependiendo del organismo, el secuestro de CsrA puede influir en determinadas rutas metabólicas, siendo en general afectadas vías reguladoras generales (adaptación a estrés, *quorum sensing*, virulencia, etc)



**Fig. 6. *Trans*-sRNAs que modifican la actividad de proteínas.** A) Representación de la estructura secundaria del 6S de *E. coli*. La doble flecha indica el sitio preferente de unión de la RNA polimerasa (RNAP). Tomado de Wassarman and Storz, 2000. B) Estructura secundaria del sRNA CsrB de *E. coli*, señalando en rojo los 18 motivos de unión a la proteína CsrA. El terminador de la transcripción independiente de Rho también se indica. Modificado de Babitzke and Romeo, 2007

#### 4.2.2. *Trans*-sRNAs que actúan por apareamiento de bases

Como ya se ha anticipado, el mecanismo más común de la *trans* ribo-regulación es el de la interacción por complementariedad de bases entre el sRNA y el mRNA diana. A diferencia de lo que ocurre con los RNAs antisentido en *cis* este apareamiento no suele ser completo, sino que implica a series cortas y discontinuas de nucleótidos en ambas moléculas (Vogel and Wagner, 2007). En este caso, la complementariedad es con frecuencia imperfecta de forma que además de las interacciones Watson-Crick pueden existir otras no canónicas y, en algunos casos, incluso de tipo terciario. El efecto del apareamiento *trans* sRNA-mRNA es la alteración de la tasa de traducción y/o la estabilidad del mRNA diana siendo éste negativo en la mayoría de los casos mejor documentados de ribo-regulación post-transcripcional, es decir, la actividad del sRNA generalmente produce una disminución en la expresión del gen diana. Todos los sRNAs de función conocida que actúan en *trans* mediante este mecanismo necesitan en diferente medida de su asociación con la chaperona de RNA Hfq para ser activos. La biología de esta

proteína será tratada en profundidad en el Capítulo 2 de esta memoria (Valentin-Hansen *et al.*, 2004).

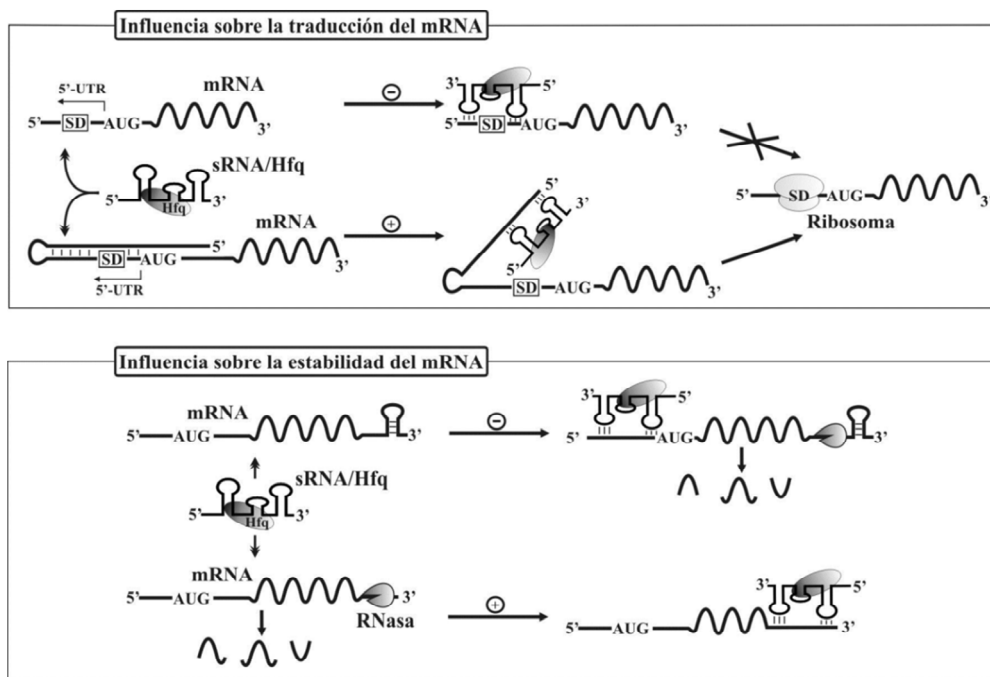


Fig 7. Efectos ejercidos por los *trans*-sRNAs que funcionan por apareamiento de bases sobre sus mRNAs diana.

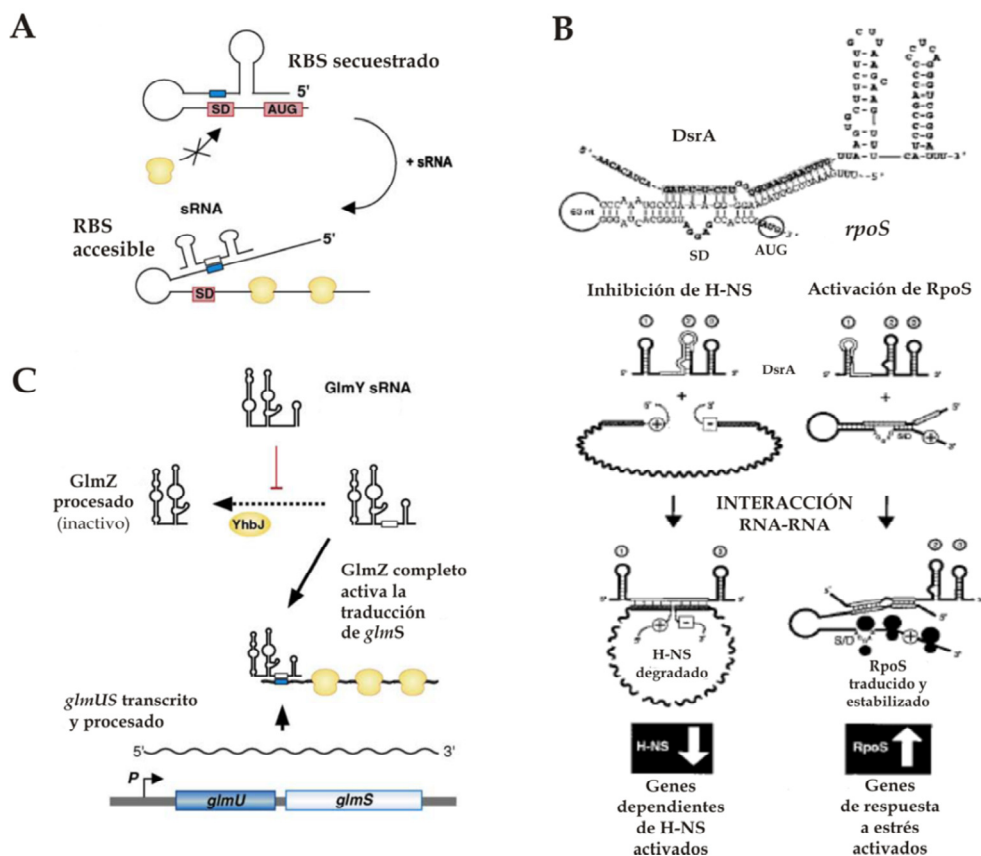
En bacterias, la traducción es iniciada en los ribosomas por el reconocimiento de secuencias específicas en el RNA mensajero como el sitio Shine-Dalgarno (SD) o de unión a ribosoma (RBS, *Ribosome Binding Site*) y el codón de iniciación AUG. En general, los sRNAs interaccionan con sus mRNA diana en las regiones 5'-UTR de éstos y más concretamente en el entorno del RBS mediante secuencias de reconocimiento que incluyen motivos anti-SD. De esta forma se bloquea el acceso de los ribosomas al mRNA y la subsiguiente traducción de éste. En casi todos los casos el bloqueo de la traducción va también acompañado de un efecto negativo sobre la estabilidad del mRNA de forma que el dúplex sRNA-mRNA suele actuar como señal de reconocimiento de ribonucleasas celulares como RNasaE o RNasaIII, que podrían ser reclutadas por Hfq.

Sin embargo, la región líder de algunos mRNAs se estructura de forma que secuestra el RBS haciéndolo inaccesible a los ribosomas. Como ya se ha descrito, esta

región 5'-UTR funciona como *riboswitch* en algunos mRNAs, pero en otros la interacción con un *trans*-sRNA es la que elimina la estructura secundaria inhibitoria haciendo posible la traducción del mensajero. La inducción de la traducción mediada por sRNAs es menos frecuente que el bloqueo de ésta pero no es excepcional, de forma que son también numerosos los casos de interacción sRNA-mRNA con efectos positivos sobre la expresión génica (Urban and Vogel, 2007).

Tampoco es excepción sino característica común a muchos de los *trans*-sRNAs mejor caracterizados, que contribuye a su reconocida versatilidad funcional, su capacidad para regular múltiples mRNAs, incluso con efecto diferente sobre la traducción de cada uno de ellos. Por ejemplo, DsrA, un sRNA dependiente de Hfq que se acumula en *E. coli* en respuesta a la baja temperatura estimula la traducción de RpoS mediante apareamientos con su región 5'-UTR que imposibilitan la formación de horquillas inhibitorias en torno al RBS. Sin embargo, este sRNA contiene otras series de nucleótidos que son también complementarias a secuencias situadas en la región líder del mRNA de *hns*, que codifica un regulador transcripcional general asociado a nucleóide, y su interacción con éstas inhibe su traducción (Repoila *et al.*, 2003). Aunque RpoS y Hns son ambas proteínas reguladoras de la transcripción, no tienen una aparente relación funcional. Otros sRNAs, sin embargo, utilizan un único motivo o serie de nucleótidos para interactuar con secuencias conservadas en las regiones 5'-UTR de mRNAs que codifican proteínas muy relacionadas funcionalmente. Es el caso de GcvB, identificado en *Salmonella*, que reconoce elementos ricos en C/A conservados en las regiones 5'-UTR de muchos mRNAs que codifican el componente periplásmico de transportadores ABC de péptidos y aminoácidos que en ausencia de este sRNA podrían actuar como amplificadores de la traducción. En las condiciones en las que GcvB se expresa su interacción en esa zona de los mensajeros inhibe la traducción de todas esas proteínas de transporte (Sharma *et al.*, 2007). Por el contrario, la activación directa de múltiples mRNAs por un único sRNA es un fenómeno no descrito todavía (Papenfort and Vogel, 2009).

Tampoco es infrecuente, y está bien caracterizada en muchos casos, la regulación de un mRNA por múltiples sRNAs. En *E. coli*, RpoS, además de ser regulado positivamente por DsrA, lo es también mediante un mecanismo similar por RprA. Por el contrario, el sRNA OxyS inhibe la traducción de RpoS por un mecanismo todavía desconocido que probablemente implique también interacción con su región 5'-UTR (Repoila *et al.*, 2003). De forma similar, en *Vibrio cholerae*, hasta cuatro sRNAs (Qrr) inhiben la expresión del regulador de quórum sensing *hapR* por oclusión del RBS en el mRNA (Tu and Bassler, 2007).



**Fig 8. Ejemplos de regulación de *trans*-sRNAs que actúan mediante apareamiento de bases.** A) Esquema general de activación de la traducción del mRNA diana mediante desorganización de estructuras inhibitoras situadas su la 5'-UTR. B) Efecto dual del sRNA DsrA sobre los mRNAs H-NS y RpoS. En la parte superior se muestra la interacción DsrA-RpoS en el motivo Shine-Dalgarno. La sección inferior muestra el mecanismo de inhibición de H-NS (izquierda) por occlusión del sitio de unión a ribosoma, que desencadena la rápida degradación del mRNA y la represión de los genes derivados del mismo. Por el contrario, la unión con RpoS supone la disociación de la estructura inhibitora de la región 5'-UTR y la estabilización del mRNA, desencadenando la inducción de los genes relacionados con la respuesta a estrés mediados por *rpoS*. C) Mecanismo de activación de la traducción del mRNA policistrónico *glmUS* mediado por GlmZ (explicación en el texto). Figura adaptada de Frohlich and Vogel, 2009 (A y C) y de Repoila and Gottesman, 2003 (B).

Además de la interacción con regiones 5'-UTR de transcritos monocistrónicos o de los primeros genes de operones, los sRNAs también pueden reconocer secuencias complementarias en mensajeros policistrónicos para inhibir o activar selectivamente la síntesis de algunas proteínas codificadas por el operón. Es el caso del sRNA Spot42, determinante de la descoordinación de la expresión de genes del operón *galETKM* de utilización de galactosa en *E. coli*. Concretamente, una parte de la región 5' de Spot42 establece apareamientos con *galK*, el tercer cistron de este operón, bloqueando el RBS e inhibiendo su traducción sin alterar los niveles de las proteínas GalE y GalT codificadas por los dos primeros genes (Moller *et al.*, 2002). De forma



inversa, el sRNA GlmZ, cuya función ha sido investigada en *Salmonella*, utiliza una región de 15 nt para establecer interacciones entre los dos cistrones integrantes del operón del metabolismo de amino-azúcares *glmUS* estabilizando y favoreciendo así la traducción de GlmS, cuyo RBS está normalmente secuestrado por una horquilla. En este caso la acción de GlmZ posibilita una tasa similar de traducción de GlmU y GlmS. (Kalamorz *et al.*, 2007; Urban and Vogel, 2008).

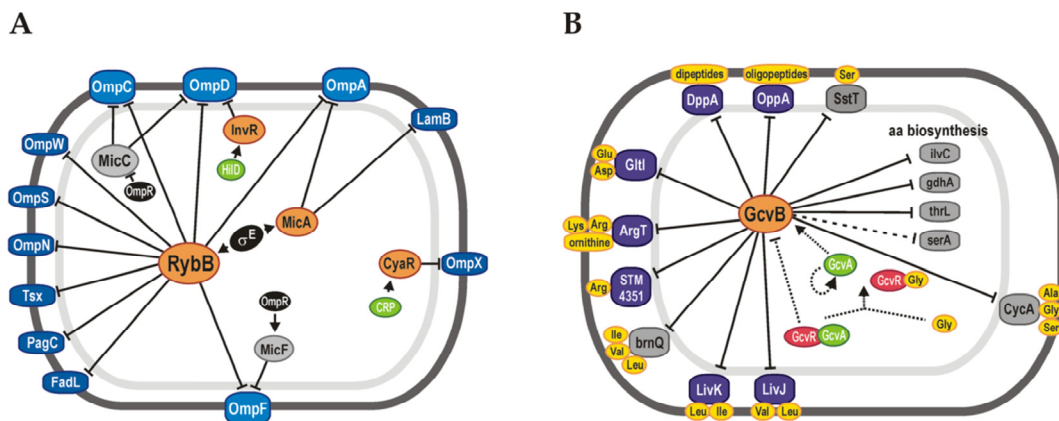
## 5. FISIOLÓGÍA DE LA RIBO-REGULACIÓN

La versatilidad del mecanismo de acción de los sRNAs se traduce en su participación en el control de multitud de procesos, de forma que su actividad tiene un gran impacto en la fisiología celular. En general, una amplia mayoría de los sRNAs de los que se conoce su función están implicados en procesos relacionados con la respuesta fisiológica frente a estrés a tres niveles: *i*) modificación de la expresión de proteínas de la membrana, *ii*) reprogramación del metabolismo o *iii*) modulación la síntesis de factores de transcripción.

### i) Efecto sobre proteínas de membrana

Este grupo engloba el mayor número de sRNAs con función asignada. La capacidad de adaptarse a un medioambiente cambiante requiere entre otras respuestas del adecuado ensamblaje, reordenamiento y reciclado de las proteínas estructurales de la membrana externa (OMPs; Outer Membrane Proteins). En *S. typhimurium* este proceso es regulado por al menos seis sRNAs; i.e. RybB, MicC, InvR, MicF, MicA y CyaR. RybB y MicA son componentes del regulón del factor sigma alternativo  $\sigma^E$  que se induce ante estreses diversos como el shock térmico, la hiperosmolaridad, la entrada en fase estacionaria, la exposición a péptidos antimicrobianos o el pH ácido. RybB inhibe rápida y simultáneamente la síntesis de muchas de las principales OMPs (OmpA, C, D y F) y varias de las minoritarias (e.g. FadL, Tsx o PagC) mediante apareamiento de bases con las 5'-UTR de sus mensajeros, mientras que MicA reprime OmpA y LambB por el mismo mecanismo (Bouvier *et al.*, 2008; Vogel and Papenfort, 2006; Papenfort *et al.*, 2006; Udekwu *et al.*, 2005; Figueroa-Bossi *et al.*, 2006). La transcripción de MicC y MicF, sin embargo, depende del regulador OmpR y su acumulación en la célula reprime la expresión de OmpC/D y OmpF, respectivamente. Por su parte, el factor de transcripción HilD promueve la expresión de InvR para contribuir a la inhibición de la traducción y a la degradación del mensajero de *ompD*. Finalmente, CyaA es inducido por un receptor de cAMP (CRP) para reprimir específicamente la síntesis de OmpX, conectando así procesos de dinámica de los componentes de membrana con el metabolismo de fuentes de energía (Fig. 9A). La inhibición masiva de nuevos componentes

estructurales de la membrana previene el ensamblaje de éstos sobre proteínas dañadas por el estrés liberando la maquinaria celular de translocación y plegamiento de polipéptidos para optimizar el funcionamiento de ésta en condiciones que requieren el remodelado rápido y eficiente de la cubierta celular. En bacterias no estresadas la expresión basal de RybB y MicA contribuye al mantenimiento de la homeostasis de la membrana garantizando así la autorregulación de  $\sigma^E$  (Papenfort *et al.*, 2006).



**Fig. 9. Riboregulación de proteínas de membrana y transportadores ABC en *Salmonella*.** A) Red de control de producción de las proteínas de membrana externa (OMPs, círculos azules). Los sRNAs reguladores (en naranja) muestran su papel inhibitor (línea quebrada) sobre su diana. Los factores de transcripción de estos sRNAs se representan en negro o en verde, según el caso (CRP reguladores generales, HliD reguladores específicos de virulencia). B) mRNAs afectadas por GcvB. En morado se muestran los componentes periplásmicos afectados por este sRNA. Los círculos grises indican genes adicionales implicados en la síntesis de aminoácidos. Los aminoácidos y péptidos afectados se muestran en amarillo. Tomado de Vogel, 2009.

Otro nivel fundamental en el control de la respuesta frente a cambios en el medio extracelular es la modulación de la expresión de los transportadores ABC. Uno de los sRNAs mejor caracterizados e implicados en este proceso es el ya mencionado GcvB (Fig. 9B). Este ribo-regulador inhibe directamente la expresión de hasta diez mRNAs que codifican las proteínas periplásmicas de unión a sustrato de transportadores ABC de aminoácidos y péptidos diversos en *E. coli* y en *Salmonella* (Urbanowski *et al.*, 2000; Sharma *et al.*, 2007). GcvB se acumula preferentemente en condiciones de crecimiento rápido, cuando la concentración de nutrientes no es limitante y no existe necesidad de un transporte específico (Sharma *et al.*, 2007).

### ii) Efecto sobre la reprogramación del metabolismo

Los sRNAs no sólo regulan los procesos de captación de nutrientes sino que también pueden controlar directamente la concentración de enzimas del metabolismo, contribuyendo a la reprogramación de éste ante cambios externos en la

disponibilidad de fuentes de energía. GcvB, por ejemplo, extiende su actividad hasta el control de sistemas de biosíntesis y utilización de aminoácidos, como la glicina (Urbanowski *et al.*, 2000).

Otro regulador que interviene en la síntesis de enzimas metabólicas es el ya mencionado Spot42, cuya expresión depende de CRP, que es reprimido en condiciones de carencia de glucosa (Moller *et al.*, 2002). Igualmente, el sRNA FnrS cuya transcripción depende del regulador de la nitrato reductasa (FNR), se induce en condiciones de baja presión de oxígeno, reprimiendo enzimas como la lactato deshidrogenasa que no son necesarias durante el crecimiento anaeróbico (Boysen *et al.*, 2010; Durand and Storz, 2010).

El sRNA RsaE, identificado en *S. aureus* y en *Bacillus subtilis*, controla la expresión de PycA (piruvato carboxilasa), Folds (proteína bifuncional Folds) y Fhs (formato-tetrahidrofolato ligasa), todas enzimas clave del metabolismo general. En *B. subtilis* RsaE está además involucrado en el control de CtsA (proteína inducida en carencia de carbono) y GalK (galactoquinasa), responsables de la utilización de carbohidratos y azúcares, respectivamente (Geissmann *et al.*, 2009).

En *Pseudomonas aeruginosa*, la oscilación de los niveles del sRNA CrcZ y el mRNA AmiE (acilamida amidohidrolasa) son regulados de forma precisa por varios sustratos energéticos. Normalmente, en condiciones no limitantes de crecimiento, la bacteria suele utilizar preferentemente succinato como fuente de carbono, manteniéndose el mensajero del gen *amiE* bajo represión catabólica. En condiciones de crecimiento sobre manitol, el factor sigma *rpoN* y el sistema de dos componentes CbrA/CbrB pueden activar la expresión de CrcZ y éste, gracias a sus motivos CA es capaz de reclutar a la proteína Crc, que inhibe la traducción de AmiE (de forma análoga al sistema CsrA/CsrB,C). El aumento de los niveles de AmiE alivia la represión catabólica (Sonnleitner *et al.*, 2009).

### iii) Efecto sobre la síntesis de factores de transcripción

La síntesis del factor de fase estacionaria  $\sigma^{38}$ , codificado por el gen *rpoS* fue uno de los primeros mensajeros que se catalogaron como ribo-regulados en *E.coli*. Este componente de la RNA polimerasa es fundamental en la respuesta a estrés y representa el punto de confluencia de varias cascadas de regulación en las que participan sRNAs cuya expresión depende de cambios en el medio externo.

Como ya se ha mencionado RpoS es regulado positivamente por DsrA y RprA y negativamente por OxyS (Majdalani *et al.*, 2005). Recientemente se ha encontrado que

otro sRNA, denominado ArcZ, incrementa los niveles intracelulares de  $\sigma^{38}$  (Mandin and Gottesman, 2010). DsrA y RprA son inducidos en la respuesta a bajas temperaturas y al estrés celular de superficie, respectivamente, mientras que OxyS es inducido en condiciones de estrés oxidativo y ArcZ se reprime en anaerobiosis. Otros sRNAs regulan sistemas de dos componentes, como por ejemplo OmrA y OmrB, que afectan negativamente la expresión del componente quinasa del regulador *ompR*, que a su vez controla la síntesis de los dos sRNAs en un circuito retro-alimentado (Guillier and Gottesman, 2008). Finalmente, otro ejemplo destacable en este grupo de sRNAs es el ya citado MicA, que regulado por  $\sigma^E$ , conecta la vía de respuesta al estrés de membrana externa con el regulón PhoPQ mediante la inhibición de la traducción de la quinasa sensora (Coornaert *et al.*, 2010).

### 5.1. sRNAs Y PATOGÉNESIS

El éxito ecológico de las bacterias patógenas depende en gran medida de su capacidad de adaptación medioambiental tanto dentro como fuera del huésped. La respuesta de los patógenos a los cambios en el medio afecta a procesos celulares tan diversos como la captación de nutrientes, la resistencia al estrés oxidativo, el control de la homeostasis del hierro, la biosíntesis de toxinas o la supervivencia intracelular, que en muchos casos son ribo-regulados. La presencia en un genoma de determinados grupos de sRNAs reguladores de alguno de estos procesos puede incluso estar específicamente ligado al carácter patógeno de una especie. Un ejemplo serían los genes sRNA denominados *rli* (*rliA*, *rliC*, *rliF*, *rliG*, y *rliB*) que se encuentran en *Lysteria monocytogenes* pero no en especies del mismo género como *L. innocua* que no lo son. No obstante, muchos de los sRNAs identificados y caracterizados en bacterias patógenas participan en la regulación de procesos de adaptación generales que no son necesariamente específicos de un fenotipo de virulencia. Es el caso de la reprogramación del metabolismo o la dinámica de las proteínas de membrana ya descritos. A continuación se describe con más detalle cómo son ribo-regulados los procesos que afectan más directamente la virulencia de algunas bacterias patógenas (Fig. 10).

#### **5.1.1. Regulación del quorum sensing**

Las bacterias utilizan el *quorum sensing* para sensar la densidad de la población y comunicarse entre ellas de forma que la expresión coordinada de determinados genes a nivel poblacional es clave en la virulencia. El patógeno humano *V. cholerae* posee varios sistemas de *quorum sensing*; i.e. CAI-1-CqsS, AI-2-LuxPQ, y VarS/VarA-CsrA/BCD (Fig. 10, sector superior derecho). La cascada consta de un sistema sensor LuxP acoplado a una quinasa con capacidad de autofosforilarse y transferir el grupo

fosfato, en condiciones de baja densidad celular, al regulador de respuesta, LuxO, que en su forma fosforilada, junto con el factor sigma alternativo  $\sigma^{54}$ , activan cuatro sRNAs Qrr (*quorum regulatory RNAs*). Estos sRNAs inhiben la expresión de *hapR*, que codifica el regulador principal de la cascada de *quorum sensing*. A concentraciones celulares altas, se producen niveles de auto-inductores que transforman la actividad quinasa de los reguladores Lux en actividad fosfatasa. En esas condiciones los sistemas dependientes de LuxO reducen la expresión de los sRNAs Qrr aumentando la concentración del mRNA de *hapR*, con la consecuente activación de genes implicados en la formación de biopelículas y en la virulencia (Toledo-Arana *et al.*, 2007).

Los sRNAs Qrr también controlan en *V. cholerae* el sistema VarS/VarA-CsrA y modifican la actividad de LuxO por otro mecanismo independiente, el cual es desconocido actualmente, aunque guarda semejanzas con el sistema CsrB-CsrC/CrA de *E. coli* (Lenz *et al.*, 2005).

### **5.1.2. Regulación de la homeostasis de hierro**

Durante la infección del huésped por el patógeno la disponibilidad de hierro para éste suele estar muy limitada. El hierro, sin embargo, es un elemento traza indispensable para la mayoría de los organismos de forma que su concentración intracelular está finamente controlada para evitar los efectos tóxicos de una excesiva entrada en la célula. Uno de los sRNAs que controla la homeostasis de hierro es RyhB, que es reprimido por el regulador Fur. En *E. coli*, el hierro libre intracelular se asocia a Fur inhibiendo la expresión de los genes de la adquisición de hierro (*suf*) y la de RyhB lo que, a su vez, des-reprime la expresión de mRNAs diana de este sRNA que codifican proteínas esenciales para el almacenamiento de hierro como las ferritinas (*bfr*) u otras no esenciales que contienen hierro como la aconitasa B (*sodB*) o la succinato deshidrogenasa (Masse *et al.*, 2005) (Jacques *et al.*, 2006).

Se han encontrado ortólogos de RyhB en *V. cholerae*, *Shigella flexneri* o *Erwinia chrysanthemi* por lo que la ribo-regulación mediada por este sRNA parece ser un mecanismo general de control de la homeostasis del hierro en un gran número de patógenos (Davis *et al.*, 2005; Oglesby *et al.*, 2005; Boughammoura *et al.*, 2008).

### **5.1.3. Regulación de múltiples factores de virulencia**

Su capacidad para regular múltiples mRNAs con diferentes funciones hace que en algunos casos la actividad de un sRNA afecte simultáneamente a diversos factores de virulencia. Es el caso de RNAIII de *S. aureus*, el primer RNA no codificante

identificado en un patógeno. (Fig. 10, sector superior izquierdo). La virulencia de esta bacteria es determinada fundamentalmente por el sistema *agr* que consta de dos vías transcripcionales divergentes denominadas RNAII y RNAIII. RNAII consta del típico regulador de dos componentes (i.e. AgrA, regulador de respuesta y AgrC, componente sensor con actividad quinasa), un pro-péptido ArgD y una peptidasa AgrB. Esta última procesa AgrD, dando lugar a un péptido pequeño autoinducible, que une a las formas fosforiladas de AgrC. AgrA, AgrB, AgrC y AgrD, que conforman el sistema de *quorum sensing*. A altas concentraciones del autoinductor, se induce la transcripción de RNAIII, que es la molécula efectora del sistema *arg* (Novick and Geisinger, 2008). RNAIII controla la expresión de al menos tres mRNAs, el correspondiente al gen de la  $\alpha$ -hemolisina *hla*, el de la proteína A (Spa) y el del factor de transcripción Rot, induciendo la expresión del primero y reprimiendo la traducción de los otros dos. Otra particularidad de RNAIII es que parte de su secuencia codifica un péptido pequeño, la  $\delta$ -hemolisina (Hld). La expresión de ambas hemolisinas, junto con la represión de Spa produce la liberación de gran número de sustancias extracelulares, unidas a cambios profundos en factores de membrana (reprimidos por Rot) que hacen que esta bacteria tenga un gran potencial patógeno.

*P. aeruginosa* es una bacteria saprofítica que puede ser patógeno oportunista de la especie humana, causando infecciones severas en pacientes inmunocomprometidos. Forma biopelículas en multitud de sustratos dado que posee sistemas de adherencia muy eficientes (sistemas de secreción tipo III) y es capaz de producir un amplio espectro de sustancias extracelulares (toxinas) reguladas por N-acil-homoserín-lactonas, así como de metabolitos secundarios. El control de todos estos procesos está basado en numerosos sistemas interconectados, incluyendo las vías GacS-GacA-RsmY-Z-RsmA (Heurlier *et al.*, 2004; Burrowes *et al.*, 2005; Ventre *et al.*, 2006) (Fig. 10, sector inferior izquierdo). El sistema GacS-GacA-RsmY-Z-RsmA es homólogo al BarA-UvrY-CsrB-C-CsrA de *E. coli* y está conservado en numerosas especies bacterianas. Este sistema controla procesos relacionados con el metabolismo del carbono y la formación de biopelículas. En *P. aeruginosa*, el sistema de dos componentes GacS-GacA controla positivamente la expresión de dos sRNAs, RsmY y RsmZ, que son homólogos a CsrB-C de *E. coli* (Kay *et al.*, 2006; Kulkarni *et al.*, 2006). RsmZ está además bajo en control de LadS y RetS, dos proteínas sensoras capaces de actuar positiva y negativamente sobre GacA, respectivamente (Bejerano-Sagie and Xavier, 2007). Por su parte, RsmA es un regulador transcripcional que actúa tanto positiva como negativamente sobre la expresión de variedad de genes importantes en la virulencia y en la supervivencia, como los del metabolismo del carbono, quorum

sensing y sistemas de secreción tipo III, así como los requeridos para la formación de biopelículas (Heurlier *et al.*, 2006; Kay *et al.*, 2006; Mulcahy *et al.*, 2006).

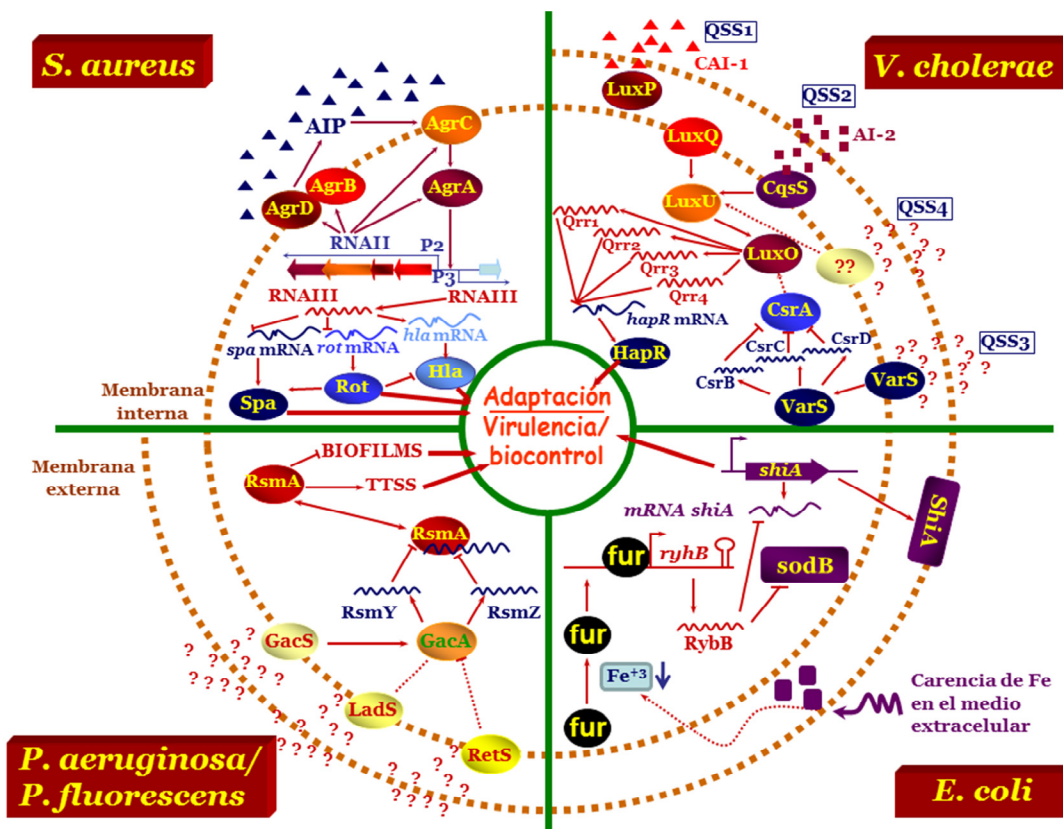


Fig. 10. Ribo-regulación de la patógenesis. Redes reguladoras implicadas en la virulencia de (de izquierda a derecha y de arriba a abajo) *S. aureus* (RNA III), *V. cholerae* (Qrr sRNAs), *P. aeruginosa* (RsmY, Z) y *E. coli* (RyhB). Los detalles de cada proceso se detallan en el texto.

## 6. SIMBIOSIS RHIZOBIUM-LEGUMINOSA

Las  $\alpha$ -proteobacterias constituyen un grupo heterogéneo de especies de bacterias Gram-negativas aunque muchas de ellas tienen en común la capacidad para establecer infecciones crónicas de distinta naturaleza en eucariotas superiores. (Batut *et al.*, 2004) (Fig. 11). En él podemos encontrar bacterias marinas oligotróficas (*Caulobacter sp.*), bacterias que actúan como promotores del crecimiento vegetal (*Rhodospirillum*, *Azospirillum*), patógenos de plantas (*Agrobacterium*) y parásitos intracelulares de animales (*Rickettsia*, *Brucella*, *Ehrlichia*).

Otras especies de  $\alpha$ -proteobacterias pertenecientes a los géneros *Rhizobium*, *Allorhizobium*, *Ensifer* (antes *Sinorhizobium*), *Mesorhizobium*, *Bradyrhizobium*, *Blastobacter*, *Azorhizobium*, *Devosia* y *Methylobacterium* del orden Rhizobiales (colectivamente denominadas rizobios o *Rhizobium*) son capaces de establecer simbiosis mutualistas específicas con plantas leguminosas. Estas interacciones culminan con la formación en la raíz (y excepcionalmente el tallo) de las plantas de nuevos órganos de histología organizada, denominados nódulos, donde el microsimbionte se establece en el interior de las células vegetales y se diferencia morfológicamente a bacteroide, la forma especializada de la bacteria con capacidad para reducir el dinitrógeno molecular a una forma combinada (amonio) asimilable por el vegetal. Este proceso está catalizado por una metaloenzima denominada nitrogenasa, cuyo funcionamiento precisa de un ambiente microaeróbico y consume grandes cantidades de energía, que las bacterias obtienen de la oxidación de dicarboxilatos suministrados por la planta (Prell and Poole, 2006; Trainer and Charles, 2006).

Una importante peculiaridad de esta interacción es su elevado grado de especificidad, de modo que, generalmente, cada rizobio es capaz de interactuar con una o muy pocas especies vegetales, normalmente muy relacionadas filogenéticamente. Sin embargo, hay también rizobios que presentan un rango de hospedador amplio, como *Rhizobium sp.* NGR234, capaz de nodular al menos 112 géneros de leguminosas y una especie no leguminosa como *Parasponia* (Pueppke and Broughton, 1999). La especificidad de la simbiosis es el resultado del reconocimiento entre los dos socios, basado en un intercambio de señales bioquímicas que desencadenan la inducción de programas genéticos específicos en ambos organismos.



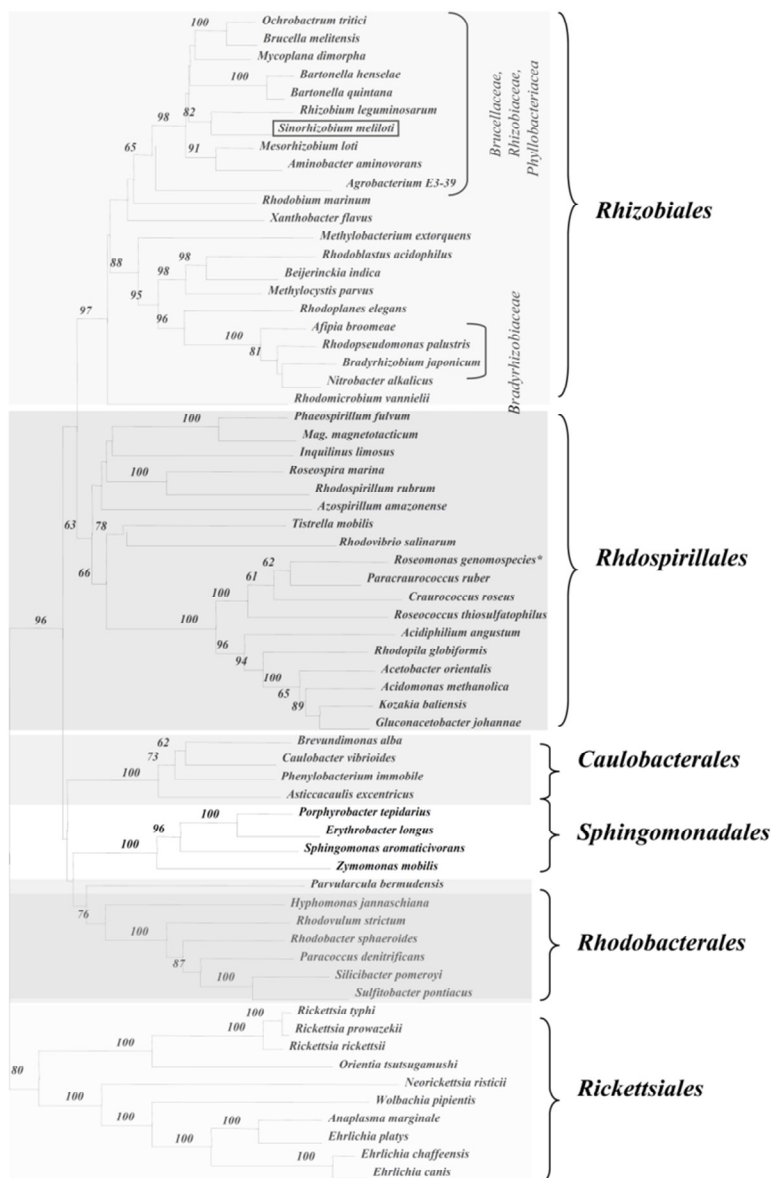


Fig. 11. Árbol filogenético 16S del grupo de las α-proteobacterias.

### 6.1. EL PROCESO SIMBIÓTICO

En el ciclo vital de los rizobios se suceden, por tanto, dos etapas; una de vida libre en suelo y rizosfera y otra endosimbiótica en el interior de las células vegetales. El mecanismo más general y mejor conocido para el establecimiento de la simbiosis consta de una secuencia de etapas que se ilustran en la Fig. 12 y se describen a continuación.

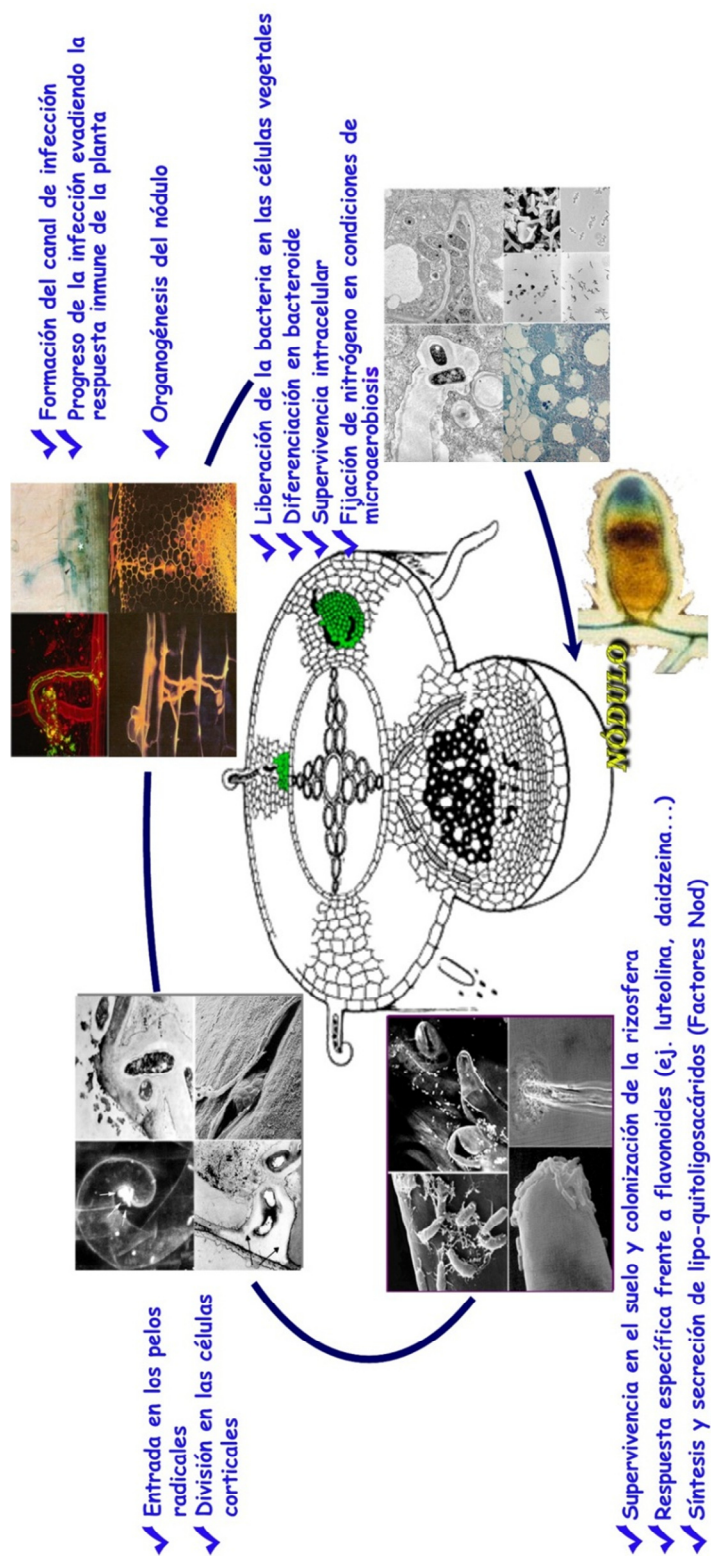


Fig. 12. Etapas de la simbiosis *Rhizobium-leguminosa*. La imagen muestra el corte transversal de una raíz esquematizando las etapas de la organogénesis nodular que se ilustran con las fotografías obtenidas por diferentes técnicas microscópicas.

### **6.1.1. Colonización rizosférica e infección**

La competencia saprofítica y simbiótica de *Rhizobium* en suelo y rizosfera va a estar en gran medida determinada por la capacidad de adaptación de las bacterias a los diferentes estreses abióticos a los que debe enfrentarse. Estas adaptaciones van generalmente acompañadas de cambios en los componentes estructurales de pared celular así como de sus polisacáridos asociados, que participan en diferente medida en los procesos de osmoadaptación, adhesión a superficies o comunicación intercelular.

Como microorganismos del suelo los rizobios son metabólicamente versátiles, lo que les permite utilizar una amplia variedad de nutrientes. En este sentido se han identificado y caracterizado un gran número de genes de transporte, catabólicos o biosintéticos de moléculas diversas como aminoácidos y derivados, azúcares, compuestos aromáticos, rizopinas o incluso compuestos antibióticos que en diversa medida confieren a las bacterias ventajas competitivas en la colonización de la rizosfera y por la nodulación.

En los primeros estadios de la simbiosis las bacterias mejor adaptadas colonizan eficientemente la rizosfera de su leguminosa huésped y al alcanzar cierta densidad de población pueden responder colectivamente a los estímulos medioambientales mediante *quorum sensing*. En *Rhizobium* se han identificado ortólogos de las proteínas LuxI y LuxR que definen sistemas de *quorum sensing* en las diferentes especies que responden a las acil homoserín lactonas u otras moléculas autoinductoras para la regulación de genes que especifican funciones simbióticas diversas como la síntesis de bacteriocinas, la transferencia de plásmidos, la síntesis de exopolisacáridos (EPS) o la expresión de proteínas rizosféricas en etapas simbióticas tempranas y la nodulación, el desarrollo del simbiosoma, o la fijación de nitrógeno en etapas intermedias y tardías de la interacción simbiótica (Jones *et al.*, 2007).

En la proximidad de los pelos radiculares los rizobios perciben los flavonoides exudados por la raíz que inducen la expresión de los genes de nodulación mediada por el activador transcripcional, de tipo LysR, conocido como NodD. Los productos de los genes de nodulación están implicados en la síntesis y secreción de lipoquitooligosacáridos (factores Nod), constituidos por un esqueleto común de entre 4 y 6 unidades de N-acetilglucosamina, y diversas sustituciones (e.g. ácidos grasos, fucosa, grupo sulfato y otros), tanto en el extremo reductor como en el no reductor de la molécula. Estas decoraciones son los determinantes esenciales de la especificidad de la simbiosis. A su vez, los factores Nod inducen en la planta una serie de respuestas morfológicas que culminan con la organogénesis del nódulo, entre ellas,

la curvatura de los pelos radiculares, la formación de los cordones de pre-infección y la división de las células corticales de la raíz para la formación del primordio nodular (Jones *et al.*, 2007).

Las bacterias quedan atrapadas en la curvatura de los pelos radiculares y penetran en los mismos a través de erosiones localizadas en su pared celular, mediadas por celulasas de origen bacteriano (Robledo *et al.*, 2008). Posteriormente, se desplazan y dividen a lo largo del pelo radicular en los cordones de infección. Las leguminosas responden a la infección por *Rhizobium* mediante la producción de especies reactivas de oxígeno (ROS), como peróxido de hidrógeno o el radical superóxido, que las bacterias pueden neutralizar mediante la activación de potentes sistemas antioxidantes en los que, entre otras, catalasa y superóxido dismutasa son las enzimas efectoras. La función de ROS en la reacción de defensa es doble; por una parte media el fortalecimiento de la pared celular para formar barreras frente a la penetración de los patógenos y por otra, y más importante, participa como señal intermediaria en la activación de gran parte del resto de reacciones de defensa que, junto a ROS, conforman la denominada reacción de hipersensibilidad. Entre éstas se pueden citar reacciones locales como la producción de fenilpropanoides, fitoalexinas y diferentes proteínas de defensa, que provocan necrosis y muerte celular en el sitio de infección, y reacciones sistémicas como la acumulación de ácido salicílico, que media en la adquisición de la resistencia sistémica. Este segundo grupo de respuestas no parece producirse en interacciones *Rhizobium*-leguminosa compatibles pero es evidente en infecciones abortadas, en determinadas combinaciones no compatibles, y fundamentalmente después de la inoculación de mutantes alterados en la síntesis de lipopolisacáridos (LPS), EPS y glucanos cíclicos. Este conjunto de moléculas forma parte, por tanto, de otro circuito no caracterizado de señalización planta-bacteria, para la inhibición o control de las respuestas de defensa de la planta que permita el progreso de la simbiosis (Soto *et al.*, 2006).

### **6.1.2. Diferenciación morfológica del bacteroide y fijación de nitrógeno**

Una infección efectiva culmina con la liberación de las bacterias en las células del primordio nodular y su subsiguiente diferenciación a bacteroides que, en su forma madura, quedan finalmente rodeados por una membrana de origen vegetal formando así el simbiosoma (Limpens *et al.*, 2005). La disección genético-molecular de este proceso es una de las grandes lagunas en nuestro conocimiento de la simbiosis. Su biología está caracterizada por un cese de la división celular acompañada de una activa endorreduplicación del material genético y un drástico

cambio en la morfología de las bacterias que se establecen en el medio intracelular al que deben adaptarse (Van de Velde *et al.*, 2010).

El control del ciclo celular de los rizobios está escasamente caracterizado aunque por su homología con los de otros organismos se han identificado varios genes que podrían estar involucrados en el proceso como *fts*, que se expresa durante la mitosis en *E. coli*, *dnaA*, cuya función está relacionada con la iniciación de la replicación del DNA o *ctrA*, un regulador transcripcional que controla diversas etapas del ciclo celular en *C. crescentus*, por citar algunos de ellos. No obstante la función precisa de los mismos en el inicio de la diferenciación del bacteroide es desconocida.

Por su parte, el establecimiento intracelular y la diferenciación morfológica requieren, entre otras, adaptaciones metabólicas y nuevos cambios estructurales en la cubierta celular. En el simbiosoma los procesos de metabolismo y fijación de nitrógeno están desacoplados a nivel genético, por lo que se produce la inhibición de los genes de asimilación de amonio y la inducción de otros que controlan la utilización de fuentes de energía alternativas como los ácidos dicarboxílicos (Fischer, 1994).

El marcador molecular clásico de este proceso de diferenciación morfológica es el gen *bacA*. Los mutantes de *Rhizobium* en este gen presentan un fenotipo simbiótico caracterizado por una infección abortada en el momento en que las bacterias son liberadas en las células del primordio nodular. *BacA* codifica una proteína que inicialmente fue caracterizada como transportador de péptidos con actividad antibiótica, que confiere resistencia a pH ácido y en *Rhizobium* determina una distribución específica de las cadenas lipídicas de los LPS. Recientemente se ha demostrado que son precisamente péptidos de origen vegetal los desencadenantes de este proceso de diferenciación morfológica (Van de Velde *et al.*, 2010).

La simbiosis culmina con la formación de los nódulos radiculares de los que planta y bacteria se benefician mutuamente. Los nódulos maduros pueden clasificarse en general en dos tipos estructurales; determinados, típicos de leguminosas de origen tropical, de forma redondeada sin meristemo apical y ocupados por una zona central de fijación de nitrógeno en la que bacterias y células vegetales presentan el mismo estadio de diferenciación, e indeterminados, de forma alargada propios de leguminosas de clima templado, que presentan un meristemo apical activo o Zona I y otras zonas subyacentes II-IV a través de las que se observa un gradiente de estados de diferenciación de bacterias y células vegetales, siendo la zona III la de fijación de nitrógeno ocupada por bacteroides completamente diferenciados.

Entre todos los productos de la actividad fotosintética de las plantas los simbiosomas utilizan preferentemente los ácidos dicarboxílicos como fuente nutritiva para generar mucha de la energía que demanda el proceso de fijación de nitrógeno. El nitrógeno fijado es exportado posteriormente a la planta como amidas en nódulos indeterminados o ureidos en nódulos determinados. La reducción del nitrógeno molecular a amonio por la nitrogenasa es, por tanto, el proceso biológico más relevante que ocurre en el nódulo. La sensibilidad de este enzima al oxígeno determina que la microaerobiosis sea la señal medioambiental clave en la regulación de los genes que controlan este proceso. Dos mecanismos esenciales contribuyen a la reducción de la concentración de oxígeno en la zona de fijación de nitrógeno; una barrera a la difusión del gas a lo largo del nódulo y la expresión de la leghemoglobina, proteína con una alta afinidad por el oxígeno, que regula la cesión de éste a los simbiosomas.

La histidín kinasa FixL percibe la baja concentración de oxígeno en la membrana del bacteroide y transduce esta señal mediante FixJ cuya forma fosforilada activa la transcripción de los reguladores NifA y FixK de los que depende la expresión de los genes que codifican las proteínas estructurales de la nitrogenasa, las que se requieren para su adecuado ensamblaje y las que integran la cadena específica que transporta los electrones que utiliza la nitrogenasa para la reducción del nitrógeno molecular. *Rhizobium* codifica una Ferro-Molibdo nitrogenasa formada por dos componentes, el homodímero nitrogenasa reductasa y el heterotetrámero nitrogenasa. Este complejo utiliza un total de 8 electrones y 16 moléculas de ATP para la reducción de una molécula de nitrógeno que origina 2 moléculas de amonio y una de hidrógeno, que en algunos rhizobia como *Bradyrhizobium japonicum* o *Rhizobium leguminosarum* *bv. viciae* es reciclado por los sistemas Hup o Hidrogen uptake para la mejora de la eficiencia global del proceso (Fischer, 1994; Schwarz *et al.*, 2009).

## 6.2. SINORHIZOBIUM MELILOTI

*Sinorhizobium meliloti* es una bacteria Gram-negativa, de un tamaño de 0,5 - 1,0 µm de ancho por 1,2 - 3,0 µm de largo, perteneciente a la familia Rhizobiaceae dentro del orden Rhizobiales de las α-proteobacterias. Se trata de microorganismos aerobios que generalmente presentan gránulos de poli-β-hidroxiburato y no forman esporas. Son móviles gracias a la presencia de un flagelo polar o subpolar o de uno a seis flagelos peritricos. Forman colonias circulares, convexas, semitranslúcidas y mucilaginosas, normalmente con un diámetro de 2 - 4 mm después de 3 - 5 días de cultivo en medio completo YMA (Jordan, 1984). Está presente tanto en el suelo como en la rizosfera induciendo la formación de nódulos indeterminados en raíces de un

conjunto de especies de *Medicago*, incluyendo *Medicago sativa* (alfalfa), dentro de los cuales fija nitrógeno atmosférico. La simbiosis *S. meliloti*-*M. truncatula* es uno de los sistemas modelo de referencia para el estudio de estas interacciones procariota-eucariota.

Estudios filogenéticos recientes muestran que las especies del género *Sinorhizobium* están muy relacionadas con *Ensifer adhaerens* (Casida, 1982), por lo que el género *Sinorhizobium* ha sido reclasificado como *Ensifer* (Young, 2003), debido a que las reglas de la nomenclatura establecen que un género descrito con anterioridad tiene prioridad sobre los nombres posteriores del mismo. No obstante, la controversia aún persiste y, aunque el Comité Judicial de Taxonomía ha resuelto que *E. adhaerens* no puede pasar a llamarse *S. adhaerens* como habían propuesto algunos autores (Willems *et al.*, 2003), el subcomité de taxonomía de *rhizobia* ha propuesto que sigan utilizándose ambos nombres. Por ello, en este trabajo hemos preferido utilizar el nombre de *Sinorhizobium* por ser el más conocido y aceptado entre los rizobiólogos.

### 6.2.1. El genoma de *S. meliloti* 1021

Como microorganismo modelo, la secuencia genómica de la cepa de referencia *S. meliloti* 1021 fue publicada en el año 2001 (Galibert *et al.*, 2001). Esta bacteria posee un genoma compuesto por tres replicones; el cromosoma (3.65 Mb) (Capela *et al.*, 2001) y dos megaplásmidos denominados pSymA (1.35 Mb) (Barnett *et al.*, 2001) y pSymB (1.68 Mb) (Finan *et al.*, 2001) en los que se localizan la mayoría de los genes necesarios para el establecimiento de la interacción simbiótica. El primer análisis de la secuencia de este genoma resultó en la anotación de 6.206 ORFs, 54 tRNAs, 3 operones de rRNAs, una copia de tmRNA y múltiples secuencias de inserción (ISs) y elementos genéticos repetitivos (REP). Entre las posibles proteínas codificadas por este genoma, el 59,7% presenta homología funcional mientras que para el 40,3% restante no se puede predecir función en las bases de datos y de éstas el 8,2% están codificadas por genes sin homología y, por tanto, únicos en esta bacteria. Llama la atención que casi un 12% de los genes anotados codifica factores de transcripción y proteínas de transporte, lo que dota a esta bacteria de la versatilidad de adaptación que demanda su complejo ciclo vital.

Las características más relevantes de este genoma aparecen recogidas en la Tabla 2.

Tabla 2. Características del genoma de *S. meliloti* 1021

Característica	Cromosoma	pSymA	pSymB	Genoma
Tamaño (pb)	3.654.135	1.354.226	1.683.333	6.691.694
Proporción G+C	62,7%	60,4%	62,4%	62,1%
Regiones codificantes (ORFs)	85,8%	83,2%	88,6%	85,9%
tRNAs	51	2	1	54
tmRNA	1	0	0	1
Operones ribosomales	3	0	0	3
Genes codificantes de proteínas	3.341	1.293	1.570	6.204
Tamaño medio de genes codificantes de proteínas (pb)	938	871	950	927
Genes con asignación funcional	59%	56,5%	64,4%	59,7%
Genes Orphan (porcentaje del total de genes codificantes)	5%	11,5%	12,3%	8,2%
Genes reguladores (porcentaje del total de genes codificantes)	7,2%	10,4%	10,5%	8,7%
Secuencias de inserción y fagos	2,2%	3,6%	0,9%	2,2%
Elementos RIME	185	6	27	218
Palíndromos A, B y C	253	0	5	258

### 6.2.2. Las cepas *S. meliloti* 1021 y 2011

*S. meliloti* 1021 y 2011 son cepas que provienen del mismo progenitor (*S. meliloti* SU47) y han sido clásicamente consideradas isogénicas aunque el genoma de la segunda no ha sido secuenciado, utilizándose indistintamente como cepa modelo de referencia en los estudios de la biología de esta bacteria (Casse *et al.*, 1979).

Sin embargo, la realidad es que aunque no haya indicios de divergencias genómicas notables hay diferencias fenotípicas evidentes entre ambas. Por ejemplo, datos empíricos (no publicados) revelan que en cuanto a crecimiento y comportamiento simbiótico, *S. meliloti* 1021 y 2011 son diferentes, ya que la segunda crece y nodula más rápido que la primera.

El estudio de la variación de los niveles de calcio en los pelos radiculares (*calcium spiking*) ante la infección también revela ciertas divergencias entre ambas cepas (Wais *et al.*, 2002). Concretamente 2011 no es capaz de inducir este fenómeno durante la interacción simbiótica con *M. truncatula*, al contrario que 1021, que sí produce un *calcium spiking* evidente (Wais *et al.*, 2002).

La respuesta de ambas cepas frente una situación de carencia de fosfatos en el medio extracelular también difiere (Krol and Becker, 2004). En *S. meliloti* la adaptación al estrés producida por fosfato se regula por el sistema *phoB*. La comparación del transcriptoma de ambas cepas en estas condiciones reveló que 1021 y 2011 ven afectados genes dependientes de *phoB* diferentes, sugiriendo una



regulación desigual de la respuesta a este estrés. Esta evidencia se ve apoyada por una mutación en el gen *pstC* (transportador de membrana implicado en la asimilación del fosfato) en 1021, que posee una versión truncada de este gen (Krol and Becker, 2004).

En cualquier caso, en este trabajo de investigación se han seguido considerando los genomas de ambas cepas como equivalentes, sin aparentes inconsistencias en los resultados experimentales más relevantes obtenidos.

# **OBJETIVOS**



La transición de un estado de vida libre a otro de residencia intracelular crónica demanda de las bacterias una gran flexibilidad de adaptación a cambios continuos en su medio externo que implica la expresión coordinada de complejas redes de genes. En las relaciones huésped-parásito los RNAs no codificantes (sRNAs) expresados por el patógeno son elementos de regulación esenciales para la integración de señales medioambientales diversas y la coordinación de la respuesta fisiológica a las mismas asociada a la virulencia. A pesar de esta evidencia, la función de la ribo-regulación en el establecimiento de otras interacciones procariota-eucariota superior, como las endosimbiosis mutualistas entre los rizobios y las plantas leguminosas, ha permanecido inexplorada.

En el momento del comienzo de este trabajo de investigación sólo los tRNAs (54), los rRNAs (3 operones), algunas ribozimas del grupo II y el tmRNA (gen cromosómico *ssrA*) aparecían anotados como componentes del RNoma de la cepa de referencia del endosimbionte diazotrófico *S. meliloti*. Con estos antecedentes los objetivos planteados en esta Tesis Doctoral fueron los siguientes:

**1.** Identificación, caracterización molecular y catalogación de los componentes del RNoma de *S. meliloti*.

**2.** Aproximación genética y genómica a la función de la proteína Hfq como primera valoración del impacto global de la ribo-regulación en la biología de esta bacteria.

**3.** Desarrollo de herramientas genéticas para la investigación de la *trans*-ribo-regulación en este modelo experimental. Aplicación de éstas a la caracterización funcional de los sRNAs Smr7C, Smr15C1 y Smr15C2.

Esta Memoria de Tesis Doctoral se ha estructurado en tres capítulos que responden a los objetivos propuestos, además de los Materiales y Métodos y una discusión generales, que se presentan a continuación y al final de la Memoria, respectivamente. A su vez, cada capítulo consta de una pequeña introducción, de la descripción del diseño experimental relevante para la obtención de los resultados, y de la presentación y descripción de éstos. El Capítulo 1 es esencialmente metodológico y describe la predicción bioinformática, la caracterización molecular y la identificación experimental mediante secuenciación masiva como abordaje para generar un extenso catálogo de RNAs no codificantes en *S. meliloti*. En el segundo

Capítulo se presenta una aproximación genética a la función de la chaperona de RNA Hfq, para la identificación de procesos y genes potencialmente ribo-regulados. Finalmente, en el tercer Capítulo se aborda la caracterización funcional de tres de los ribo-reguladores identificados en el primer Capítulo, y se profundiza en los criterios para su caracterización funcional y determinación del mRNA diana.

# **MATERIALES Y MÉTODOS**



## M-1. ESPECIES Y CEPAS BACTERIANAS

En la tabla M-1 se muestran las principales especies y cepas bacterianas utilizadas en este trabajo:

Tabla M-1. Especies y cepas bacterianas utilizadas en este trabajo.

CEPA BACTERIANA	CARACTERÍSTICAS	REFERENCIA
<b>Cepas de <i>Sinorhizobium meliloti</i></b>		
<i>Sinorhizobium meliloti</i> 1021	Mutante espontáneo derivado de la cepa salvaje SU47, Nod+, Fix+. Sm <sup>r</sup>	Meade <i>et al.</i> , 1982
<i>S. meliloti</i> 2011	Derivado de SU47, Nod+, Fix+, Sm <sup>r</sup> ;	Casse <i>et al.</i> , 1979
<i>S. meliloti</i> B23001	Cepa derivada de 2011, en la que se ha sustituido el gen <i>expR</i> alterado por una copia funcional del mismo. Nod+, Fix+, <i>expR</i> +, Sm <sup>r</sup>	Laboratorio de A.Becker (sin publicar)
<i>S. meliloti</i> B23001 <i>hfq</i> <sup>3XFLAG</sup>	Derivada de B23001 que expresa Hfq <sup>FLAG</sup> . Sm <sup>r</sup>	Este trabajo
<i>S. meliloti</i> 1021Δ <i>hfq</i>	Mutante por delección del gen <i>hfq</i> . Sm <sup>r</sup>	Este trabajo
<i>S. meliloti</i> 1021 <i>hfq</i> <sup>3XFLAG</sup>	Derivada de 1021 que expresa Hfq <sup>FLAG</sup> . Sm <sup>r</sup>	Este trabajo
<i>S. meliloti</i> 2011 mTn5STM.4.07.G10	Mutante Tn5 del gen <i>rne</i> (ribonucleasa E)	Pobigaylo <i>et al.</i> , 2006
<i>S. meliloti</i> 1021Δ <i>smr</i> 7C	Mutante por delección de <i>smr</i> 7C, sustituido por un marcador de resistencia a Eritromicina. Er <sup>r</sup>	Este trabajo
<i>S. meliloti</i> 1021Δ <i>smr</i> 15C1	Mutante por delección de <i>smr</i> 15C1, sustituido por un marcador de resistencia a Eritromicina. Er <sup>r</sup>	Este trabajo
<i>S. meliloti</i> 1021Δ <i>smr</i> 15C2	Mutante por delección de <i>smr</i> 15C2, sustituido por un marcador de resistencia a Eritromicina. Er <sup>r</sup>	Este trabajo
<i>S. meliloti</i> 1021 Δ <i>smr</i> 15C1/C2	Doble mutante <i>smr</i> 15C1 y <i>smr</i> 15C2, en el que se han sustituido los dos <i>loci</i> por el marcador Er <sup>r</sup>	Este trabajo
<b>Cepas de <i>Escherichia coli</i></b>		
<i>Escherichia coli</i> DH5α	F-, ø80 <i>dlacZ</i> ΔM15, Δ( <i>lacZYA-argF</i> )U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (rK-, mK+), <i>phoA</i> , <i>supE44</i> , λ-, <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Bethesda Research Lab.
<i>E. coli</i> HB101	<i>supE44</i> , Δ( <i>mcrC-mrr</i> ), <i>recA13</i> , <i>ara-14</i> , <i>proA2</i> , <i>lacY1</i> , <i>galK2</i> , <i>rpsL20</i> , <i>xyl-5</i> , <i>mtl-1</i> , <i>leuB6</i> , <i>thi-1</i>	Promega Corporation



## M-2. PLÁSMIDOS Y VECTORES DE CLONAJE

En la tabla M-2 se muestran las construcciones plasmídicas finales más relevantes utilizadas en este trabajo. Las correspondientes a etapas intermedias de la generación de las mismas se adjunta en los anexos correspondientes a cada capítulo de esta tesis doctoral.

Tabla M-2. Principales plásmidos utilizados en este trabajo.

PLÁSMIDO	CARACTERÍSTICAS	REFERENCIA
pGEM-T Easy	Vector de clonaje para productos de PCR. Ap <sup>r</sup>	Promega
pBlueScript-KS II	Plásmido multicopia para transcripción <i>in vitro</i> . Ap <sup>r</sup>	Stratagen
pKS <sub>smr7C</sub>	pBlueScript para transcripción <i>in vitro</i> de <i>smr7C</i> . Ap <sup>r</sup>	Este trabajo
pKS <sub>smr15C1</sub>	pBlueScript para transcripción <i>in vitro</i> de <i>smr15C1</i> . Ap <sup>r</sup>	Este trabajo
pKS <sub>smr15C2</sub>	pBlueScript para transcripción <i>in vitro</i> de <i>smr15C2</i> . Ap <sup>r</sup>	Este trabajo
pBBR1MCS-2	Derivado de pBBR1MCS que contiene la resistencia a Km <sup>r</sup>	Kovach <i>et al.</i> , 1994
pBBSYN7C	pBBR1MCS-2 que expresa <i>smrC7</i> bajo el promotor constitutivo <i>syn</i> .	Este trabajo
pBBSYN15C1	pBBR1MCS-2 que expresa <i>smrC15</i> bajo el promotor constitutivo <i>syn</i> .	Este trabajo
pBBSYN15C2	pBBR1MCS-2 que expresa <i>smrC16</i> bajo el promotor constitutivo <i>syn</i> .	Este trabajo
pSRKKm	Derivado de pBBR1MCS-2, conteniendo el promotor <i>lac</i> y <i>lacI<sup>q</sup></i> , <i>lacZa<sup>+</sup></i> y Km <sup>r</sup>	Khan <i>et al.</i> , 2008
pSRK_C	Derivado del pSRKKm en el que se ha eliminado el operador <i>lacI<sup>q</sup></i> . Km <sup>r</sup>	Este trabajo
pSRK7C	pSRK_C que expresa <i>smrC7</i> bajo el promotor <i>lac</i> . Km <sup>r</sup>	Este trabajo
pSRK15C1	pSRK_C que expresa <i>smrC15</i> bajo el promotor <i>lac</i> . Km <sup>r</sup>	Este trabajo
pSRK15C2	pSRK_C que expresa <i>smrC16</i> bajo el promotor <i>lac</i> . Km <sup>r</sup>	Este trabajo
pK18 <i>mobsacB</i>	Vector suicida en <i>S. meliloti sacB</i> . oriV, Km <sup>r</sup>	Schafer <i>et al.</i> , 1994
pK18Δ <i>smr7C</i>	pk18 para la delección de <i>smr7C</i> y sustitución por el marcador de resistencia a Eritromicina. Er <sup>r</sup> , Km <sup>r</sup>	Este trabajo
pK18Δ <i>smr15C</i>	pk18 para la delección de <i>smr15C1</i> y sustitución por el marcador de resistencia a Eritromicina. Er <sup>r</sup> , Km <sup>r</sup>	Este trabajo

pK18 $\Delta$ smr15C2	pk18 para la delección de <i>smr15C2</i> y sustitución por el marcador de resistencia a Eritromicina. Er <sup>r</sup> , Km <sup>r</sup>	Este trabajo
pK18 $\Delta$ hfq	pk18 para la delección de <i>hfq</i> sin marcador de resistencia. Km <sup>r</sup>	Este trabajo
pK18hfq <sup>3XFLAG</sup>	pk18 que contiene la región genómica <i>hfq</i> con el mismo gen modificado con la copia Hfq <sup>3XFLAG</sup> . Km <sup>r</sup>	Este trabajo
pJB3Tc19	Plásmido de bajo número de copias, IncP, Ap <sup>r</sup> , Tc <sup>r</sup>	Blatny <i>et al</i> , 1997
pJBHfq	Plásmido que expresa el ORF <i>hfq</i> desde su propio promotor Ap <sup>r</sup> , Tc <sup>r</sup>	Este trabajo
pBBSYNeGFP	Vector derivado de pBBR1MCS-2 que expresa la proteína verde fluorescente desde el promotor syn. Km <sup>r</sup>	Este trabajo
pJBHfq:GFP	Plásmido que expresa la fusión traduccional Hfq:GFP desde su propio promotor Ap <sup>r</sup> , Tc <sup>r</sup>	Este trabajo

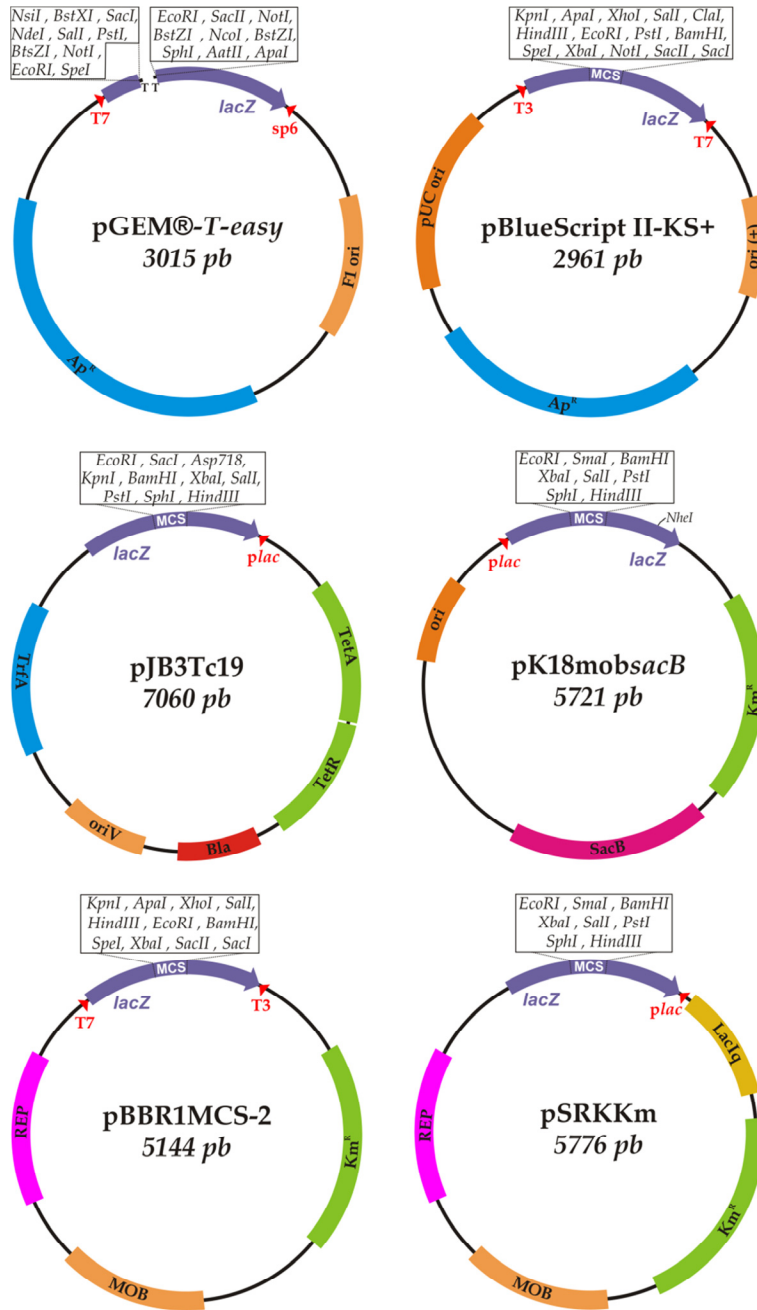
Los diagramas correspondientes a los vectores de destino más relevantes se muestran en la Fig. M-1.

### M-3. OLIGONUCLEÓTIDOS EMPLEADOS EN ESTE ESTUDIO

Los oligonucleótidos que se han utilizado como cebadores para PCR o como sondas para la hibridación Northern se detallan específicamente en los anexos correspondientes a cada capítulo. Los cebadores universales empleados sí se detallan en esta sección en la tabla M-3:

Tabla M-3. Oligonucleótidos universales utilizados en este trabajo.

OLIGONUCLEÓTIDO	SECUENCIA
T7	5'-TAATACGACTCACTATAGGGCGA -3'
SP6	5'- GTATTCTATAGTGTACCTAAATAGC -3'
T3	5'- AATTAACCCTCACTAAAGGGAGA -3'
M13-Forward	5'- TGTAACAACGACGGCCAGTG -3'
M13-Reverse	5'- TCATGGTCATAGCTGTTCC -3'



**Fig. M-1.** Principales plásmidos de destino utilizados en este trabajo. Las flechas rojas indican el sitio de anillamiento de cebadores universales.

## M-4. TÉCNICAS MICROBIOLÓGICAS

### M-4.1. MEDIOS DE CULTIVO Y ANTIBIÓTICOS

#### M-4.1.1. MEDIOS DE CULTIVO

El medio completo utilizado para el crecimiento rutinario de *S. meliloti* fue **TY** (Beringer, 1974) preparado en agua desionizada:

CaCl <sub>2</sub> × 2H <sub>2</sub> O	0,9 g/l
Triptona (DIFCO)	5 g/l
Extracto de levadura (DIFCO)	3 g/l

Para la preparación del medio sólido se añadió agar (PANREAC) al 1,6%. La esterilización se realizó en autoclave durante 20 min a 120°C.

En protocolos en los que resulta necesario reducir el crecimiento de *E. coli* sin alterar el de *S. meliloti* se utilizó la siguiente modificación del **medio mínimo (MM)** descrito por Robertsen *et al.*, 1981 preparado en agua desionizada:

K <sub>2</sub> HPO <sub>4</sub>	0,3 g/l
KH <sub>2</sub> PO <sub>4</sub>	0,3 g/l
MgSO <sub>4</sub> .7H <sub>2</sub> O	0,15 g/l
CaCl <sub>2</sub>	0,05 g/l
FeCl <sub>3</sub>	0,006 g/l
NaCl	0,05 g/l
Glutamato sódico	1,1 g/l
Manitol	10 g/l
Biotina	0,0002 g/l
Pantotenato cálcico	0,0001 g/l

Para solidificar el medio se añadió agar purificado (OXOID) al 1,6%. El pH se ajustó a 6,8 - 7,0 y se esterilizó en autoclave durante 20 min a 120°C.

Como medio de cultivo rutinario de las cepas de *Escherichia coli* se utilizó el medio de **Luria-Bertani (LB)** (Sambrook *et al.*, 1989) preparado con agua desionizada:

NaCl	5 g/l
Triptona (DIFCO)	10 g/l
Extracto de levadura	5 g/l

Para solidificar el medio se añadió agar (PANREAC) al 1,6%. El medio se esterilizó en autoclave a 120°C durante 20 min.

Como medio de cultivo restrictivo para coliformes se utilizó el **Endo Agar™ LES** (DIFCO). Se utilizó para la verificación de posibles contaminaciones por *E. coli* tras las conjugaciones. Se prepara disolviendo 51 g por litro de agua desionizada, calentando hasta disolución completa y luego esterilizado finalmente en autoclave a 120°C durante 20 min.

#### M-4.1.2 ANTIBIÓTICOS

La adición de antibióticos a los medios de cultivo se hizo a partir de soluciones 100 veces concentradas preparadas en agua desionizada (Km, Ap) o en agua-EtOH al 50% (Tc). Las soluciones preparadas en agua se esterilizaron por filtración utilizando unidades Minisart® NML (Sartorius) de 0,2 µm de tamaño de poro. Las concentraciones empleadas para los distintos antibióticos se muestran en la siguiente tabla

Tabla M-4. Antibióticos utilizados en este trabajo y su concentración empleada en *E. coli* y *S. meliloti*.

ANTIBIÓTICO	EFEECTO	<i>E.coli</i>	<i>S. meliloti</i>
Estreptomocina, Sm (Roche)	Bactericida	200 mg/l	200 mg/l
Ampicilina, Ap (Sigma)	Bacterioestático	200 mg/l	200 mg/l
Kanamicina, Km (Sigma)	Bactericida	50 mg/l	180 mg/l
Tetraciclina, Tc (Sigma)	Bactericida	10 mg/l	10 mg/l
Rifampicina, Rf (Sigma)	Bactericida	50 mg/l	50 mg/l
Eritromicina, Er (Sigma)	Bacterioestático	60 mg/l	100 mg/l
Ácido nalidixico (Sigma)	Bactericida	-	10 mg/l

#### **M-4.2. CONDICIONES DE CULTIVO**

Las distintas cepas de *S. meliloti* se crecieron a 28°C, siendo su tiempo de generación en medio líquido de aproximadamente 2,5 h a 190 r.p.m. en agitador orbital. *E. coli* se incubó a 37°C y su tiempo de generación en medio líquido es aproximadamente 30 min a 190 r.p.m. en agitador orbital.

#### **M-4.3. CONSERVACIÓN DE CULTIVOS BACTERIANOS**

Para la conservación prolongada de los cultivos bacterianos se utilizó la congelación. Para el mantenimiento de la viabilidad celular durante el periodo de conservación se utilizó glicerol 20% (v/v) como crioprotector. El glicerol estéril

dispuesto en criotubos se mezcló con cultivos crecidos hasta fase logarítmica tardía. Los viales se congelaron rápidamente y se almacenaron a una temperatura de  $-80^{\circ}\text{C}$ .

#### **M-4.4. MÉTODOS DE MOVILIZACIÓN DE DNA EXÓGENO EN BACTERIAS**

##### M-4.4.1. TRANSFORMACIÓN EN *E. COLI*

###### *M-4.4.1.1. Preparación de células competentes*

Para introducir el DNA circularizado se utilizan un tipo de células que están preparadas para captar DNA exógeno, las células competentes. Su preparación requiere de métodos químicos o físicos que alteran la pared celular de la bacteria.

###### M-4.4.1.1.1. Métodos químicos: RbCl

Se cultivaron las células de *E. coli* en 100 ml de medio LB hasta que se alcanzó una  $A_{600}$  de 0,4 (fase logarítmica). Entonces se detuvo el crecimiento incubando el cultivo en hielo durante 15 min. Transcurrido este tiempo, las bacterias se sedimentaron por centrifugación (6000 r.p.m., 10 min,  $4^{\circ}\text{C}$ ). Con cuidado, se resuspende el sedimento en 32 ml de solución RF1 estéril y preenfriada a  $4^{\circ}\text{C}$  [RF1 (por 100 ml): 1,2 g de RbCl; 0,99 g de  $\text{MnCl}_2 \times 4 \text{H}_2\text{O}$ ; 0,294 g de acetato potásico; 0,15 g de  $\text{CaCl}_2 \times \text{H}_2\text{O}$ ; 11,9 ml de glicerol; agua destilada hasta 100 ml, pH 5,8]. Se incubaron en hielo durante 15 min y después se centrifugaron a 6.000 r.p.m. durante 10 min. Se retiró el sobrenadante y las células se resuspendieron en 4 ml de solución RF2 preenfriada a  $4^{\circ}\text{C}$  [RF2 (por 50 ml): 0,1046 g de ácido morfolino propanosulfónico (MOPS); 0,06 g de RbCl; 0,55 g de  $\text{CaCl}_2 \times \text{H}_2\text{O}$ ; 5,95 ml de glicerol; agua destilada hasta 50 ml, pH 6,8]. Se repartieron en alícuotas de 100  $\mu\text{l}$  en tubos preenfriados a  $4^{\circ}\text{C}$  y se congelaron inmediatamente con nitrógeno líquido. Se conservaron por un periodo limitado a  $-80^{\circ}\text{C}$ . El rango de eficiencia de transformación se encuentra entre  $1-5 \times 10^5$  células/ $\mu\text{g}$  DNA.

###### M-4.4.1.1.2. Células electrocompetentes

Para este protocolo se precisa la eliminación completa de las sales del medio de cultivo de manera que no interfieran con la electroporación. A su vez, se añade un agente crioprotector (glicerol) que va a permitir la viabilidad de las células conservadas a  $-80^{\circ}\text{C}$ . Consiste en sucesivos lavados con volúmenes cada vez más reducidos de glicerol 10%; partimos de 500 ml de cultivo a una  $A_{600}$  de 0,5 cuyo metabolismo se retrasa por incubación en hielo durante 20 min. Se recogen a 6.000 r.p.m. durante 15 min y  $4^{\circ}\text{C}$  y se resuspenden en 100 ml de glicerol 10% frío,

volviéndose a centrifugar. En el siguiente paso se resuspenden en 20 ml de glicerol para acabar, tras este segundo lavado, con las bacterias resuspendidas en 2 ml de glicerol 10%. Se reparten en alícuotas de 50  $\mu$ l que inmediatamente se congelan en N<sub>2</sub> líquido. Se conservan a -80°C durante un periodo limitado. El rango de eficiencia de transformación es aproximadamente de 10<sup>6</sup> células/ $\mu$ g DNA.

#### ***M-4.4.1.2. Transformación de células competentes***

##### **M-4.4.1.2.1. Transformación por choque térmico**

Las células competentes preparadas por métodos químicos se sacan a hielo hasta su descongelación total (~20 min). Se añade en condiciones de esterilidad el DNA, procedente de ligación o no, y se deja incubando con las células durante otros 20-25 min. Las células se someten a un choque térmico de 42°C durante 90 s, para seguidamente devolverlas a hielo durante otros 5 min en los que se permitirá la captación del DNA. Posteriormente se añaden 900  $\mu$ l de LB fresco durante 1 h a 37°C (en agitación a bajas revoluciones) para la recuperación de las células. Transcurrido este tiempo, se siembra parte o todo el cultivo en medio LB sólido, dependiendo de la eficiencia de transformación que esperemos y la cantidad de DNA utilizada y se seleccionan los transformantes en medio LB con sus correspondientes antibióticos.

##### **M-4.4.1.2.1. Electrotransformación**

En este protocolo también se debieron descongelar las células en hielo tras lo cual se añadió el DNA con el que se quiere transformar. Se transfirió la suspensión celular a una cubeta de electroporación y se sometió a un pulso eléctrico según las recomendaciones hechas por el fabricante. En nuestro caso se utilizó un electroporador *Eppendorf 2510* que recomienda un pulso de 1800 V durante 3-5 milisegundos. Como en el caso anterior, se restauró la estabilidad celular por la adición de 950  $\mu$ l de LB fresco. A continuación se procede como en el apartado anterior. Es vital que el DNA con el que queremos transformar esté libre de sales; para eliminar las sales presentes tras una reacción de digestión o ligación se dializan las muestras frente a agua utilizando filtros de nitrocelulosa VSWP de 0,025 mm (MILLIPORE).

#### **M-4.4.2. CONJUGACIONES EN *S. MELILOTI***

La movilización de plásmidos entre la cepa receptora de *S. meliloti* y la cepa donadora de *E. coli* se realizó por conjugación triparental (Ditta *et al.*, 1980). En este tipo de conjugación, se requieren los genes de movilización presentes la cepa *helper* *E. coli* HB101:pRK2013.

Las conjugaciones se llevaron a cabo por mezcla de cantidades equivalentes de las tres bacterias durante 12-16 h en medio sólido TY incubadas a 28°C. Los transconjugantes se seleccionaron en placas de MM suplementadas con los correspondientes antibióticos. Se incubaron a 28°C hasta la aparición de colonias (2-3 días) y se volvió a realizar una nueva selección en MM pero ya de colonias aisladas, comprobando en paralelo la posible contaminación con coliformes mediante el uso de Endo-Agar sólido. Además, en los casos que fue necesario, se hicieron pases sucesivos en TY suplementado con ácido nalidíxico combinado con estreptomicina (600 mg/l) para eliminar toda presencia de *E. coli*.

Para los casos en los que se debían introducir dos plásmidos compatibles, se movilizó el primero siguiendo el proceso normal, y seguidamente se procedió a hacer lo propio con el segundo, haciendo la selección final en MM suplementado con los antibióticos correspondientes a la resistencia de ambos plásmidos.

#### M-4.4.3. SELECCIÓN DE INTERCAMBIO ALÉLICO Y CO-INTEGRACIÓN

La introducción de modificaciones genómicas se llevó a cabo por doble recombinación según el método descrito para bacterias Gram negativas (Schafer *et al.*, 1994). Para ello se llevó a cabo la modificación/delección un fragmento interno del gen que se deseaba mutar, previamente clonado en el plásmido pGEM®-T-Easy, con los sitios de restricción adecuados. Todos los pasos del diseño de estas construcciones se detallará en la sección de técnicas de biología molecular en este bloque de materiales y métodos.

Posteriormente, el fragmento con el gen mutado se subclonó en el plásmido suicida pK18*mobSacB*, con el que se transformó la cepa a mutar por conjugación triparental. Los transformantes que por una recombinación simple (co-integración), habían integrado el plásmido en el cromosoma se seleccionaron mediante crecimiento en placas de TY suplementado con Km y sacarosa al 10% (Panreac).

Para favorecer el evento de la segunda recombinación, las colonias que presentaron sensibilidad a sacarosa se cultivaron en medio TY líquido a 28°C y con agitación, hasta alcanzar una densidad óptica (D.O.) de 1. De este cultivo se sembraron diluciones en placas de TY con 10%de sacarosa, hasta obtener colonias separadas. Las colonias crecidas se replicaron en TY y TYKm, para finalmente seleccionar las sensibles a kanamicina. La pérdida de la copia original del gen en las colonias seleccionadas fue posteriormente confirmada por PCR y Southern blot.

En el caso de la presencia de marcadores de resistencia en las cepas portadoras del alelo modificado, se añadió el correspondiente antibiótico (eritromicina en



nuestro caso) al medio selectivo de la segunda recombinación a una concentración de 100 mg/ml.

## **M-5. TÉCNICAS DE BIOLOGÍA MOLECULAR**

### **M-5.1. MANIPULACIÓN DE ÁCIDOS DESOXIRRIBONUCLEICOS (DNA)**

#### M-5.1.1. AISLAMIENTO DE DNA

##### ***M-5.1.1.1. Extracción de DNA total mediante kit comercial***

Para este propósito se utilizó el kit comercial *Real Pure Extrac. DNA Genómico* (REAL). Se recogió 1ml de cultivo de *Rhizobium* y se lavó con 0,2 ml de Sarcosil al 0,1% en TE. La lisis se llevó a cabo por la adición de 0,6 ml de "Lysis Solution" incubando 5 min a 80°C. Se dejó enfriar a temperatura ambiente y se añadió 3 µl de "Solución RNase". Se mezcla invirtiendo y se incuba una h a 37°C. Pasado este tiempo se adicionan 0,2 ml de "Protein Precipitation" y agitamos vigorosamente durante 20 seg. Se incuba en hielo 10 min. La muestra se centrifuga a máximas revoluciones durante 5 min y se recoge el sobrenadante en un tubo nuevo. El DNA se precipita añadiendo 0,6 ml de isopropanol frío invirtiendo hasta la mezcla total durante 10 min a -20°C. Se centrifuga 10 min y se lava el sedimento con 0,2 ml de EtOH 70%. Se resuspende con 50-100 µl de agua destilada.

##### ***M-5.1.1.2. Lisados de colonia***

Es una estrategia muy rápida para obtener DNA de una calidad suficiente para su amplificación por PCR. En *E. coli* consiste en repartir los 25 µl de la mezcla de reactivos de PCR y, en lugar de poner el DNA o el lisado, se pica directamente una colonia de la placa en esterilidad utilizando un palillo, agitando dentro de la mezcla y sometiendo ese lisado a un programa de PCR estándar.

En el caso de *S. meliloti* se pican las colonias seleccionadas en esterilidad y se resuspenden en 100 µl de sarcosil 0,1% en TE. Se sedimentan las células mediante centrifugación 2 min a 13000 r.p.m., se descarta el sobrenadante y se resuspende el pellet celular en el mismo volumen de agua bidestilada. Se hierve la mezcla durante 3 min y se vuelve a centrifugar. Se descarta el precipitado y el sobrenadante es apto para utilizarlo en reacciones de PCR.

### ***M-5.1.1.3. Extracción de DNA plasmídico mediante lisis alcalina***

El aislamiento de DNA plasmídico se realizó siguiendo el protocolo modificado por Sambrook y colaboradores (Birnboim and Doly, 1979; Ish-Horowicz and Burke, 1981; Sambrook *et al.*, 1989). Se recogen 1,5 ml de cultivo bacteriano crecido hasta fase exponencial y se centrifugan a 13000 r.p.m. durante 2 min. Se lavan las células con 500  $\mu$ l de sarcosil 0,1% en TE y se centrifugan de nuevo, retirando finalmente el sobrenadante. Se resuspenden las células en 100  $\mu$ l de lisozima 4 mg/ml disuelta en solución I (glucosa 50 mM, Tris-HCl 25 mM pH 8, EDTA 10 mM; conservada a 4°C). Se incuba 5 min a temperatura ambiente, tras lo cual se añaden 200  $\mu$ l de solución II (NaOH 0,2 M, 1% SDS). Se homogeniza la preparación agitando el tubo por inversión varias veces y se mantiene en hielo durante 5 min. Entonces se adicionan 150  $\mu$ l de solución III (acetato potásico 5 M pH 4,8) que ayudará a la neutralización de la solución y la precipitación de proteínas. Se mezcla invirtiendo el tubo y se incuba en hielo 5 min para minimizar el posible daño con DNAsas celulares. Se centrifuga a 13.000 r.p.m. durante 5 min para sedimentar los restos celulares y se recoge el sobrenadante en un tubo nuevo. Se añade un volumen (~400  $\mu$ l) de Fenol-Cloroformo-Isoamílico 25:24:1 ( $\Phi$ CIA) y se mezcla vigorosamente. Se centrifuga 5 min a 13.000 r.p.m. y se recoge la fase superior acuosa y se mezcla seguidamente con un volumen de CIA (elimina posibles restos de fenol en las preparaciones). Se vuelve a mezclar y se centrifuga durante 5 min. Se pasa la fase acuosa a un tubo nuevo y se precipita el DNA con 2,5 volúmenes de EtOH 100% frío. Se puede optar por precipitar 30 min a -20°C o 15 min a -70°C, pero lo más habitual para evitar la precipitación de DNA genómico son 5 min a temperatura ambiente. Se centrifuga el DNA precipitado y se lava el sedimento con 200  $\mu$ l de EtOH al 70%. Se deja secar y se resuspende en 25  $\mu$ l de agua desionizada en la que se diluye previamente una solución concentrada (100x) de RNasa A (10  $\mu$ g/ml de concentración final). La *solución RNasa 10 mg/ml* se prepara diluyendo, en las cantidades indicadas, la enzima (liofilizada, Sigma) en una solución con 10 mM Tris-HCl pH 7.5 y 15 mM NaCl. Se hierve durante 15' y se deja enfriar a RT. Se conserva a -20°C (Sambrook *et al.*, 1989).

### ***M-5.1.1.4. Minipreparación de DNA plasmídico por precipitación con sales de magnesio***

Esta técnica está basada en el método de Studier, 1991 que permite el aislamiento rápido de plásmidos de *E. coli*. Se recogen 1,5 ml de cultivo bacteriano por centrifugación y, tras retirar el sobrenadante, se resuspenden las células en 100  $\mu$ l de agua desionizada. La lisis tiene lugar por la adición de 100  $\mu$ l de una solución compuesta por NaOH 0,1 M, EDTA 10 mM y 2% SDS. Se homogeneiza por agitación y se hierve durante 2 min. El DNA lineal, normalmente cromosómico, se precipita

por la adición de 50 µl de MgCl<sub>2</sub> 1 M. Tras homogeneizar completamente (aparición de precipitado blanco), se incuba en hielo 2-5 min. Se centrifuga 1 min para sedimentar el precipitado en el fondo del tubo y añadimos 50 µl de acetato potásico 5 M para eliminar proteínas. El tubo invertido se mezcla con vórtex evitando que se levante el precipitado blanco del fondo y se incuba 5 min en hielo. Se centrifuga durante 5 min y se pasa el sobrenadante a un tubo nuevo donde se precipita el DNA plasmídico con 0,6 ml de EtOH 100% enfriado -20°C. Se deja 5 min a temperatura ambiente y se centrifuga 5 min. Se lava el sedimento con 200 µl de EtOH al 70% y dejamos secar a vacío. Se resuspende en el volumen adecuado dependiendo de la concentración deseada; lo más habitual para un plásmido de alto número de copias es resuspender en 25 µl de agua desionizada con RNasa a una concentración de 10 µg/ml.

#### M-5.1.2. EVALUACIÓN DE LA CANTIDAD Y CALIDAD DEL DNA EXTRAÍDO

El cálculo de la concentración y pureza de los ácidos nucleicos se estimó siguiendo el método espectrofotométrico descrito en Sambrook *et al.*, 1989, utilizando un espectrofotómetro NanoDrop® ND-100, que integra todos estos datos a partir de 1-2 µl de nuestra muestra.

La relación D.O. <sub>260nm</sub> / D.O. <sub>280nm</sub> se usó como estimador de la pureza de la preparación, considerándose valores por debajo de 1,8 como indicadores de contaminación por proteínas o fenol.

#### M-5.1.3. DIGESTIÓN DE DNA MEDIANTE ENDONUCLEASAS DE RESTRICCIÓN

La digestión completa del DNA se realizó siguiendo las indicaciones de temperatura y tampón óptimos recomendadas por los proveedores de las endonucleasas de restricción (Roche y New England Biolabs). Lo más común es utilizar 1-5 U de enzima/µg DNA a digerir. En el caso de digestiones simultáneas con 2 enzimas de restricción se utiliza un tampón compatible para ambas (instrucciones del proveedor) y, en caso de no ser posible, se realiza primero la digestión con la enzima que precise el tampón con menor fuerza iónica. Si aún así no fuese factible, el DNA se limpia de sales y de la enzima tras la primera digestión, y se procede con la segunda. Las sales se eliminaron mediante el uso de kits comerciales como *QIAquick PCR Purification Kit* (Qiagen), o por fenolización y precipitación del DNA.

#### M-5.1.4. REACCIÓN DE DEFOSFORILACIÓN

En ocasiones es necesaria la eliminación de los grupos fosfato presentes en los extremos 5' de vectores linearizados para evitar que vuelvan a ligarse. La defosforilación de 5 µg de plásmido digerido precisa su incubación con 1 U de fosfatasa alcalina (CIP, Calf Intestine Phosphatase; Roche) en su tampón correspondiente durante 1 h a 37°C. La reacción se para añadiendo 55 µl de una solución compuesta por tampón STE 1X [1 mM Tris-HCl pH8, 100 mM NaCl y 1 mM EDTA] y SDS 1% e incubando 15 min a 68°C. Para obtener mejor rendimiento en la ligación es conveniente fenolizar y precipitar.

#### M-5.1.5. PCR (POLIMERASE-CHAIN-REACTION)

Dependiendo de la finalidad se utilizaron distintas polimerasas: *Phussion*<sup>®</sup> DNA polimerase (NEB), *Accuprime* (Invitrogen), *Certhamp*<sup>®</sup> (Biotools) o *Pfu* polimerasa (Biotools), si era necesaria una baja tasa de errores en el producto final, mientras que para comprobaciones de rutina se usó una *Taq* polimerasa purificada en el laboratorio. La amplificación se realizó en un termociclador *Eppendorf Mastercycler*<sup>®</sup> y el programa empleado varió dependiendo del objetivo.

Las reacciones se llevaron a cabo generalmente en un volumen final de 25 µl conteniendo 0,1 a 100 ng del DNA molde, 25 pmol de cada uno de los cebadores específicos, 0,1 mM de desoxinucleótidos trifosfato (1 M si se trata de DNA total o de lisados celulares), el tampón especificado para cada aplicación y 2 unidades de DNA polimerasa.

La purificación de los productos de amplificación (eliminación de cebadores y sales) se realizó mediante columnas *MicroSpin*<sup>™</sup> S-300HR (GE Healthcare) o usando el *QIAquick PCR Purification Kit* (Qiagen).

#### M-5.1.6. CLONAJE DE FRAGMENTOS

En este apartado se incluyen distintos protocolos utilizados habitualmente para el clonaje de insertos de interés en vectores.

##### ***M-5.1.6.1. Aislamiento de fragmentos de DNA de gel de agarosa***

Para el aislamiento de fragmentos de restricción separados por electroforesis horizontal en gel de agarosa se utilizó el kit comercial *QIAquick Gel Extraction Kit* (Qiagen) y se siguió el protocolo recomendado por el proveedor. El paso previo para el uso de kit precisa cortar la banda de interés, para lo que nos servimos del trans-

iluminador UV una vez se ha teñido el gel con bromuro de etidio (aunque se procura minimizar la exposición del DNA a la luz UV).

#### ***M-5.1.6.2. Ligación de fragmentos de restricción***

La unión covalente de dos moléculas de DNA lineal se lleva a cabo mediante la adición de 1 U T4 DNA ligasa (Roche) en su tampón adecuado. La proporción molar más habitual vector:inserto suele ser 1:3. Las reacciones se llevaron a cabo en un volumen final de 10-20 µl a 14-16°C durante toda la noche.

#### ***M-5.1.6.3. Adenilación de amplificadores de PCR***

En ocasiones el producto que se precisa ligar no es un fragmento de restricción sino un producto de PCR. En el mercado se encuentran disponibles vectores que permiten la clonación de amplificadores con extremos A protuberantes que deja la *Taq* polimerasa, pero algunas polimerasas con actividad exonucleasa 3'-5' dejan extremos romos, por lo que es un procedimiento habitual la adición de adeninas en los extremos de estos fragmentos. Esta reacción está catalizada por la *Taq* polimerasa (5 U) en presencia de 0,2 mM de dATP (70°C durante 30 min). Este producto puede ser utilizado directamente en la reacción de ligación.

### M-5.1.7. SECUENCIACIÓN Y ANÁLISIS DE SECUENCIAS

Para la secuenciación Sanger clásica se recurrió al servicio de secuenciación de DNA/Genómica del Instituto de Parasitología y Biomedicina "López-Neyra" o al de la Estación Experimental del Zaidín. Estos servicios amplifican las muestras con un termociclador *PE-9600* (Perkin-Elmer) y las secuencian mediante el secuenciador *ABI 373 XL Stretch* (Perkin-Elmer) usando el kit *ABI PRISM Big Dye™ Terminator Cycle Sequencing Ready Reaction* (Perkin-Elmer).

La visualización de los picos de secuencia se realizó mediante el programa *Chromas Lite 2.01* y *Sequence Scanner v1.0*. El trabajo rutinario con las secuencias (búsqueda de dianas de restricción, diseño de cebadores, comparativa de secuencias, búsqueda de ORFs, etc.) se llevó a cabo con el programa *Clone Manager Professional Suite®*.

### M-5.1.8. ELECTROFORESIS DE DNA

Consiste en la separación de macromoléculas en función de su tamaño y carga. Los ácidos nucleicos, con carga negativa, migran hacia el polo positivo.

#### ***M-5.1.8.1. Electroforesis en geles de agarosa no desnaturalizantes***

El análisis de productos de PCR, de DNA total o plasmídico o de fragmentos de restricción se llevó a cabo mediante electroforesis horizontal en geles de agarosa (agarosa *SeaKem® LE*, Cambrex/Iberlabo) al 1% en TAE 1x: [TrisHCl 40 mM, EDTA Na<sub>2</sub> 2 mM, ácido acético glacial 0,1142%(v/v)]. El voltaje habitual de trabajo fue de 120 V aunque determinados procedimientos requirieron velocidades inferiores. Como tampón de carga se utilizó una solución 6x (1µl por cada 5 de muestra) de 0,25% (p/v) de azul de bromofenol y glicerol 45% (v/v).

El revelado de los geles de agarosa se realizó por inmersión de los mismos en una solución de bromuro de etidio 1 mg/ml en agua durante 15 a 20 min. Este intercalante permite visualizar DNA o RNA en un transiluminador que emite luz UV de 365 nm de longitud de onda. La toma de fotografías y procesamiento de imágenes de los geles se realizaron con *Gelprinter* y *Gelstation* de TDI.

#### ***M-5.1.8.2. Electroforesis en geles de poliacrilamida no desnaturalizantes***

Para la resolución de fragmentos inferiores a 200 pb se recurre a electroforesis vertical en geles de poliacrilamida. En este caso el tampón de electroforesis suele ser TBE 1X [Tris-HCl 0,089 M pH 8, ácido bórico 0,089 M y EDTA 0,002 M]. Se trabajó con disoluciones 40% (29:1) de acrilamida:bis-acrilamida (Bio-Rad). El porcentaje y el grosor del gel varía según las necesidades y con éstos también cambian las proporciones relativas del agente polimerizante (TEMED, Sigma) y el catalizador (persulfato amónico, APS; Sigma). El voltaje al que se realizó la electroforesis osciló entre los 150-200V hasta la salida del naranja G.

Para el revelado de estos geles se utilizó la tinción con nitrato de plata (Heuer *et al.*, 1997). Se fija con una solución compuesta por EtOH 10% (v/v) y ácido acético glacial 0,5% (v/v) (2 veces durante 3 min); a continuación se incubaba en agitación durante 30 min con una solución de nitrato de plata 0,2% en agua. Se lava 4 veces con agua para eliminar el exceso de plata y se revela con una mezcla compuesta por tetrahidrobórato sódico 10% (p/v)/formaldehído 0,4% (v/v)/NaOH 1,5% (p/v) controlando la aparición de las bandas. Se lava con agua destilada y el revelado se para por la incubación con carbonato sódico 0,75%. Previo a ser secado es conveniente mantener el gel toda la noche en una solución de EtOH 25% y glicerol 10%.

#### M-5.1.9. MARCADORES DE PESO MOLECULAR DE DNA

Los marcadores de peso molecular fueron en todos los casos suministrados por casas comerciales:

- **Marcador II:** DNA del fago  $\lambda$  digerido con la enzima *HindIII*. Está compuesto por 8 fragmentos que abarcan desde los cientos de nucleótidos hasta más de 20 Kb (Roche; Universidad Autónoma de Madrid).
- **Marcador III:** DNA del fago  $\lambda$  digerido con las enzimas *HindIII* y *EcoRI*. En este caso el patrón de digestión consta de 13 fragmentos que abarcan desde los cientos de nucleótidos hasta más de 20 Kb (Roche).
- **Marcador  $\Phi$ 29:** DNA del fago con el mismo nombre digerido con *HindIII* que libera 14 fragmentos: 72, 156, 273, 453, 579, 611, 759, 1150, 1331, 1933, 2201, 2498, 2899 y 4370 pb (Universidad Autónoma de Madrid).
- **Marcador pGEM:** DNA del plásmido pGEM-T digerido con las enzimas *Hinfl* y *EcoRI* que generan 15 fragmentos que se encuentran entre los 36 nucleótidos hasta las ~2500 pb (Promega).

#### M-5.1.10. HIBRIDACIÓN DNA-DNA (SOUTHERN BLOT)

Una vez separado en geles de agarosa el DNA total previamente digerido con las enzimas oportunas, se transfiere y se fija a una membrana de Biodine® cargada positivamente. Para ello se utilizó un sistema de transferencia alcalina (NaOH 1 M) por vacío (sistema VacuGene™ XL de Pharmacia). Previamente se rompe el DNA por la exposición prolongada (15-20 min) de los geles a los rayos UV en el transiluminador. Para evitar el depósito de sales, la membrana se embebe en agua desionizada para luego equilibrarla con 20XSSC [1XSSC: NaCl 150 mM, citrato sódico 15 mM, pH 7,0] durante 5 min. Se prepara el sistema de transferencia según las instrucciones de los proveedores con cuidado de no formar burbujas de aire entre la membrana y el gel. Se conecta el vacío y, una vez comprobada la estanqueidad del sistema, se cubre con NaOH 1M. Debemos vigilar que el sistema se mantenga constante en torno a 55 mbar durante 1-2 h. Posteriormente se lava la membrana con una solución de 2XSSC durante 10 min en suave agitación y se deja secar a temperatura ambiente. La fijación del DNA a la membrana se llevó a cabo por calor a vacío envuelta en papel Whatman 3MM (120°C 35 min, 70 cm Hg). Se puede conservar a temperatura ambiente.

##### ***M-5.1.10.1. Marcaje radiactivo de la sonda***

En este tipo de experimentos el marcaje se lleva a cabo con dCTP [ $\alpha$  -  $^{32}$ P] (PerkinElmer). La sonda se sintetizó mediante el kit comercial “MegaPrime™ DNA labeling system” de Amersham-GE healthcare™. Para ello se tomaron 50 ng de la

sonda y se anillaron 10 min a 65°C con cebadores aleatorios proporcionados por el kit. Al producto anillado se le incorporó el isótopo radioactivo mediante la adición de 1 U de Klenow y posterior incubación 1 h a 37°C. La sonda marcada se purificó con columnas S300 (GE-healthcare) y se midió la cantidad de marcaje (cuentas por minuto, c.p.m.) en un contador de centelleo Liquid Scintillation Analyzer Tri-Carb 1500, Δ Packard.

#### ***M-5.1.10.2. Hibridación, lavados y revelado***

Este proceso se llevó a cabo en un horno de hibridación Hybridiser® HB-1D (Techne), siguiendo el protocolo de Church and Gilbert, 1984. Se incubó el filtro en una solución de pre-hibridación (20 ml/250 cm<sup>2</sup>) [0,5 M Na<sub>2</sub>HPO<sub>4</sub> pH 7.2 y 0,34 % H<sub>3</sub>PO<sub>4</sub> v/v (*Church Buffer*), 1 mM EDTA, 0,5 % SDS p/v], durante 2 h a 68°C. La sonda se desnaturaliza durante 10 min a 95°C y posterior paso a hielo (5-10 min). En la solución de pre-hibridación se añade todo el volumen (50 ng) de la sonda desnaturalizada con la que se incubaba a 68°C durante toda la noche.

Al día siguiente se procedió con los lavados y el revelado de la hibridación. Se realiza un primer lavado rápido con 2XSSC/SDS al 0,1% (p/v), para eliminar la sonda no incorporada. Seguidamente se procede con lavados sucesivos de 15 min a 68°C con la misma solución; 1XSSC/SDS 0,1% (p/v) y 0,1XSSC/SDS 0,1% (p/v) para eliminar la sonda unida inespecíficamente a la membrana.

Para el revelado se envuelve la membrana en papel de plástico transparente y se expone en una pantalla (*Imagin Plate 2040, Fujifilm*), durante 4 h (exposición corta) y 24 h (exposición larga).

### **M-5.2. MANIPULACIÓN DEL RNA**

#### **M-5.2.1. EXTRACCIÓN DE RNA TOTAL DE *S. MELILOTI***

##### ***M-5.2.1.1. Extracción a partir de cultivos bacterianos***

Partimos de 25 ml de cultivo bacteriano, cuya actividad biológica se para por adición de 1/5 volúmenes de solución STOP (95% EtOH, 5% fenol, v/v). A continuación se centrifuga 10 min a 5000 r.p.m. a 4°C y se congelaron los pellets en nitrógeno líquido. En este punto los precipitados se pueden almacenar congelados a -80°C.

Se resuspenden las bacterias en 1 ml de solución de lisis (1.4% SDS, 4mM EDTA, 500 µg proteinasa K) precalentada a 65°C (3 ml para cultivos en fase estacionaria). Se



incubaban a 65°C durante 10 min, agitar esporádicamente en vórtex para lisar las bacterias y se pasan a hielo. Se añaden 500 µl (1.5 ml para lisados de 3 ml) de NaCl 5M a 4°C para favorecer la precipitación de proteínas. Se mantiene 10 min en hielo para proceder a una centrifugación de 10000 r.p.m. a 4°C durante 20 min. Se toma la fase acuosa en un nuevo tubo y se precipita con 4.5ml de EtOH 100% frío (7 ml para extracciones en fase estacionaria). Se deja al menos 1 h a -80°C. Se centrifuga a 10000 r.p.m. a 4°C durante 30 min y se retira totalmente el EtOH (EtOH) para proceder al tratamiento con DNasa I.

#### ***M-5.2.1.2. Extracción a partir de nódulos***

Se pulverizan 1-2 g de nódulos en mortero con nitrógeno líquido. Se dejan descongelar y se lisan directamente en el mortero con 4 ml de NTES (100mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1% SDS, 100mM β-mercaptoEtOH). Se añaden primero 2-3 ml y se termina de recoger con el resto de solución de lisis. Se pasa el lisado a un Falcon de 15 ml y se fenoliza con ΦCIA (pH 8 en Tris Buffer). Se mezcla y se da un ligero vórtex para centrifugar 15 min a 5000 r.p.m. y a 4°C. La fase acuosa se pasa un tubo nuevo y se precipita con 300 µl de AcNa 3M (pH 5.2) y 5 ml de isopropanol. Se mantiene al menos 1 h a -80°C y se centrifuga 20 min a 10000 r.p.m. para proceder al tratamiento con DNasa I.

#### ***M-5.2.1.3. Tratamiento con DNasa I***

Se resuspenden los precipitados obtenidos en las lisis tanto de bacterias como de nódulos en 440 µl de agua milliQ, se pasan a tubos de 1.5 ml y se añade por reacción 50 µl de tampón DNasa 10X (500 mM Tris-Cl ph 7.5, 10 mM MgCl<sub>2</sub>), 5 µl de RNase Inhibitor RNasin 40U/µl (Roche) y 5 µl de DNasa I (Roche). Se incuba 1 h a 37°C.

La reacción se fenoliza con 1 volumen de ΦCIA pH 4.3, se centrifugar 10 min a 13000 r.p.m. a 4°C, y la fase acuosa se pasa a un tubo nuevo. Se realiza una nueva extracción con 500 µl con cloroformo isoamílico (24:1) frío, se centrifuga 5 min a máxima velocidad y a 4°C y la fase acuosa se divide en dos tubos de 1.5 ml. Se precipita el RNA con 20 µl de AcNa 3M pH 5.2 y 1 ml de EtOH 100% frío. Se deja al menos 1 h a -80°C. Se centrifuga 30 min a máxima velocidad y a 4°C, se retira el EtOH y se lava el pellet con 500 µl de EtOH al 70%. Se centrifuga 10 min y se seca el pellet a temperatura ambiente en *speedvac*, para posteriormente resuspender en 50-100 µl de agua milliQ y cuantificar la cantidad de RNA que hemos obtenido. Para la estimación de la concentración de ácidos nucleicos en cada muestra se utilizó el espectrofotómetro *NanoDrop® ND-1000*.

#### **M-5.2.1.4. Extracción de RNA total de *S. meliloti* mediante kit comercial**

Para el conseguir marcar con éxito los RNAs utilizados para la hibridación de los *microarrays*, fue precisa la extracción de RNA total mediante el uso del kit comercial *RNeasy Protect Bacteria Mini Kit* (Qiagen).

*S. meliloti* es una bacteria bastante mucosa por lo que para mejorar la lisis se utilizan en este caso 100  $\mu$ l de una solución de lisozima 0,4 mg/ml en TE (por cada tubo correspondiente a 1,5 ml de cultivo) que ayuda a digerir la pared y mejorar la lisis celular. Se incuba 5 min a temperatura ambiente (RT) y se añaden 350  $\mu$ l de tampón RLT (al que añadimos en el momento  $\beta$ -mercaptoEtOH a la concentración indicada por la casa comercial), se agita vigorosamente y se centrifuga 2 min para eliminar los restos celulares. Se pasa el sobrenadante a un tubo nuevo con 250  $\mu$ l de EtOH 100%, mezclando con pipeta. Se procede a fijar el ácido nucleico sobre las columnas (por cada columna se pasa el sobrenadante correspondiente a 3 ml de cultivo) centrifugando 30 s a máximas revoluciones. Se lavaron las columnas con 700  $\mu$ l de tampón RW1. Se trataron dos veces sucesivas con 500  $\mu$ l de tampón RPE y se llevó a cabo una última centrifugación de 2 min para eliminar cualquier resto de EtOH de las columnas que reduce la eficiencia de elución. Finalmente se eluyeron los ácidos nucleicos en dos pasos, adicionando 30  $\mu$ l de agua sobre la resina e incubando un min a RT para posteriormente centrifugar 1 min.

Por lo general, la tecnología en la que se basan la columna minimiza la cantidad de DNA que es capaz de permanecer pegado a la resina. Sin embargo, este protocolo precisa de un RNA completamente libre de DNA (que puede interferir en la eficiencia de hibridación de *microarrays*). Para ello, se llevó a cabo la digestión del RNA resultante con DNAsa (Qiagen). A los 120  $\mu$ l de la solución anterior se añadieron 14  $\mu$ l de tampón RDD y 6  $\mu$ l de DNAsa. Se digirió durante una h a 28°C (la incubación podría alargarse media h más). A los 140  $\mu$ l de reacción se añadieron 490  $\mu$ l de tampón RLT (con  $\beta$ -mer) y 350  $\mu$ l de EtOH 100%. Se mezcla pipeteando, y se dispone sobre una columna y centrifugamos 30 seg. Se lava en dos ocasiones: primero con 350  $\mu$ l de RW1 y luego, con 500  $\mu$ l de RPE. Al igual que se describe anteriormente, se eliminó cualquier resto de soluciones de lavado mediante una centrifugación adicional de 2 min y se eluye en dos para obtener un volumen final de 50  $\mu$ l.

Para concentrar el RNA obtenido se utilizaron columnas Microcon 30 (Millipore) que, además de concentrar nuestra muestra, también purifican el RNA. Se llevó la muestra hasta un volumen final de 500  $\mu$ l con agua, se puso sobre la parte gris de la columna y se mezcló pipeteando. Se centrifugó hasta reducir el volumen a  $\sim$ 35  $\mu$ l

(generalmente, se utilizaron en torno a 8 min pero depende de la muestra) y se invirtió la columna sobre un tubo nuevo para centrifugar finalmente 2 min.

Aunque el RNA es estable a -20°C es aconsejable conservarlo a -80°C.

### M-5.2.2. ELECTROFORESIS DE RNA

#### ***M-5.2.2.1. Electroforesis en geles de agarosa desnaturalizantes***

Para visualizar la calidad y la cantidad de las muestras de RNA total obtenidas se llevó a cabo una separación en geles de agarosa desnaturalizantes al 1,3% (agarosa SeaKem® LE, Cambrex/Iberlabo) en MOPS 1X [4X: MOPS 80 mM, acetato sódico 20 mM y EDTA 4 mM; ajustar a pH 7 con NaOH] y Formaldehído al 1.875% (v/v). Es necesario tener la precaución de fundir la agarosa sin MOPS, pues éste es sensible a la luz y al calor. Una vez fundida, se pueden añadir el MOPS y el Formaldehído, en campana de extracción de gases, ya que este último es tóxico por inhalación. La electroforesis se corre en tampón MOPS 1X, siendo el voltaje habitual de trabajo de 100 V. Como tampón de carga se utilizó una solución 6X (1µl por cada 5 de muestra) de 0,25% (p/v) de azul de bromofenol y glicerol 45% (v/v). A éste tampón se le añadió el agente intercalante (bromuro de etidio o Bio-safe gel Red®) en una proporción del 1% v/v.

El revelado y la fotografía de estos geles se llevaron a cabo mediante un proceso similar al realizado para geles de agarosa para separar DNA.

### M-5.2.3. HIBRIDACIÓN DNA-RNA (NORTHERN-BLOT)

Para la detección de transcritos específicos en una muestra de RNA (total o procedente de una fracción co-inmunoprecipitada) de *S. meliloti* se realizó mediante hibridación Northern con una sonda específica marcada radioactivamente.

#### ***M-5.2.3.1. Electroforesis y transferencia del RNA***

El RNA se separa en geles desnaturalizantes de urea-poliacrilamida (6% acrilamida, 7M urea, 1X TBE). Las dimensiones del gel fueron de 16x16 cm. Para la preparación de las muestras se cargó en cada pocillo 10-20 µg de RNA y resuspendidos en tampón de carga 2X (0.3% azul bromofenol, 0.3% xileno-cianol FF, 10 mM EDTA ph 7.5, 97.5% formamida desionizada). Como marcador se utilizó el pGEM marcado con  $\gamma^{32}$ -ATP y mezclado con el tampón de carga en proporción 1:1.

Se realiza una pre-electroforesis a 450 V en 1X TBE durante 30-60 min hasta que el gel alcance una temperatura aproximadamente de 50°C. Se desnaturalizan las

muestras, antes de cargarlas en el gel, 5 min a 95°C y se pasan a hielo rápidamente. La electroforesis se realiza 1 h a 450 V, hasta que el azul de bromofenol alcance el final del gel. Se retira el gel y se corta el carril de los marcadores para secarlo y posteriormente exponerlo en el PhosphorImager.

Para la transferencia se cortan 6 papeles Whatman 3MM y la membrana (Zeta-Probe, Hybond N<sup>+</sup> Amersham Biosciences) al tamaño del gel y se humedecen en 1X TBE. Uno de los papeles 3MM se pega al gel evitando la formación de burbujas y sobre él se ponen los otros dos 3MM. Con la ayuda de los tres papeles se retira el gel del cristal y se termina de formar el *sandwich* poniendo sobre el gel la membrana y los otros 3 papeles 3MM. Se coloca todo en la orientación adecuada (transferencia hacia el polo positivo) en la unidad de transferencia LKB-Amersham. Se transfiere a 0.5-3 mA/cm<sup>2</sup> de membrana durante 30-60 min. (150 mA 60 min sería una transferencia estándar). Terminada la transferencia se fija el RNA a la membrana calentando el filtro 1 h a 80°C y a vacío.

#### ***M-5.2.3.2. Hibridación***

Como sonda se utilizaron oligonucleótidos marcados en su extremo 5' con  $\gamma^{32}$ -dATP. El marcaje se realizó con la polinucleótido quinasa (NEB) según la siguiente reacción: oligo (50 pmol),  $\gamma^{32}$ -ATP (20 $\mu$ Ci), tampón 10X polinucleótido quinasa y la enzima (5U), llevados a un volumen final de 10  $\mu$ l. Se incubó 1 h a 37°C, se llevó a un volumen de 25  $\mu$ l y se purificó con columna G-25 (Bio-rad).

La prehibridación se realizó a 42°C durante 30 min con 20 ml de solución de prehibridación (0.5 M tampón fosfato Na<sup>+</sup>, pH 7.2, 7% SDS, 10 mM EDTA) para una membrana de 16x16 cm. Se desnaturalizó la sonda calentando 5 min a 95°C y se añadieron 10<sup>6</sup> cpm/ml a la solución de hibridación (normalmente todo el volumen de un marcaje típico de 50 pmol de oligo) para hibridar a 42°C durante toda la noche. Se retiró la sonda y se lavó la membrana sucesivamente con 2XSSC/0.1%SDS 1 minuto, 2XSSC/0.1%SDS 15 min, 1XSSC/0.1%SDS 15 min, 0.1XSSC/0.1%SDS 15 min, todos ellos a 42°C. Se cubrió la membrana con plástico y se expuso durante toda la noche en pantalla PhosphorImager. Los resultados se visualizan en escáner *Personal Molecular Imager® FX* sustentado por el programa *Quantity One®*.

#### **M-5.2.4. HIBRIDACIÓN DE MICROARRAYS**

##### ***M-5.2.4.1. Los microarrays Sm14KOLI***

La universidad de Bielefeld diseñó los *microarrays* de DNA Sm14kOLI, que contienen 14.976 sondas de hibridación duplicadas y fijadas sobre una resina epoxi,

distribuidas en 48 gradillas de 26 filas y 24 columnas. Estas sondas de hibridación representan 6.208 oligonucleótidos de 70 nt complementarios a los distintos genes codificados en el genoma de *S. meliloti*; a su vez, también contienen 8.080 oligonucleótidos de 50 a 70 nt de longitud complementarios a 2881 regiones intergénicas, algunas de las cuales están representadas por más de un oligonucleótido complementario a ambas hebras del DNA. Los *microarrays* presentan una serie de controles que se corresponden con 17 oligonucleótidos de 70 nt que determinan la especificidad de hibridación (la identidad de sus secuencias varía en un rango de 70-90%); el control negativo son 5 oligonucleótidos 70-mer de secuencias no presentes en *S. meliloti*; así mismo, también se incluyen posiciones vacías y algunas que contienen el tampón de para la fijación de los oligonucleótidos al soporte. Finalmente, también se incluyen 4 oligonucleótidos 70-mer complementarios a distintos transgenes (*gusA*, *lacZ*, *nptII*, *aacC1*).

#### **M-5.2.4.2. Marcaje de las muestras**

El protocolo utilizado es una adaptación del descrito por Becker and Rüberg en el Manual de *S. meliloti* Sm6KOLigo Microarray.

Se procede con el anillamiento de 10-20 µg de RNA con 10 ng de *random* hexámeros en un volumen final de 16 µl (15 µg) durante 10 min a 70°C. Se incubó en hielo 5 min y posteriormente se añade la siguiente mezcla de reacción:

5x first strand buffer	6 µl
0.1 M DTT	3 µl
50x dNTPs (con aa-dUTP)	0,6 µl
RNase inhibitor (40 U/µl)	0,5 µl
SuperScript II (200 U/µl)	1,5 µl

El *Stock* de 50X dNTPs con aminoalil-dUTP se preparó a partir de una solución madre 100 mM aa-dUTP: 1 mg aa-dUTP+17 µl H<sub>2</sub>O+0,68 µl 1 N NaOH. Se toma una gota y se mide el pH hasta ajustar a pH 7 (si es necesario se pueden seguir añadiendo 0,1-0,2 µl 1 N NaOH). Guardar a -20°C. Las proporciones fueron 4:1 aa-dUTP/dTTP mix; 100 mM dATP 19,4 µl (concentración final 25 mM); 100 mM dCTP 19,4 µl (concentración final 25 mM); 100 mM dGTP 19,4 µl (concentración final 25 mM); 100 mM dTTP 3,9 µl (concentración final 5 mM); 100 mM aa-dUTP 15,5 µl (concentración final 20 mM). Se preparan alícuotas de 4 µl y se guardan a -20°C.

La RT sintetiza el cDNA durante 1 h a 42°C, momento en que se añade 1 µl adicional de SuperScript II manteniendo la reacción a 42°C durante 1 h más. En este punto se podrían guardar a -20°C

Se procede a eliminar el RNA mediante la adición de 15  $\mu$ l NaOH 0,2 M (mezclar pipeteando) manteniendo a 70°C durante 10 min. Posteriormente la solución se neutralizó añadiendo 15  $\mu$ l HCl 0,2 M. Rápidamente se purifica el cDNA mediante la utilización de CyScribe GFX Purification kit, GE. Se añadieron 450  $\mu$ l de *capture buffer* al tubo y se transfirieron a la columna CyScribe GFX; se lavó la columna 3 veces con 600  $\mu$ l EtOH 80% y el cDNA se eluyó con 60  $\mu$ l NaHCO<sub>3</sub> 0,1 M (pH 9,0).

Esta técnica se basa en un de marcaje indirecto del cDNA por lo que debemos acoplar el fluoróforo correspondiente en los grupos aminoalil presentes en nuestro cDNA. Los fluoróforos se preparan diluyendo la solución proporcionada por la casa comercial. Se añaden 10  $\mu$ l de DMSO (libre de agua), se mezclan muy bien (cuidado porque no se disuelven) y se reparten en alícuotas de 1  $\mu$ l en tubos opacos para posteriormente desecar durante 1 h en el *speed-vac*. Se guardan en pareja Cy3/Cy5 en bolsas con perlitas desecantes y a -20°C. También se pueden usar directamente con el DMSO. Se resuspende el fluoróforo en los 60  $\mu$ l del paso anterior y se incuba entre 1 y 2 h en oscuridad a T<sup>a</sup> ambiente. El bloqueo de los grupos reactivos no acoplados se lleva a cabo mediante el tratamiento durante 15 min en oscuridad a T<sup>a</sup> ambiente con 4,5  $\mu$ l hidroxilamina 4 M (mezclar pipeteando). Se hace una limpieza de los fluoróforos no incorporados mediante CyScribe GFX Purification kit, GE: se añadieron 600  $\mu$ l *capture buffer* a la muestra marcada con Cy5, se mezcla pipeteando y se pasa esta mezcla a la muestra marcada con Cy3 para acabar pasando por la columna CyScribe GFX; se lava 3 veces con 600  $\mu$ l *washing buffer* (kit) y se eluye con 60  $\mu$ l *elution buffer* (kit) al centro del filtro. Se incuba durante 5 min en posición vertical en oscuridad.

Para reducir el volumen de la sonda se deseca en *speed-vac* hasta un volumen aproximado de 5  $\mu$ l y se añaden 45  $\mu$ l de DIG Easy Hyb solution+1  $\mu$ l de DNA de esperma de salmón (5  $\mu$ g/ $\mu$ l).

#### ***M-5.2.4.3. Pretratamiento de los microarrays***

Previo a la hibridación debemos tratar los *microarrays*. Se preparan las soluciones frescas con cada hibridación. Las incubaciones se realizan en oscuridad y en continua agitación; 5 min a T<sup>a</sup> ambiente en Solución 1 (Triton X-100 0,001%), dos veces 2 min a T<sup>a</sup> ambiente en Solución 2 (HCl 0,0032%), 10 min a T<sup>a</sup> ambiente en Solución 3 (KCl 100 mM), 1 min a T<sup>a</sup> ambiente en agua miliQ. Se mantiene 15 min en solución de bloqueo (1x QMT Blocking solution (Nexterion) + 0,0075% HCl) a 50°C previamente atemperada.

#### **M-5.2.4.4. Hibridación y lavados**

Las muestras marcadas se desnaturalizaron durante 5-10 min a 60°C, para posteriormente incubar durante 30 min a 37°C. Se aplicaron sobre la parte central del portaobjetos (previamente caliente 50°C) evitando formar burbujas. Al poner el cubre la solución debe distribuirse homogéneamente entre ambos cristales. Se colocó en el interior de la cámara de hibridación y se incubó en un baño de agua a 42°C durante al menos 16h.

Finalizada la hibridación, se realiza un primer lavado rápido que nos permite retirar el cubre en 2XSSC/0,2%SDS a 42°C pudiendo mantener la agitación hasta 30 seg. Posteriormente los portaobjetos se incuban en agitación en una solución 2XSSC/0,2%SDS a 42°C durante 5 min. Se hicieron lavados sucesivos de 5 min en una solución 0,2XSSC/0,1%SDS a 42°C (agitando arriba-abajo) y dos veces más en 0,2XSSC a RT; cada lavado de 1min. Finalmente se hizo un lavado adicional de 1 min con 0,1XSSC frío (18°C). Inmediatamente se procede a secarlos centrifugando 5 min a 1200 r.p.m con rotor A-2-DWP.

#### **M-5.2.4.5. Revelado y procesamiento de los datos**

La adquisición de la imagen con los consiguientes valores de intensidad media de los fluoróforos para cada señal de hibridación se realizó con un escáner Axon 4100A y el programa de cuantificación GenePix Pro 5.1. Aunque este programa permite la normalización de los datos, se utilizó la plataforma *on-line* diseñada en el centro de biotecnología (CeBiTec) de la universidad de Bielefeld expresamente para la interpretación de los datos derivados de los *microarrays* utilizados, EMMA 2.8.2 (<https://www.cebitec.unibielefeld.de/groups/brf/software/emma/cgi-bin/emma2.cgi>; Dondrup *et al.*, 2003). Este programa permite, además de la normalización de los datos por distintos métodos, su tratamiento estadístico.

Los dos parámetros más significativos a tener en cuenta durante el análisis de los datos son el valor de A, que determina la intensidad media correspondiente a un punto de hibridación calculado en función de las diferentes replicas, y el valor de M, que refleja el balance entre las dos muestras fluorescentes. El cálculo de A se realiza de acuerdo a la siguiente fórmula:  $A = \log_2 (R_i \times G_i) 0,5$ , donde  $R_i$  se corresponde con la intensidad de la señal menos la intensidad inespecífica del fondo para la muestra problema y  $G_i$  es igualmente la resta de intensidades, pero en la muestra control. M se calcula aplicando el logaritmo en base 2 al cociente entre  $R_i$  y  $G_i$ .

### M-5.2.5. REVERSO-TRANSCRIPCIÓN (REACCIÓN RT)

La presencia de transcritos específicos en una muestra de RNA se valoró mediante RT-PCR, en algunos casos. El proceso general de anillamiento y reacción RT es similar al comentado en el apartado 4.2.4.1., en el marcaje de la sonda para *microarrays*, con la única salvedad es que para la síntesis del cDNA se utilizó una mezcla de dNTPs carente del nucleótido modificado aa-dUTP, que se sustituyó por dUTP estándar. La hidrólisis del RNA no se realizó mediante un procesos alcalino sino que se utilizaron 10 U de RNasa H (Roche).

El cDNA resultante se conservó a 4°C y fue susceptible de ser amplificado por PCR estándar (RT-PCR).

### **M-5.3. MANIPULACIÓN DE PROTEÍNAS**

#### M-5.3.1. EXTRACCIÓN DEL EXTRACTO PROTEICO TOTAL DE *S. MELILOTI*

Para obtener la fracción de proteínas totales de células de *S. meliloti* en cultivo, se crecieron en 4 ml de medio TY hasta alcanzar una densidad óptica específica del experimento. El cultivo celular se sometió a centrifugación durante 2 min a 12.000 r.p.m. y las bacterias se resuspendieron en 100 µl de una solución que contiene 4 mM Tris-HCl pH 8, 30 mM KCl, 0,2 mM MgCl<sub>2</sub>, 1 mM DTT y 0,2 mM de fluoruro de fenilmetilsulfonilo. Seguidamente se sonicó la muestra en un Sonicador Bransson, a 4°C y con 3 pulsos de 15-20 seg cada uno. El extracto total se preservó en hielo o se almacenó a -20°C en los casos en los que su utilización no fuera inmediata.

#### M-5.3.2. FRACCIONAMIENTO CELULAR DE *S. MELILOTI*

Para separar las distintas fracciones celulares (periplasma, citoplasma y membrana) de células en vida libre de *S. meliloti* se siguió el protocolo de fraccionamiento celular descrito por Eggenhofer *et al.*, 2004. Las muestras resultantes del fraccionamiento se utilizaron para Western-blot y análisis proteómico.

Se parten de 200 ml de cultivo bacteriano en medio completo TY que se centrifuga a 5000xg, 5min a 4°C. Las células se resuspenden en 2 ml de TEXBuffer (50 mM Tris-HCl pH 8, 3 mM EDTA, 0.1% Triton X-100) y se incuban durante 45 min en hielo. Posteriormente se centrifugan a la misma velocidad y se lavan de nuevo con 2 ml de TEX Buffer. El sobrenadante de ambas centrifugaciones es la **FRACCIÓN DE PROTEÍNAS PERIPLASMICAS**.

El pellet sobrante se resuspende en 4 ml de un tampón de nucleasas (10mM MgCl<sub>2</sub>, 20 mM Tris-HCl pH 8, 50µg/ml RNasa I, 50µg/ml DNasa A) y se incuba en



hielo 30 min. Se pasa 3 veces por la prensa de French (alternativamente se puede sonicar 3 veces 1 min) y el extracto resultante se somete a centrifugación a 2000xg 2 min a 4°C para eliminar las células no lisadas. El extracto resultante se precipita por ultracentrifugación a 160.000xg 1,5 h a 4°C. El sobrenadante representa la **FRACCIÓN DE PROTEÍNAS CITOPLÁSMICAS** y el pellet como la **FRACCIÓN NO SOLUBLE DE PROTEÍNAS DE MEMBRANA**.

El mismo procedimiento se puede realizar sin tratamiento con TEX Buffer como control (opcional).

En cada fracción las proteínas se precipitan mediante ácido tri-cloroacético (TCA). Para ello, se añadieron 12,5 µl de una solución de 2% (w/v) de desoxicolato sódico pH 8,4 a 500 µl de volumen de muestra incubándose 20 min en hielo. Se añaden 24 µl de una solución al 50% (w/v) de TCA (ácido tricloroacético) y se incuba 15 min en hielo (se puede precipitar a -80°C toda la noche, ON). Las proteínas precipitadas se recogen por centrifugación a 12000rpm 5-10 min. Se lava dos veces el pellet con acetona fría al 80% (se guarda a -20°C). Finalmente, se seca al aire y se resuspende en 30 µl de 1X de tampón de carga con 0.5% β-MercaptoEtOH o DTT.

### M-5.3.3. ELECTROFORESIS DE PROTEÍNAS EN GELES SDS-PAGE

Para separar extractos proteicos mediante su peso molecular se recurrió a electroforesis vertical en geles de acrilamida-bisacrilamida desnaturalizantes, con SDS (SDS-PAGE), según la metodología propuesta por Laemmli, 1970.

Los geles se utilizaron para visualización directa de las proteínas mediante tinción o para transferirlos y someterlos a análisis mediante Western-blot.

#### *M-5.3.3.1. Preparación de los geles.*

Los volúmenes descritos a continuación (en ml) corresponden a 2 geles de 0,75 mm de espesor en el sistema REAL® Sub-Mini 10x10 cm Dual System (REAL), para condiciones desnaturalizantes SDS-PAGE:

Tablas M-5 y M-6. Cantidades recomendadas para geles SDS-PAGE.

<b>GEL DE RESOLUCIÓN (Inferior)</b>	<b>10%</b>	<b>12%</b>	<b>15%</b>
Acrilamida/bis-acrilamida (40%)	3	3,6	4,5
Agua milliQ	5,8	5,2	4,14
Tris-HCl 1,5M pH 8,8	3	3	3
SDS 10%	0,12	0,12	0,12
APS (10%)	0,12	0,12	0,12

TEMED	0,012	0,012	0,012
-------	-------	-------	-------

<b>GEL DE EMPAQUETAMIENTO (Superior)</b>	<b>5%</b>
Acrilamida/bis-acrilamida (40%)	0,375
Agua milliQ	2,225
Tris-HCl 1,0M pH 6,8	0,38
SDS 10%	0,03
APS (10%)	0,03
TEMED	0,005

Primeramente, se prepara el gel inferior mezclando los componentes en el orden de redacción seguido en la tablas M-5 y M-6. A continuación se le añaden rápidamente tres gotas de Isopropanol (permite una polimerización totalmente horizontal en su superficie) y se deja polimerizar 30 min. Posteriormente, se prepara el gel superior de igual manera y se coloca el peine con los pocillos, dejándose polimerizar otros 30 min.

#### ***M-5.3.3.2. Electroforesis***

Como tampón de electroforesis en condiciones desnaturizantes SDS-PAGE para el sistema de electrodos en cubetas REAL® Sub-Mini 10x10 cm Dual System (REAL) se empleó el descrito por Laemmli, 1970:

##### Tampón de desarrollo, *Running Buffer* (10X)

0,124 M Tris-HCl (30,25 g)

1,252 M glicina (144 g)

5 % SDS (20 g)

H<sub>2</sub>O c.s.p. 1 l

Ajustar con HCl concentrado hasta pH 8,8

Las muestras se prepararon adicionando en proporción 3:1 (v/v) tampón de carga con la siguiente composición:

##### Tampón de carga, *Sample Buffer* (3X)

0,01 % (p/v) azul de bromofenol

12 % SDS (p/v)

300 mM Tris-HCl pH 6,8

60 % glicerol (v/v)

Las muestras proteicas se mezclaron con este tampón de carga suplementado con 10 % DTT o 600 mM de  $\beta$ -mercaptoEtOH, y se desnaturalizaron inmediatamente antes de ser cargadas en el gel mediante ebullición durante 3 min y posterior pase a hielo durante 5 min. Las electroforesis se corrieron normalmente a 150-200V hasta que el azul de bromofenol alcanzara el frente del gel.

Para la estimación aproximada del peso molecular de las proteínas resueltas en los genes se utilizaron dos patrones de peso molecular: el primero fue un marcador estándar de bajo rango, sin teñir suministrado por Bio-Rad®, en el que se observa las bandas correspondientes a las proteínas: fosforilasa b de 97,4 kDa, Seroalbúmina bovina de 66,2 kDa, Ovoalbúmina de 45 kDa, anhidrasa carbónica de 31 kDa, Tripsinógeno de 21,5 kDa y lisozima de 14,1 kDa; el segundo fue un marcador de amplio rango (*Kaleidoscope Molecular Marker*, Bio-Rad®), que comprende desde 10 kDa hasta 250 kDa.

#### ***M-5.3.3.3. Detección de proteínas en geles de poliacrilamida***

La detección de proteínas se llevó a cabo mediante los siguientes métodos de fijación y tinción:

##### **Tinción con Azul de Coomassie**

Tras el desarrollo de la electroforesis, el gel se sumergió en una solución de EtOH:ácido acético:agua en proporciones 46:45:9 junto con azul Coomassie R-250 Brilliant Blue (Bio-Rad®) 0,25 % donde se mantuvo durante 15 - 30 min a temperatura ambiente en agitación suave. El desteñido se realizó en una solución de EtOH : ácido acético : agua en proporciones 5 : 7 : 78. Para facilitar la eliminación del color azul y reducir el gasto de solución se añadió un trozo de esponja que fue retirado y sustituido conforme atrapó la coloración. El proceso de desteñido se continuó hasta la completa decoloración del gel y la visualización precisa de las bandas de proteína.

##### **Tinción con nitrato de plata**

Tras el desarrollo de la electroforesis, el gel se sometió a fijación y tinción siguiendo una modificación del procedimiento descrito en Blum *et al.*, 1987. El gel se sumergió, durante 30 min, con agitación suave, en 100 ml de una solución de EtOH al 40 %, ácido acético al 10 % y 50 % agua milliQ. Posteriormente, el gel se incubó durante 30 min en 100 ml de una solución de sensibilización que contiene tiosulfato sódico 0,2 % (p/v), EtOH al 30%, acetato sódico 6,8 % (p/v) y agua. Tras el pretratamiento se lavó tres veces durante 5 min con agua milliQ. Se continuó

incubando el gel durante 20 min en 100 ml de una solución extemporánea de nitrato de plata al 0,25 % en agitación suave y en oscuridad. Se retiró la impregnación de plata con dos nuevos lavados de 1 minuto exacto con agua milliQ. Finalmente se reveló añadiendo 100 ml de una solución de carbonato sódico al 2,5% y formaldehído al 0,004 % de una solución al 37 %. En el momento en que aparecieron las bandas de proteínas, se añadió la solución de parada (EDTA sódico al 1,46% (p/v) en la que el gel se mantuvo un mínimo de 10 min para detener completamente el revelado. Finalmente se hicieron 3 lavados de 5 min cada uno con agua milliQ.

#### ***M-5.3.3.4. Conservación de geles***

Para conservar los geles durante cortos períodos de tiempo se sumergieron en agua milliQ y se mantuvieron a 4°C. En el caso de la conservación durante largos periodos de tiempo (más de 2-3 meses) se añadió al agua milliQ azida sódica al 0,2%.

#### **M-5.3.4. DETECCIÓN INMUNOLÓGICA DE PROTEÍNAS (WESTERN-BLOT)**

Para la detección de una proteína en un extracto celular se recurrió a análisis mediante Western-blot. Este protocolo consiste en una incubación con un anticuerpo primario que reconoce a la proteína de interés presente en el extracto (en nuestro caso siempre fue un anticuerpo monoclonal frente al epítipo FLAG [*Mouse anti-FLAG® antibody*, Sigma]), seguida de una segunda incubación con un anticuerpo secundario que se une al anticuerpo primario y que lleva conjugado una fosfatasa alcalina. Finalmente se revela mediante la adición de un sustrato que, en presencia de la enzima, da lugar a una señal luminosa.

#### ***M-5.3.4.1. Transferencia de proteínas a membranas de PVDF***

Tras la electroforesis según el protocolo descrito, las proteínas se transfirieron a filtros BioTRace PVDF de 0,45 µm de poro (Sigma) mediante el sistema Mini Trans-Blot® Electrophoretic Transfer Cell (Amersham-GE Healthcare™). Previamente a la transferencia la membrana de PVDF se lavó en metanol al 100 % durante 10 seg para su activación. Tanto el gel como la membrana se equilibraron en agitación suave durante 5 min en tampón de transferencia compuesto por 25 mM Tris pH 8,3, 192 mM glicina, 20 % metanol. A continuación y sobre un *cassette* de transferencia se dispuso el siguiente orden de montaje: esponja y papel Whatman 3MM empapados en tampón de transferencia. Sobre éstos, se colocó el gel de poliacrilamida previamente equilibrado, la membrana de PVDF equilibrada eliminando las posibles burbujas formadas y finalmente, otro papel Whatman y esponja empapados en tampón de transferencia. El montaje así formado se dispuso sobre el módulo de electrodos siguiendo la orientación recomendada por el fabricante y todo ello se

introdujo en una cubeta de electroforesis con tampón de transferencia. La transferencia se llevó a cabo aplicando una intensidad de 0,8 mA/cm<sup>2</sup> durante 50 min. Para comprobar la completa transferencia de proteínas, la membrana de PVDF se tiñó con una solución al 0,5% (p/v) de Rojo Ponceau (Ponceau S, SIGMA®), en ácido acético 2% (v/v) durante 5 min a temperatura ambiente. Las bandas se visualizaron destiñendo con agua destilada, hasta alcanzar el contraste adecuado. La tinción fue totalmente eliminada con sucesivos lavados de agua destilada.

#### ***M-5.3.4.3. Revelado inmunológico***

La membrana de PVDF con las proteínas transferidas, se sometió al bloqueo de sitios inespecíficos de unión al anticuerpo mediante incubación durante 1 h, con agitación suave a temperatura ambiente, en una solución que contenía leche desnatada en polvo al 1,5% en tampón de bloqueo (20 mM Tris-HCl pH 7,8, 0,12 M NaCl). Posteriormente se lavó la membrana con el mismo tampón sin leche desnatada durante 10 min 3 veces. Tras este paso se incubó durante 1 h a temperatura ambiente en agitación suave con una dilución del anticuerpo primario o anti-FLAG de 1/1000 (10 ml) en tampón de bloqueo con leche desnatada. Tras la incubación se eliminó el exceso de anticuerpos con varios lavados de 10 min con tampón de bloqueo sin leche desnatada. A continuación, la membrana fue incubada 1 h a temperatura ambiente, con una dilución del anticuerpo secundario anti-IgG de ratón marcado con fosfatasa alcalina (Sigma) de 1/50.000 (50 ml), en tampón de bloqueo con leche desnatada. Se lavó la membrana varias veces con tampón de bloqueo sin leche desnatada para eliminar el exceso de anticuerpo secundario. Se preparó el tampón luminol (50 mM Tris-HCl pH 8,6, 150 mM NaCl) y se añadieron a 10 ml del mismo en oscuridad, una concentración de 0,04 % de 3-aminofthalhidracida (luminol, Sigma), 100 µl de 4-Iodofenol (de un stock con 10 mg en 1 ml de DMSO, conservado a 4°C) y 3,2 µl de H<sub>2</sub>O<sub>2</sub> (30%). La membrana se incubó durante 2 min en esta solución y rápidamente se cubrió con papel transparente y se dispuso sobre un casete de fotografía. Sobre ésta se colocó una película autorradiográfica (Kodak X-Omat®) y se expuso (5 - 15 min) para la impresión de ésta. Transcurrido este tiempo se reveló la película mediante el revelador y el fijador de TETENAL® a las diluciones y tiempos recomendados por la casa comercial.

#### **M-5.3.5. PROTEÓMICA**

El protocolo llevado a cabo consta de dos etapas: isoelectroenfoque o primera dimensión y electroforesis SDS-PAGE o segunda dimensión.

### **M-5.3.5.1. Isoelectroenfoque**

Se utilizaron dos tipos de tiras de 13 cm (*Inmobiline™ DryStrips*, Amersham-GE healthcare™), para cubrir rangos de pH de 3 a 11 y de 4 a 7. Las tiras se rehidrataron con 325 µl una solución que contiene 8 M de urea, 2 M de tiourea, 2 % p/v de CHAPS y azul de bromofenol, además de 2 % de *DeStrip* (Amersham) y 0,5 % de *IPG Buffer* (del mismo rango de pH que la tira IPG). En esta misma solución se añadieron las diferentes muestras proteicas de estudio. En cada tira se cargaron aproximadamente de 100 - 500 mg de proteína. Una vez depositada la solución de rehidratación con la muestra en cada tira, se cubrieron por completo con aceite mineral para evitar la desecación de la tira y la consecuente precipitación de la urea. El proceso de rehidratación activa se realizó durante 23 h con un voltaje constante de 50 V y a 22 °C en una *Protean IEF cell* (Bio-rad).

Pasado ese tiempo las tiras se sometieron a un programa de isoelectroenfoque que consistió en un primer paso de 30 min a 250 V, subida gradual hasta 8000 V durante 1 h (*ramping*), 8000 V hasta 20 KW/h y mantenimiento a 250 V durante toda la noche, para evitar la disipación del gradiente de pH. Una vez finalizado el programa se guardaron las tiras a -80°C hasta la segunda dimensión.

### **M-5.3.5.2. Segunda dimensión (SDS-PAGE)**

Antes de proceder a la separación electroforética de las proteínas sepadas en las tiras IPG, se sometieron a un equilibrado que consistió en: 15 min en 5 ml de *SDS equilibration buffer* (50mM Tris-HCl pH 8,8, 6 M Urea, 30 % glicerol, 2 % SDS y azul de bromofenol) + 330 µl DTT 1M/tira. En este paso se reducen las proteínas presentes en la tira por el DTT; 15 min en 5 ml de *SDS equilibration buffer* + 225 mg de Iodoacetamida/tira. La iodoacetamida elimina los grupos carbamilados de las proteínas; 15 min en *running buffer* 1X.

Para la preparación del gel SDS-PAGE se utilizó el sistema Protean II (Bio-rad). En la siguiente tabla se muestran las cantidades para dos geles de 18 x 16 cm, dependiendo del porcentaje de acrilamida/bisacrilamida que se desee utilizar.

Tabla M-7. Cantidades recomendadas para geles SDS-PAGE destinados a proteómica.

	10 %	12 %	15 %	18 %	20 %
Acrilamida/Bis (40%)	31,25	37,5	46,88	56,25	62,5
Tris-HCl 1,5 M, pH 8,8 (375 mM)	31,25	31,25	31,25	31,25	31,25
SDS 10% (0,2%)	2,5	2,5	2,5	2,5	2,5
H <sub>2</sub> O	60	53,75	44,37	35	28,75
VOL. FINAL	125	125	125	125	125

La concentración que nosotros usamos fue del 12 %. Se añaden 1,5 ml (0,750 ml) de PSA 10% y 100 µl (50 µl) de TEMED. Añadir 100 µl de isobutanol hasta la polimerización completa (dejar 1 cm mínimo para colocar la tira IPG).

Una vez polimerizado el gel, se coloca la tira IPG equilibrada en posición horizontal sobre el gel SDS-PAGE, verificando que no se formen burbujas. Si se precisa de marcador se coloca un papel secante encima del gel con 10-20 µl del marcador de peso molecular que corresponda, colocado siempre en el polo ácido (+). Inmediatamente se añade agarosa de bajo punto de fusión al 1% con azul de bromofenol, hasta cubrir totalmente la tira. Se deja que solidifique y se monta la electroforesis. Los geles se corren 30 min a 20 mA/gel (46-52 V) y después a 35 mA/gel (85-90 V) hasta la salida del azul de bromofenol (8-10 h).

Una vez finalizada la electroforesis, los geles se sometieron a tinción con plata o con azul de Coomassie siguiendo los protocolos detallados en el apartado 5.3.3.3.

## M-6. ENSAYOS CON PLANTAS

### M-6.1. ESTERILIZACIÓN Y GERMINACIÓN DE SEMILLAS

Las semillas de *M. sativa* (alfalfa) fueron sumergidas en una solución de HgCl<sub>2</sub> al 2,5% durante 9 min en agitación. En condiciones de esterilidad, se lavaron 5 a 6 veces abundantemente con agua destilada estéril para posteriormente mantenerlas en imbibición durante 1 - 2 h también en agua destilada estéril. Finalmente se lavaron de 2 a 3 veces con agua destilada estéril y se colocaron en placas Petri sobre medio agar-agua (1.5% agar en agua desionizada). Se germinaron en oscuridad a 28 °C hasta que las plántulas alcanzaron una longitud de unos 10 mm (24 h aproximadamente).

### M-6.2. SOLUCIÓN NUTRITIVA PARA PLANTAS

Se ha empleado una solución mineral derivada de la descrita por Rigaud and Puppo, 1975, cuya composición se describe a continuación:

Macroelementos (H<sub>2</sub>O c.s.p. 1 l):

KH <sub>2</sub> PO <sub>4</sub>	0,2 g
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0,2 g
K <sub>2</sub> SO <sub>4</sub>	0,2 g
CaSO <sub>4</sub>	0,12 g
EDTA Férrico	0,05 g

Microelementos (H<sub>2</sub>O c.s.p. 1 l):

Na <sub>2</sub> MoO <sub>4</sub> · 2H <sub>2</sub> O	4 mg
H <sub>3</sub> BO <sub>3</sub>	18,6 mg
MnSO <sub>4</sub> · 4H <sub>2</sub> O	22,6 mg
ZnSO <sub>4</sub> · 7H <sub>2</sub> O	29 mg
CuSO <sub>4</sub> · 5H <sub>2</sub> O	24 mg.

El pH de la solución se ajustó a 7 con KOH. La solución se esterilizó en autoclave a 120 °C durante 20 min.

Para los ensayos en placas cuadradas con medio sólido, se utilizó el mismo medio suplementado con 1,5% agar bacteriológico.

**M-6.3. CULTIVO DE PLANTAS**M-6.3.1. CULTIVO AXÉNICO DE TUBOS

El cultivo axénico de plantas de alfalfa se realizó según la técnica descrita por Olivares *et al.*, 1980. Previamente se esterilizaron mediante autoclavado a 120 °C durante 20 min tubos de 20 x 200 mm conteniendo un soporte de papel de filtro y 10 ml de solución mineral de Rigaud y Puppó. Cuando las plántulas de alfalfa recién germinadas alcanzaron aproximadamente 1 cm de longitud (1 día de incubación a 28°C), se colocaron en condiciones asépticas en los tubos (1 planta/tubo). Se dejaron los tubos con las plantas en oscuridad durante 1-2 días. Para evitar que la luz incidiera directamente sobre las raíces, los tubos se cubrieron en su mitad inferior con papel opaco y se llevaron a la cámara de cultivo de plantas, mantenida en las siguientes condiciones: 500  $\mu$ Einstein.m<sup>-2</sup>.s<sup>-1</sup> (longitud de onda: 400-700 nm) de intensidad luminosa, fotoperíodo de 16/8 h luz/oscuridad, 23/17 °C de temperatura día/noche y 50 % de humedad relativa.

Cuando las plantas tuvieron 7 - 10 días, se inocularon con 1 ml de una suspensión de aproximadamente 10<sup>6</sup> células/ml de la cepa a estudio, así como un control no inoculado.



### M-6.3.2. CULTIVO AXÉNICO EN PLACAS DE AGAR

Para el cultivo de plantas sobre medio sólido se utilizaron placas estériles de 6x6 cm con medio Rigaud y Puppo suplementado con 1.5 % de agar bacteriológico. Se tomaron semillas esterilizadas y germinadas según el protocolo descrito en el apartado 6.1 y las plántulas de 1-2 días se colocaron sobre la placa a media altura, no colocando nunca más de 6 plantas. Las placas se pusieron en cámara de cultivo con las condiciones descritas en el apartado anterior durante 7 días y en posición vertical ligeramente inclinadas, tapando la parte de la raíz con un papel opaco.

Para su inoculación se preparó una suspensión bacteriana de  $10^6$  células/ml crecidas en medio TY. Esta suspensión se lavó 2 veces con agua destilada estéril y se inoculó a razón de 10 ml por placa. Las placas se dejaron horizontalmente en esterilidad durante 1 h, eliminándose la solución sobrante pasado ese tiempo. Las placas se volvieron a colocar en la cámara de cultivo durante 30 días, para finalmente utilizarlas en los experimentos pertinentes.

### M-6.3.3. CULTIVO EN JARRAS LEONARD

Se trata de un dispositivo de cultivo de plantas descrito por Leonard, 1943, y consistente en dos recipientes. El recipiente superior se rellenó con vermiculita y el inferior con solución mineral de Rigaud y Puppo, conectando ambos con tiras de papel de filtro y todo esterilizado mediante autoclavado a 120 °C durante 20 min. En cada jarra se sembraron semillas esterilizadas y germinadas, y se inocularon con 1 ml de una suspensión  $10^6$ - $10^7$  cel/ml de la cepa de estudio, así como un control no inoculado. Una vez inoculadas, las jarras se cubrieron con una capa de perlita estéril que evita la posible contaminación ambiental y refleja la luz, evitando un calentamiento excesivo de las semillas que pudiera alterar el desarrollo de las mismas. La junta entre el componente superior e inferior de la jarra se selló con papel *parafilm* y las jarras se envolvieron en papel opaco a la luz para evitar que ésta alcanzara las raíces.

Las jarras se trasladaron a cámara de cultivo de plantas mantenida en las condiciones del apartado 6.3.1.

### **M-6.4. ANÁLISIS DE LAS CINÉTICAS DE NODULACIÓN**

Para medir el comportamiento simbiótico de una cepa de *S. meliloti* a lo largo del tiempo se llevó a cabo un seguimiento de dos parámetros: el número de nódulos por planta y el porcentaje de plantas noduladas. Para ello se dispusieron tubos con plantas crecidas en cultivo hidropónico tal como se describe en el apartado 6.3.1. Una vez inoculadas se hizo un conteo de nódulos y de plantas en las que al menos

apareciera un nódulo (plantas noduladas) en intervalos diarios durante los primeros 15 días y cada 2-3 días en los siguientes, hasta completar el experimento (30 días).

Además, se monitorizó la aparición de nódulos fijadores (nódulos rosas), anotándose el número y el día en el que se visualizaba el primer nódulo fijador, por planta.

#### M-6.5. MEDIDA DEL GRADO DE COMPETITIVIDAD

Para medir el grado de competitividad, se inocularon de 12 a 24 tubos con plantas que tenían entre 7 y 10 días, con 1 ml de una suspensión celular que contenía una mezcla 1:1, 10:1 y 1:10 de la cepa a estudiar frente a una cepa marcada con el gen informador *gusA* (pGUS3) (García-Rodríguez and Toro, 2000), ambas cepas a una concentración final de  $10^6$  células/ml. Tras la inoculación se mantuvieron por un período de 30 días en cámara de cultivo de plantas en las condiciones anteriormente señaladas. Transcurrido este tiempo, las raíces de plantas noduladas fueron lavadas en agua destilada e incubadas sumergidas durante 24 h a 37 °C en una solución 1 mM X-gluc (5- bromo-4-cloro-3 indolil- $\beta$ -D-glucurónido), 50 mM de tampón fosfato sódico pH 7,5 con 1% SDS. Los nódulos ocupados con la cepa pGUS3 mostraron tinción azulada frente a nódulos blancos ocupados por la cepa a estudio. El grado de competitividad se calculó como el cociente de la proporción de nódulos ocupados por la cepa de estudio frente a la de nódulos azules. El error estadístico se calculó según la siguiente fórmula:

$$E = 1,96 \sqrt{\frac{P(1 - P)}{N}} \times 100$$

Donde, E = error estadístico

P = porcentaje en tanto por 1 de nódulos ocupados

N = número total de nódulos del experimento

#### M-6.6. MEDIDA DEL PESO SECO

Para determinar la eficiencia en la fijación de nitrógeno, se recurrió a la medida del peso seco, determinado después de 30 - 35 días de la inoculación. Para ello, la planta se secó en horno a 65 °C durante 2 días para asegurar la completa desecación. Finalmente las plantas fueron pesadas individualmente en una balanza de precisión.

#### M-6.7. REAISLAMIENTO DE BACTERIAS DE NÓDULOS DE LEGUMINOSAS

Se tomaron los nódulos de la planta y se esterilizaron en superficie mediante HgCl<sub>2</sub> al 0,25 % durante 5 min. Seguidamente se lavaron con agua estéril 5 ó 6 veces y

se trituraron individualmente en condiciones asépticas con una varilla de vidrio previamente esterilizada. Una gota del triturado se extendió sobre placas de medio TY adicionado de los correspondientes antibióticos, y se incubaron a 28 °C durante 4 ó 5 días.

## **M-6.8. PREPARACIÓN DE MUESTRAS DE SECCIONES DE NÓDULOS**

### **M-6.8.1. ESTUDIO DE MUESTRAS INCLUIDAS EN AGAROSA**

Para la visualización de secciones de 100 µm de grosor de nódulos maduros ocupados por la cepa de *S. meliloti* a estudiar, se utilizaron plantas de 30 días de *M. sativa* crecidas en placas Rigaud y Puppo/agar. Se preparó una solución de agarosa al 3 % en PBS. Los nódulos se lavaron con agua y se colocaron sobre un molde cilíndrico diseñado a tal efecto, se vertió la agarosa (evitando que esté muy caliente para preservar las estructuras simbióticas) y se esperó a su completa solidificación, evitando la formación de burbujas. Se eliminó la agarosa sobrante y se hicieron varios cortes transversales de 100 µm de grosor con un vibratomo Leica VT1200S. Los cortes se dispusieron en un portaobjetos para su posterior visualización en un macroscopio Nikon AZ100.

### **M-6.8.2. ESTUDIO DE MUESTRAS INCLUIDAS EN RESINAS DE ALTA RESOLUCIÓN (TECHNOVIT)**

Para la visualización de bacteroides en nódulos de distinta madurez se llevó a cabo una inclusión en resinas de alta resolución (Technovit TV7100). Este protocolo permite hacer secciones de hasta 5 µm de grosor, aunque las muestras sólo pueden ser visualizadas mediante microscopía de campo claro. Esta técnica se realiza en 6 días tal y como se detalla a continuación:

#### ***M-6.8.2.1. Primer día: recolección de los nódulos y fijación***

Para seguir la evolución de las estructuras simbióticas en el tiempo se recolectaron de 6 a 10 nódulos de cada cepa de estudio inducidos en plantas de *M. sativa* crecidas en cultivo axénico en placas de agar. Los nódulos aislados se incubaron en 1 ml de una solución 0,05 M de cacodilato sódico. Para mejorar la infiltración, las muestras se mantuvieron durante 20 min a vacío y a temperatura ambiente. Posteriormente, se reemplazó la solución de cacodilato por 1ml de glutaraldehído y se incubaron a 4°C durante toda la noche.

***M-6.8.2.2. Segundo día: deshidratación de los nódulos***

Se elimina completamente la solución de glutaraldehído de las muestras y se añade 1 ml de tampón cacodilato 0,05 M para dejar incubar durante 30 min a temperatura ambiente. Se repite esta operación dos veces y se elimina el fijador para comenzar con la deshidratación progresiva en gradientes crecientes de EtOH al 10 %, 30%, 50 % y 70 %. Todas las incubaciones se hicieron en 1 ml de cada solución durante 1 h a temperatura ambiente, exceptuando la última que se prolongó toda la noche y a 4°C.

***M-6.8.2.3. Tercer día: deshidratación de los nódulos (continuación)***

Se continúa con la deshidratación haciendo una incubación en una solución de EtOH al 90 % durante 1 h a 4°C y tres en EtOH al 100 %. Una vez que se ha terminado y se tienen los nódulos totalmente deshidratados, se lleva a cabo una infiltración progresiva de la resina. Para ello se preparan soluciones de concentraciones crecientes de Technovit al 25 %, 50 % y 75 %. Se incuban los nódulos en 3 ml de cada una de estas soluciones 1 h a 4 °C hasta terminar en una solución de la resina al 100 % que se deja toda la noche en agitación y a 4°C.

***M-6.8.2.4. Cuarto día: polimerización de la resina***

Una vez que la resina se encuentra totalmente infiltrada en las muestras de nódulos se procede a la polimerización de la misma en bloques. Para ello se colocan los nódulos infiltrados en un bloque suministrado por la casa comercial, situando en cada molde 5 nódulos como máximo. Se prepara la solución de fijación (66 µl del producto Hardener II en 1 ml de TV7100) y con ella se cubre totalmente el molde con los nódulos. Se tapa con un papel de plástico transparente y se deja 24 h a temperatura ambiente. Normalmente la polimerización se completa en ese tiempo pero ocasionalmente se puede precisar de una polimerización mayor (hasta 3 días).

***M-6.8.2.5. Quinto día: polimerización de la resina (continuación)***

Una vez que se ha polimerizado la resina en los moldes se prepara el adhesivo que va a permitir la extracción de cada molde del bloque general. Para ello se dispone de un sistema especial, el Histobloc®, que es suministrado por la casa comercial. Este sistema consta de un adhesivo y un adaptador que se va a fijar a la resina y va a permitir la extracción mecánica del molde. A tal efecto se prepara una solución compuesta por 1 ml de polvo TV3040 y 1 ml de solución TV3040. Una vez hecha la mezcla, rápidamente se coloca enzima del molde y se fija el adaptador. Se

repite con cada molde el mismo procedimiento y se deja secar durante 1 día a temperatura ambiente.

#### **M-6.8.2.6. Sexto día: realización de los cortes, tinción y visualización**

Una vez extraídos los moldes, se realizan los cortes longitudinales correspondientes a los nódulos incluidos en la resina. Los cortes se hicieron con un microtomo Leica RM2135 y fueron siempre de 7  $\mu\text{m}$  de grosor. Para ello se dispone de un baño de agua y los cortes que se van haciendo se van depositando en este baño, desechando siempre los primeros. Las secciones se pasan a un portaobjetos y se colocan encima de un bloque a 48°C para que se deshidraten y se fijen a la superficie.

Para la tinción, se preparó el colorante azul de toluidina (que tiñe específicamente células bacterianas) a una concentración de 15 mg/ml. Los portaobjetos con las secciones se sumergen en esta solución durante 10 min y, posteriormente, se lavan en agua unas 20 veces para eliminar el colorante no fijado. Para terminar, las muestras se tapan con un cubreobjetos, previa adición de una gota de fijador EUKITT que permite la conservación de las muestras durante largos períodos de tiempo.

La visualización se llevó a cabo en un microscopio Leica DMI6000 bajo campo claro y utilizando el software de imagen *Leica Image Software*®.

### **M-6.9. HIBRIDACIÓN *IN SITU* SOBRE NÓDULOS MADUROS**

Para la detección directa de transcritos Smr sobre nódulos maduros de *M. sativa* se llevó a cabo una hibridación *in situ* de una ribosonda correspondiente a la secuencia complementaria del transcrito. El proceso consta de tres etapas: síntesis de la ribosonda, fijación de los nódulos e hibridación.

#### **M-6.9.1. SÍNTESIS DE LA RIBOSONDA**

Para la síntesis de la sonda de RNA marcada con digoxigenina se utilizó el kit *DIG RNA Labeling T7/T3* (Roche). Para ello se clonó en el vector *pBluescript* KS II la secuencia completa de los genes de interés. Para generar la sonda se amplificó desde estos plásmidos utilizando la *LA-Taq* (Fermentas), con la pareja de oligonucleótidos M13 (apartado M-3 de este Materiales y Métodos). Los productos de PCR se utilizaron como molde de la reacción de transcripción *in vitro*, utilizándose la RNA polimerasa T7 para las sondas antisentido y la T3 para las sondas sentido (controles). La reacción consta de:

DNA molde	1 µg
10X NTP labeling mixture	2 µl
10X Transcription buffer	2 µl
RNase inhibitor	1 µl
RNA Polimerasa T3 o T7	2 µl
H <sub>2</sub> O c.s.p.	20 µl

La reacción se incubó 2 h a 37°C. Posteriormente se eliminó el DNA por la adición de 5 U de DNasa I (Roche), parando la reacción con 2 µl de EDTA 0,2 M pH 8. El RNA obtenido se utilizó en su totalidad para la hibridación *in situ*.

### M-6.9.2. PREPARACIÓN Y FIJACIÓN DE LOS NÓDULOS

Para este experimento se utilizaron nódulos de 30 días inducidos en *M. sativa* por la cepa salvaje *S. meliloti* 1021. La inclusión de los mismos se realizó en parafina según el protocolo de Valoczi *et al.*, 2006.

#### **M-6.9.2.1. Fijación**

Los nódulos se sumergen en una solución de paraformaldehído al 4%, 0,1 % Tween 20 y 0,1% Triton X100 en PBS (pH 7,4). Se somete a vacío durante 45 min y se incuba durante toda la noche a 4°C.

#### **M-6.9.2.2. Deshidratación**

Se realiza una incubación en solución salina (85 % NaCl) durante 30 min en hielo. Seguidamente se lleva a cabo una deshidratación progresiva en concentraciones crecientes de EtOH en solución salina (50 %, 70 % y 85 %) en pasos de 3 h de incubación en hielo, salvo el último que es a 4°C. Al día siguiente se prosigue con la deshidratación, incubándose las muestras sucesivamente en una solución al 95% de EtOH en H<sub>2</sub>O a 4°C, EtOH 100 % a 4°C durante 4 h y una nueva incubación en la misma solución a 4°C durante toda la noche.

#### **M-6.9.2.3. Inclusión en parafina, y generación de los cortes longitudinales**

En esta etapa, los nódulos se infiltran con parafina líquida previo tratamiento con un agente preservador de tejidos vegetales (Histoclear). Se llevaron a cabo los siguientes pasos: EtOH 100 % 2 h, EtOH / Histoclear 50 % 1 h, Histoclear 100 % 1 h, Histoclear 100 % 1 h e Histoclear 100 % 1 h, todos ellos a T<sup>a</sup> ambiente. El proceso finaliza con un tratamiento con Histoclear / Paraplast X-Tra 50 % durante 16 h y a 58°C.

Al día siguiente se cambia la solución Histoclear / Paraplast por parafina al 100 % y se deja otro día más. Pasado este tiempo se preparan los moldes y se pasan los nódulos en la parafina a éstos, dejándose todo el tiempo que sea necesario hasta su completa polimerización (mínimo un día a temperatura ambiente).

Para realizar los cortes se utilizó un microtomo Leica RM2135, siendo el grosor de los mismos de 10 µm. Las secciones se colocaron en portaobjetos y se utilizaron inmediatamente para la hibridación, pues el RNA presente en las muestras se puede degradar rápidamente.

### M-6.9.3. HIBRIDACIÓN *IN SITU* (ISH)

El protocolo seguido fue el propuesto en Valoczi *et al.*, 2006, que consta de los siguientes pasos:

#### ***M-6.9.3.1. Desparafinado y rehidratación de las secciones de nódulos***

Consta de una deshidratación rápida por EtOH y una progresiva hidratación utilizando concentraciones decrecientes de EtOH en la siguiente serie de tratamientos: Histoclear 100 % 10 min, Histoclear 100 % 10 min, EtOH 100 % 5 min, EtOH 100 % 2 min, EtOH / Solución salina 95 % 2 min, EtOH / Solución salina 85 % 2 min, EtOH / Solución salina 70 % 2 min, EtOH / Solución salina 50 % 2 min, EtOH / Solución salina 30 % 2 min y Solución salina 2 min. Todos los pasos se realizan a temperatura ambiente y con agitación continua.

#### ***M-6.9.3.2. Tratamiento de pre-hibridación e hibridación***

Las muestras se ponen en contacto con una serie de soluciones que van están destinadas a preservar la población de RNAs presentes. Cabe señalar que es necesaria una deshidratación y posterior hidratación de las muestras cuando se cambie una solución de incubación por otra. Los pasos son (a temperatura ambiente): Tampón Pronasa 1X 5 min, Solución Pronasa 10 min, PBS 2 min, PFA 4 % en PBS 15 min (2x), PBS 2 min (2x), Solución acético anhidro 10 min, PBS 2 min, Solución salina 2 min, EtOH / Soluc. salina 30 % 2 min, EtOH / Soluc. salina 50 % 2 min, EtOH / Soluc. salina 70 % 2 min, EtOH / Soluc. salina 85 % 2 min, EtOH / Soluc. salina 95 % 2 min y EtOH 100 % 2 min. La composición del Tampón pronasa fue 50 mM Tris-HCl pH 7,5, 5 mM EDTA; la Solución pronasa: 0,32 % pronasa en tampón pronasa; PFA: paraformaldehído; Solución acético anhidro: 25 % trietanolamina, 0,5 % HCl concentrado [37%] y 0,5 % ácido acético anhidro).

Seguidamente, se realiza una primera prehibridación con una solución de 3 M NaCl, 0,1 M Tris-HCl pH 6,8, 0,1 M Na<sub>2</sub>HPO<sub>4</sub> pH 6,8 y 50 mM EDTA en formamida,

durante 10 min a 50°C. La segunda prehibridación consta de 30 min a la misma temperatura en solución de hibridación durante 30 min [12,5 % de la solución anterior, 50 % de formamida desionizada, 25 de dextrán sulfato (al 50 %), 1,25 % de tRNA (100 mg/ml), 1,25 % Denhardt (100X) y qsp (10 %) en agua milliQ].

Se desnaturalizan las sondas 5 min a 80 °C y se colocan en las muestras antes de añadir el tampón de hibridación. Se mantiene a 50°C un mínimo de 16 h en una cámara de humedad saturada con una mezcla 1/1 de 2XSSC/formamida.

#### ***M-6.9.3.3. Lavados de hibridación***

Para eliminar los restos de sonda no incorporada se incubaron las muestras en una solución al 50 % de formamida en 2XSSC durante 30 min a 50°C seguida de otra incubación en la misma solución de 1 h. Este proceso se repitió dos veces. Posteriormente se adicionó una solución de NTE (0,5 M NaCl, 10 mM Tris-HCl, 1 mM EDTA) y se incubó durante 10 min, seguida de otra en NTE con RNasa A (20 µg/ml) durante 30 min a 37°C, 10 min de nuevo en la solución NTE y una incubación final a 50°C en formamida/2XSSC al 50 % durante 60 min.

#### ***M-6.9.3.4. Inmunomarcaje y revelado de la reacción***

A temperatura ambiente: -1XSSC 2 min, TBS 5 min (2X), Blocking reagent (Roche) en TBS (0.5 g/100ml) 30 min, Blocking *bis*: 1% BSA, 0.3% Triton X100 in TBS 30 min, α anti-DIG (1:2000) en Blocking *bis* 90 min y TBS 5 min (4X). La composición del TBS fue: 0,1 M Tris-HCl pH 7,5, 0,15 M NaCl.

Seguidamente se lleva a cabo el revelado con la incubación de las muestras 5 min en el tampón de sustrato (0,1 M Tris-HCl pH 9,8, 0,1 M NaCl) y posterior tratamiento con el mismo tampón más 2 % MgCl<sub>2</sub> (2,5 M), 0,15 % dimetilformamida (100 mg/ml) y 0,33 % fosfatasa β-CIP (50mg/ml).

Se prepara una solución fresca de MgCl<sub>2</sub> 2,5 M y se deja la reacción en oscuridad hasta la aparición de la señal. Se para la reacción con un lavado de agua.

Para realizar la contra-coloración de las muestras para poder visualizar las estructuras nodulares se llevó a cabo una deshidratación progresiva de éstas en concentraciones decrecientes de EtOH (100 %, 70 % y 40 %), para posteriormente añadir una solución de azul Alcian al 1,3 % en ácido acético durante 1 s. Se lavó inmediatamente con agua milliQ 2 ó 3 veces. Para finalizar todo el proceso se cubrieron las muestras con un cubreobjetos con una gota de fijador DPX y se visualizaron a microscopía de campo claro en un microscopio Leica DMI6000.



## M-7. TÉCNICAS ANALÍTICAS

### M-7.1. DETERMINACIÓN DE LA ACTIVIDAD $\beta$ -GALACTOSIDASA EN CULTIVOS CELULARES

La actividad  $\beta$ -galactosidasa se determinó en cepas que contenían fusiones del gen o la región génica de interés con el gen informador *lacZ* siguiendo la metodología descrita por Sambrook *et al.*, 1989. Las cepas a estudio se cultivaron en medio rico (TY) con los antibióticos correspondientes hasta alcanzar la fase exponencial. En primer lugar, de los cultivos de estudio se tomaron alícuotas de 500  $\mu$ l y se anotó el valor de D.O<sub>600</sub>. Las células fueron aisladas mediante centrifugación a 12.000 r.p.m. durante 3 min. Como muestra control se tomaron 500  $\mu$ l del medio de cultivo que siguieron el mismo tratamiento que las muestras problema. Para la lisis celular se añadieron 500  $\mu$ l de Buffer Z [ $\text{Na}_2\text{HPO}_4 \times 12 \text{ H}_2\text{O}$  (0,06 M),  $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$  (0,04 M), KCl (0,01 M),  $\text{MgSO}_4 \times 7 \text{ H}_2\text{O}$  (0,001 M),  $\beta$ - mercaptoEtOH (0,05 M)], 20  $\mu$ l de cloroformo y 5  $\mu$ l 0.1 % SDS y se agitó mediante vórtex la mezcla durante 10 s. Las muestras se incubaron durante 10 min a 30 °C. La actividad  $\beta$ -galactosidasa comenzó tras añadir 200  $\mu$ l de una solución 4 mg/ml del sustrato orto-nitrofenil-  $\beta$ -D-galactopiranosido (ONPG) en tampón Z. La incubación de los extractos enzimáticos con el sustrato se hizo a 30 °C prologándose hasta la aparición de una coloración amarilla suficientemente intensa y estable. La reacción se detuvo por la adición de 500  $\mu$ l de una solución 1 M  $\text{Na}_2\text{CO}_3$ . Se anotó el tiempo transcurrido desde la adición de ONPG y se separaron los restos celulares por centrifugación durante 5 min a 8.000 r.p.m. El sobrenadante se recogió en un nuevo tubo y se determinó la absorbancia a 420 nm, sustrayendo en las muestras problema el valor obtenido para el control.

La actividad  $\beta$ -galactosidasa se expresó en unidades Miller (U) calculada mediante la siguiente fórmula:

$$U = 1000 \times \frac{A_{420}}{t \times V \times A_{600}}$$

Donde  $A_{420}$  y  $A_{600}$  son los valores de D.O. obtenidos a la longitud de onda de 420 nm y 600 nm, respectivamente. El valor t corresponde al tiempo de incubación en min mientras que V es el volumen de cultivo utilizado en ml.

# **RESULTADOS Y DISCUSIÓN**



# **CAPÍTULO 1**

**IDENTIFICACIÓN DE RNAs NO  
CODIFICANTES EN EL GENOMA DE  
*SINORHIZOBIUM MELILOTI***



## I-1. INTRODUCCIÓN

Las primeras evidencias de la expresión en procariotas de pequeños transcritos que no se traducen a proteína, diferentes de rRNAs y tRNAs y como éstos codificados en el cromosoma, datan de finales de la década de los 60 y comienzos de los 70 cuando se identificaron los RNAs 4.5S y 6S de *Escherichia coli*, mediante marcaje de RNA total y separación electroforética (Hindley, 1967, Griffin, 1971). Hasta 2001, la lista de nuevos sRNAs identificados aumentó escasamente hasta un total de diez, todos en la enterobacteria modelo (Gottesman, 2004). La mayoría de estos sRNAs fueron descubiertos de manera fortuita por ser particularmente abundantes en la célula, co-purificar con proteínas específicas o complementar fenotipos relevantes y sus respectivos *loci* genómicos mapean entre ORFs (regiones intergénicas; RIGs). Tres de ellos tienen funciones esenciales (RNasa P, tmRNA y 4.5S), dos tenían en aquel momento función desconocida (6S, Spot42) y del resto se sospechaba que eran reguladores (MicF, OxyS, DsrA, DicF y CsrB). Sin embargo, es la era post-genómica, con el impulso de la bioinformática y las tecnologías de alto rendimiento, la que marca la verdadera eclosión de la RNómica estructural y funcional en bacterias.

### I-1.1. METODOLOGÍAS PARA LA IDENTIFICACIÓN DE sRNAs EN GENOMAS BACTERIANOS

#### I-1.1.1. ABORDAJES BIOINFORMÁTICOS Y TÉCNICAS MOLECULARES CLÁSICAS

Entre los años 2001 y 2002 se publicaron cuatro estudios pioneros en la identificación sistemática de sRNAs en *E. coli*, fundamentados todos ellos en la comparación *in silico* de genomas de enterobacterias (Argaman *et al.*, 2001; Chen *et al.*, 2002, Rivas *et al.*, 2001, Wassarman *et al.*, 2001). En tres de estos estudios, las predicciones bioinformáticas se basaron en la coexistencia de al menos dos de las siguientes características en los genes sRNA candidatos: *i*) identidad de las secuencias nucleotídicas en especies filogenéticamente relacionadas, *ii*) localización en RIGs, y *iii*) presencia de señales de transcripción (promotores y terminadores) adecuadamente espaciadas (i.e. separadas 50-250 nt) en estas regiones (Argaman *et al.*, 2001; Chen *et al.*, 2002; Wassarman *et al.*, 2001). En el cuarto estudio el criterio predictivo esencial fue la conservación de la posible estructura secundaria de los sRNAs homólogos en los genomas bacterianos comparados (Rivas *et al.*, 2001). Este criterio está implementado en el programa QRNA que identifica en los alineamientos generados por las comparaciones genómicas patrones específicos de sustituciones nucleotídicas (mutaciones) compatibles con la conservación de una estructura

secundaria correspondiente a un RNA, más que con la conservación de secuencias codificantes de proteínas u otros elementos genómicos. Con posterioridad a la publicación de estos trabajos se han desarrollado tanto mejoras de estos primeros programas como nuevos algoritmos bioinformáticos que han sido utilizados con distinto éxito para la predicción a escala genómica de *loci* sRNA tanto en *E. coli* como en otras especies bacterianas (Tabei *et al.*, 2006, Do *et al.*, 2006, Liu *et al.*, 2005, Wang *et al.*, 2006, Pavesi *et al.*, 2004). Uno de estos nuevos programas, denominado RNAz, combina la conservación de la estructura con una estimación su estabilidad termodinámica para la predicción de sRNAs en bacterias (Washietl *et al.*, 2005). Sin embargo, las predicciones bioinformáticas no son definitivas para la identificación de un sRNA si no van acompañadas de una verificación experimental de la expresión del correspondiente transcrito. Por ejemplo, utilizando escrutinios *in silico* con diferentes criterios se han llegado a predecir más de 500 regiones genómicas en *E. coli* potencialmente codificantes de un sRNA, de los que poco más de una centena han sido confirmados experimentalmente (Livny and Waldor, 2007). Para ello los experimentos más informativos son la hibridación Northern después de la resolución del RNA bacteriano mediante técnicas electroforéticas y el mapeo de los extremos 5' y 3' de los transcritos detectados. La primera es una prueba definitiva de la expresión en la célula de los sRNAs predichos por algoritmos bioinformáticos y permite la estimación aproximada de sus tamaños así como la caracterización de sus perfiles de acumulación en diversas condiciones biológicas. La determinación de los extremos de las moléculas de RNA detectadas en los geles está basada en la transcripción reversa de las mismas, su amplificación con cebadores específicos y posterior secuenciación de los extremos de los cDNAs resultantes (RACE; Rapid Amplification of cDNA Ends). Es imprescindible para la determinación precisa del inicio de transcripción y de los tamaños de los sRNAs y, en consecuencia, para la adecuada anotación de estos nuevos genes en los genomas bacterianos (Altuvia and Wagner, 2000).

#### I-1.1.1. TECNOLOGÍAS DE ALTO RENDIMIENTO: RNÓMICA

El desarrollo de la genómica funcional de alto rendimiento ha puesto a disposición de los investigadores nuevas tecnologías que han sido también aprovechadas para la identificación experimental de sRNAs expresados por las bacterias en diversas condiciones biológicas sin la necesidad de su predicción bioinformática previa (RNómica propiamente dicha). La primera de estas alternativas experimentales utilizada en *E. coli* fue la hibridación de *microarrays* de alta densidad (*high-density* o *tilling microarrays*), que incluyen sondas de ambas hebras del DNA tanto para los ORFs como para las RIGs de este genoma (Wassarman *et al.*, 2001; Tjaden *et al.*, 2002; Zhang *et al.*, 2003). No obstante, la interpretación de las señales de

hibridación en las RIGs, en la proximidad de ORFs o en la cadena complementaria de secuencias codificantes de proteínas y su asignación a una determinada categoría de sRNA es compleja y habitualmente imprecisa, requiriendo de verificaciones independientes mediante técnicas moleculares clásicas como las descritas (Wassarman *et al.*, 2001). Esta metodología es, por tanto, menos definitiva que Northern y RACE en la primera caracterización molecular de los sRNAs pero aporta datos globales sobre los perfiles de expresión de éstos como primera aproximación a su función biológica (Sharma and Vogel, 2009). Una segunda alternativa de alto rendimiento, también utilizada inicialmente para el estudio de la transcripción del genoma de *E. coli*, consiste en la reverso transcripción del RNA, la posterior secuenciación de las moléculas de cDNA generadas y su mapeo final en el genoma de referencia. Los primeros estudios de este tipo requirieron la amplificación del cDNA y el clonaje masivo de los productos de PCR (*shotgun cloning*) previos a la secuenciación de un número amplio pero, en cualquier caso limitado, de moléculas mediante el clásico método Sanger (Kawano *et al.*, 2005, Vogel *et al.*, 2003). La Fig. I-1 recoge la evolución en la identificación de sRNAs bacterianos desde 1966 hasta 2006 mediante la combinación de predicción bioinformática y validación experimental y secuenciación Sanger de bancos de cDNA (Livny and Waldor, 2007). Sin embargo, el desarrollo más reciente de nuevas tecnologías de secuenciación masiva (NGS, Next Generation Sequencing Technologies) en cualquiera de sus plataformas, Roche/454 o Illumina/Solexa, ha multiplicado en varios órdenes de magnitud el número de secuencias que pueden ser obtenidas, sin necesidad del clonaje previo de las moléculas cDNA, en muy poco tiempo y a costes razonables (Metzker, 2010). Esta aproximación para el estudio del transcriptoma, conocida como RNA-Seq, aúna las ventajas del mapeo RACE y de la hibridación de *microarrays* pues a la par que permite determinar de forma precisa los extremos de cada transcrito presente en la muestra posibilita la cuantificación de su abundancia relativa en la misma. En la Tabla I-1 se relacionan las bacterias cuyo RNoma ha sido investigado mediante tecnologías de alto rendimiento hasta comienzos de 2010 y la metodología empleada en cada caso.

Cualquiera que sea la tecnología de alto rendimiento empleada para la caracterización del RNoma bacteriano se suelen utilizar dos subpoblaciones diferentes de transcritos de partida, supuestamente enriquecidas en sRNAs, para generar los bancos de cDNA. Éstas son las moléculas de RNA de pequeño tamaño (50-350 nt) y las que se asocian a la chaperona de RNA Hfq, aisladas por co-inmunoprecipitación (CoIP) con esta proteína. Zhang y colaboradores publicaron en el año 2003 un estudio pionero que combinaba con éxito la CoIP de RNAs con Hfq utilizando anticuerpos específicos para esta proteína con el análisis del CoIP-RNA



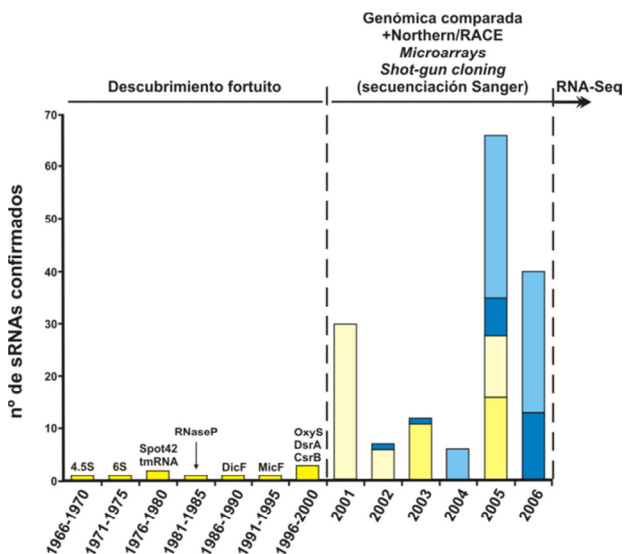


Fig. I.1. Evolución de la identificación de sRNAs desde 1966 hasta 2006. Las barras amarillas indican los sRNAs identificados en *E. coli* mediante abordajes bioinformáticos (amarillo oscuro) o experimentales (amarillo claro). Las barras azules indican el número de los sRNAs físicamente confirmados en otras especies por métodos experimentales (azul oscuro) o bioinformáticos (azul claro). Adaptado de Livny and Waldor, 2007.

mediante hibridación de *microarrays* de alta densidad para la identificación de cientos de sRNAs en *E. coli*. Sin embargo, la necesidad de disponer tanto de *microarrays* de alta densidad como de anticuerpos anti-Hfq específicos de la especie bacteriana en estudio ha condicionado la utilización de este protocolo para la caracterización del RNoma de bacterias diversas. Solventando estas limitaciones, Sittka y colaboradores (2008) modificaron el gen cromosómico que codifica Hfq en *S. enterica* serovar Typhimurium para expresar una proteína con una extensión en su extremo C-terminal (FLAG) reconocida por un anticuerpo monoclonal comercial y analizaron por secuenciación en plataforma 454 los transcritos co-inmunoprecipitados (CoIP-RNA) con la proteína modificada (Hfq<sup>FLAG</sup>). Este abordaje resultó en el aislamiento con alto grado de especificidad de sRNAs asociados a Hfq y permitió el mapeo de los mismos en el genoma con una resolución inalcanzable para otras tecnologías. En consecuencia, la RNA-Seq de RNA total o, preferentemente de la subpoblación de transcritos de pequeño tamaño o del RNA asociado a Hfq aislado con anticuerpos anti-FLAG comienzan a imponerse como abordajes genéricos de elección para la caracterización de la fracción no codificante de los genomas de un número creciente de especies bacterianas. Esta tecnología permite además la adecuada catalogación de estos transcritos en sus diversos tipos; regiones *cis*-reguladoras de mRNAs (*riboswitches* y otras regiones 5'-UTR), *trans*-sRNAs, *cis*-sRNAs y RNAs antisentido de regiones codificantes de proteínas, anticipando así una diversidad y plasticidad

funcional del RNA, imprevistas hace escasamente cinco años por la genómica comparada, que debe ser explorada en el futuro inmediato.

En este capítulo se describe y discute con detalle nuestra contribución a la caracterización del RNoma de *S. meliloti* utilizando abordajes bioinformáticos, experimentos Northern y RACE y RNA-Seq de la fracción de transcritos asociados a la proteína Hfq.

**Tabla I-1. RNomas bacterianos y tecnología utilizada para su caracterización<sup>1</sup>**

Bacteria	Componente experimental	Referencia
<b>Bacteria Gram-negativa</b>		
<i>Escherichia coli</i>	Arrays de oligonucleótidos	Wassarman <i>et al.</i> , 2001
<i>Escherichia coli</i>	RNómica	Vogel <i>et al.</i> , 2003
<i>Escherichia coli</i>	Co-immunoprecipitación Hfq, <i>microarrays</i>	Zhang <i>et al.</i> , 2003
<i>Escherichia coli</i>	Clonaje directo	Kawano <i>et al.</i> , 2005
<i>Salmonella typhimurium</i>	Co-immunoprecipitación Hfq, RNA-Seq	Sittka <i>et al.</i> , 2008, 2009
<i>Prochlorococcus marinus</i>	Microarrays	Steglich <i>et al.</i> , 2008
<i>Vibrio cholerae</i>	RNA-Seq	Liu <i>et al.</i> , 2009
<i>Aquifex aeolicus</i>	RNómica	Willkomm <i>et al.</i> , 2005
<i>Pseudomonas aeruginosa</i>	RNómica	Sonnleitner <i>et al.</i> , 2008
<i>Sinorhizobium meliloti</i>	RNA-Seq	Schlüter <i>et al.</i> , 2010
<i>Caulobacter crescentus</i>	Tiling array	Landt <i>et al.</i> , 2008
<i>Burkholderia cenocepacia</i>	RNA-Seq	Yoder-Himes <i>et al.</i> , 2009
<i>Francisella tularensis</i>	RNómica	Postic <i>et al.</i> , 2010
<i>Rhodobacter sphaeroides</i>	Co-immunoprecipitación Hfq, RNA-Seq	Berghoff <i>et al.</i> , 2011
<b>Bacteria Gram-positiva</b>		
<i>Bacillus subtilis</i>	Microarrays	Silvaggi <i>et al.</i> , 2006
<i>Bacillus anthracis</i>	RNA-Seq	Passalacqua <i>et al.</i> , 2009
<i>Listeria monocytogenes</i>	Co-immunoprecipitación Hfq, RNA-Seq	Christiansen <i>et al.</i> , 2006
<i>Listeria monocytogenes</i>	Tiling array	Toledo-Arana <i>et al.</i> , 2009
<i>Streptomyces coelicolor</i>	RNómica	Vockenhuber <i>et al.</i> , 2011
<i>Mycobacterium tuberculosis</i>	RNA-Seq	Arnvig <i>et al.</i> , 2011

<sup>1</sup> Actualizada de Sharma and Vogel, 2009.

En este capítulo se describe con detalle nuestra contribución a la caracterización del RNoma de *S. meliloti* utilizando abordajes bioinformáticos, experimentos Northern, RACE y RNA-Seq de la fracción de transcritos asociados a la proteína Hfq.

Durante la realización de este trabajo, se publicaron varios estudios que añadieron nuevos sRNAs al genoma de *S. meliloti*, utilizando otras aproximaciones bioinformáticas (Valverde *et al.*, 2008), hibridación de *microarrays* (Ulve *et al.*, 2007) o RNA-Seq (Schlüter *et al.*, 2010), cuyos resultados serán discutidos también en este capítulo.

## I-2. DISEÑO EXPERIMENTAL

### I-2.1. ANÁLISIS BIOINFORMÁTICO

Esta parte del trabajo se realizó en colaboración con las Dras. Elena Rivas (Howard Hughes Medical Institute) y Coral del Val (Universidad de Granada)

#### I-2.1.1. SECUENCIAS GENÓMICAS, EXTRACCIÓN DE RIGs

Las secuencias y anotaciones de los siguientes genomas de  $\alpha$ -proteobacterias fueron descargados de la base de datos NCBI (National Center for Biotechnology Information) (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>): *S. meliloti* 1021, *Mesorhizobium loti* MAFF303099, *Bradyrhizobium japonicum* USDA110, *Agrobacterium tumefaciens* C58 (versions Cereon and WashU), *Brucella melitensis* 16 M, *B. suis* 1330, *Caulobacter crescentus* CB15, *Rickettsia conorii* Malish7, y *R. prowazekii* Madrid E. Los genomas de los simbioses de leguminosas *Rhizobium etli*, *R. leguminosarum* bv. viceae y *S. medicae* no estaban disponibles en el momento del análisis computacional, por lo que no se incluyeron en el fichero inicial.

Para generar un archivo con la secuencia de las RIGs, los genes anotados de *S. meliloti* en la base de datos (<http://bioinfo.genopoletoulouse.prd.fr/annotation/iANT/bacteria/rhime/>), se extrajeron desde la secuencia genómica completa. Los genes conocidos fueron agrupados en diferentes categorías: 6.309 genes codificantes de proteínas (ORFs y secuencias de inserción), 66 RNAs (54 tRNAs, 3 copias del operón rRNA [5S, 16S y 16S], tmRNA y 2 copias del intrón del grupo II RmInt1 asociado a la secuencia de inserción IS*Rm2011-2*), 374 elementos repetidos (RIMES y secuencias Sm), y una miscelánea de 267 secuencias (motivos y otras secuencias). Tres genes codificantes de los RNAs SRP (4.5S), RNasa P e IncA no se encontraban anotados en el momento del análisis y se mantuvieron en el fichero de las RIGs como controles positivos.

#### I-2.1.2. ANÁLISIS CON LOS PROGRAMAS eQRNA Y RNAz

El análisis mediante QRNA fue realizado utilizando la versión del programa eQRNA 2.0.3c. (<ftp://selab.janelia.org/pub/software/qrna/>). eQRNA usa tres modelos probabilísticos diferentes para examinar el patrón de mutación en un alineamiento de secuencias nucleotídicas. Distingue entre un patrón de sustitución relacionado con la estructura secundaria de un RNA (RNA), un patrón correspondiente a un RNA mensajero codificante (COD), y otro de evolución independiente (OT) (Fig. I-2). El alineamiento es después clasificado como RNA, secuencia codificante u otra secuencia (i.e. no asignable a otros grupos), según la probabilidad Bayesiana posterior de cada modelo. La nueva versión de eQRNA

utilizada en este estudio introduce un modelo evolutivo probabilístico explícito, que incorpora además inserciones y deleciones (Rivas, 2005). Para cada comparación, eQRNA selecciona los parámetros del modelo en base al grado de divergencia observado en el alineamiento. Además, para las predicciones también se utilizó el software RNAz, versión 0.1.1 (<http://www.tbi.univie.ac.at/~wash/RNAz>) (Washietl *et al.*, 2005).

Los alineamientos generados fueron analizados conjuntamente mediante eQRNA y RNAz, utilizando un tamaño de ventana de 150 nucleótidos con desplazamientos sucesivos de 50 nt. Las parejas de alineamientos fueron clasificadas como RNA si se obtenía una puntuación de 3,5 bits o más. Esta puntuación representa la probabilidad de los datos se ajusten mejor al modelo de RNA que a la combinación de los modelos COD y OT. Es medida en bits y definida como el logaritmo en base dos de la relación entre la probabilidad del modelo RNA y de las probabilidades de los modelos COD y OT.

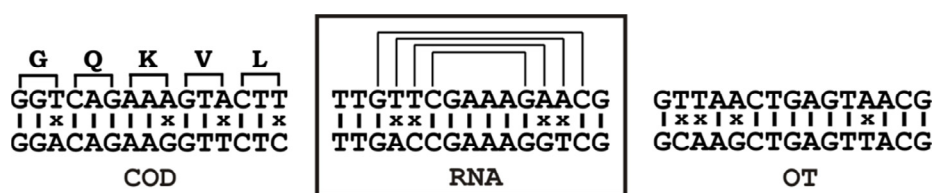


Fig. I-2. Modelos probabilísticos en los que se basa eQRNA. RNA, las mutaciones se ajustan a modelos de covarianza de bases que conservan la estructura secundaria; COD, mutaciones en mRNAs localizadas preferentemente en el tercer nucleótido de cada codón; OT, patrón de mutaciones aleatorio que no se ajusta a ninguno de los otros dos modelos.

eQRNA seleccionó ventanas (i.e. secuencias de 150 nt) de estas parejas de alineamientos con una puntuación RNA de 3,5 bits o más. Para RNAz, se seleccionaron ventanas con una probabilidad RNA de 0,95 o más. El subconjunto de secuencias (sin especificidad de cadena), predicha por ambos métodos como correspondiente a un RNA, se sometió a análisis posteriores. Los valores de corte para cada programa fueron seleccionados por la imposición de tener como mucho 10 falsos positivos en las comparaciones intergénicas entre los genomas. Para eQRNA estimamos la forma de la distribución de las puntuaciones de las ventanas aleatorias, por generación de gran número de ventanas mezcladas desde RIGs. El 10% de la distribución de las puntuaciones de los alineamientos aleatorios correspondientes a las RIGs fueron validadas según una distribución de Gumbel ( $l = 0,6123$  y  $m = -5,7247$ ). Los valores (logaritmos pares) asignados como RNA por debajo de 3.5bits fueron inferidos para obtener la mayoría de los falsos positivos del método (que contiene 2.289 ventanas). Para RNAz, la puntuación de las ventanas barajadas

aleatoriamente no siguió una distribución conocida. Para ello, determinamos experimentalmente que la probabilidad RNA límite es 0,95, dado que reproduce fielmente el número de los falsos positivos estimados en todo el método.

### I-2.1.2. MODELOS DE COVARIANZA

Para el análisis de la distribución de los genes *smr* y la búsqueda de homólogos en otros grupos bacterianos se llevaron a cabo, en colaboración con la Dra. Coral del Val, análisis mediante modelos de covarianza. En un primer paso, las secuencias de los genes *smr* se compararon con las bases de datos mediante BLASTN y contra todos los genomas disponibles (1.615 secuencias en el 20 Abril 2011; <http://www.ncbi.nlm.nih.gov>). Las regiones que mostraban homologías significativas (78 - 79 % de similaridad) se utilizaron para generar alineamientos automáticos en el programa **Infernal** (Nawrocki *et al.*, 2009). Estos alineamientos iniciales fueron inspeccionados manualmente para deducir la estructura consenso de cada familia. La estructura consenso fue también predicha por el programa locARNATE (Will *et al.*, 2007), que presenta una manera automática de relacionar estructuras secundarias conservadas. Los alineamientos ya curados de los homólogos más cercanos fueron introducidos de nuevo en Infernal para interrogar al mismo set de genomas bacterianos, buscando nuevas secuencias que se ajustaron a los modelos. El proceso de alineamiento se repitió durante tres iteraciones. Los candidatos obtenidos con los modelos de Infernal se seleccionaron como miembros de una familia de acuerdo a su *E-value*, siendo condición que fuera de un valor  $e10^{-03}$  o menor, o después de una inspección manual para los *E-values* mayores correspondientes a Infernal.

## **I-2.2. MAPEO DE EXTREMOS 5' Y 3' DE TRANSCRITOS MEDIANTE AMPLIFICACIÓN (RACE)**

### I-2.2.1. RACE 5'

Este protocolo se utilizó para la determinación del sitio de inicio de transcripción de los transcritos *Smr* de *S. meliloti* 1021. Se utilizaron algunos componentes del kit comercial *First Choice® RLM-RACE* (Ambion) siguiendo las instrucciones del proveedor. Este procedimiento se basa en la transformación de los extremos 5' trifosfato de los transcritos primarios en 5' monofosfato por tratamiento del RNA total con la pirofosfatasa ácida de tabaco (TAP), lo que permite la ligación posterior de un adaptador RNA a estos extremos.

La defosforilación con TAP se llevó a cabo en la siguiente reacción: 15  $\mu$ g RNA+ 5U de TAP (1  $\mu$ l), en un  $V_F$  de 10  $\mu$ l. Paralelamente, se llevaron a cabo reacciones

control sin tratamiento con TAP. Posteriormente el adaptador RNA (5A, 5'-GCU GAU GGC GAU GAA CAC UGC GUU UGC UGG CUU UGA AA-3') se ligó al extremo 5' de los transcritos mediante el uso de una T4 RNA ligasa (New England Biolabs) (1  $\mu$ l de la reacción anterior en un  $V_F$  de 10  $\mu$ l). Tras 1 h a 37°C, se tomó un quinto de la reacción de ligación y se sometió a reverso transcripción con *random hexamers* (Ambion). Aunque el kit recomienda el uso de la enzima para la RT M-MLV RT, comprobamos empíricamente que el uso de *Thermoscript*<sup>TM</sup> RT (Invitrogen) es más apropiado para estos experimentos, ya que, al tener una temperatura de polimerización superior, soslaya mejor las estructuras secundarias del RNA (del Val *et al*, 2007). Las condiciones de incubación fueron modificadas en función de las condiciones aconsejadas por el proveedor de la enzima: 20 min a 55°C, 20 min a 60°C, 20 min a 65°C y 5 min a 85°C. Los productos de la reacción de RT se sometieron a la acción de la RNasa H (Invitrogen) que elimina las moléculas de RNA en los híbridos RNA-DNA. Posteriormente se amplificó el cDNA generado mediante 2 rondas sucesivas de PCR, en las que la segunda PCR sitúa los cebadores en posiciones interiores respecto a la primera (PCR anidada). Las parejas de cebadores incluyen, en ambos casos, un cebador que hibrida en el adaptador (proporcionado en el kit) y un cebador complementario a una secuencia de la región 5' de cada gen, diseñado de acuerdo a la predicción bioinformática (ver anexo tablas de oligonucleótidos). Los fragmentos generados se resolvieron tanto en geles de agarosa como en geles de poliacrilamida no desnaturalizantes. Las bandas correspondientes a los productos mayoritarios se aislaron de gel, fueron clonados en pGEM-T *easy* (Promega, Ap<sup>R</sup>) y los plásmidos con los insertos de interés, seleccionados por PCR de colonia y digestión, fueron secuenciados.

#### I-2.2.2. RACE 3'

Utilizamos este protocolo para la determinación del extremo 3' del transcrito Smr22C (6S). Las muestras (20  $\mu$ g de RNA en un  $V_F$  = 20  $\mu$ l) fueron sometidas a un tratamiento con 3  $\mu$ l de fosfatasa alcalina (CIP, 1U/ $\mu$ L) durante 1 h a 37°C para eliminar los grupos 5' monofosfato de las moléculas de RNA, de manera que no puedan ligarse con el adaptador en 3'. Un adaptador (0,3  $\mu$ g) diseñado con este fin (3A, 5'-P-UUC ACU GUU CUU AGC GGC CGC AUG CUC-idT-3') se incubó con 5  $\mu$ l del RNA previamente tratado en presencia de 10 U de T4 RNA ligasa (New England Biolabs) durante 1 h a 37°C. La RT se realizó de forma idéntica a la descrita para el RACE 5' pero con un cebador complementario a la secuencia del adaptador. De forma similar, la PCR anidada se realizó con parejas de cebadores complementarias a la secuencia del adaptador y a la región 3' del cDNA, según la predicción bioinformática.

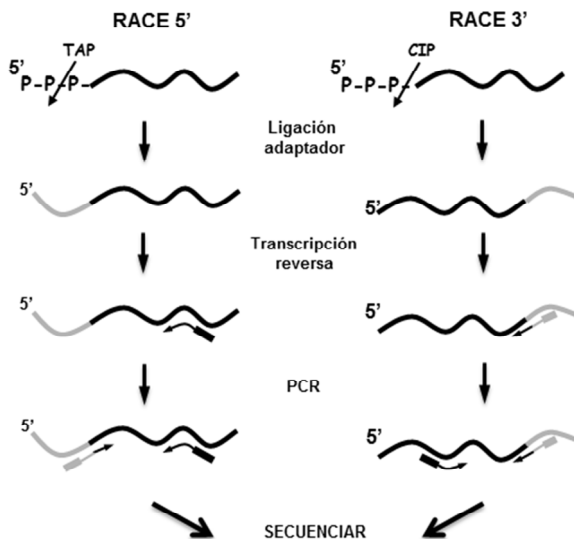


Fig. I-3. Pasos fundamentales en la técnica de mapeo de transcritos mediante RACE 5' y RACE 3'.

### I-2.3. CONSTRUCCIÓN DE LA CEPA *S. MELILOTI* 2011 B23001 $hfq^{FLAG}$

El protocolo que se siguió para modificar la copia cromosómica del gen *hfq* de la cepa *S. meliloti* 2011 B23001, y expresar la variante de la proteína marcada en su extremo C-terminal con el epítipo FLAG, se describe con detalle a continuación y se esquematiza en la Fig. I-4.

Un fragmento de DNA de doble cadena correspondiente al triple tándem del epítipo FLAG (3XFLAG; Sigma-Aldrich) se generó por anillamiento de los oligonucleótidos 3xFlag y 3xFlag-i, ambos de una longitud de 69 nucleótidos. Los extremos de estos cebadores fueron específicamente diseñados para ser complementarios a los extremos 5' protuberantes correspondientes a las dianas de restricción *XbaI* y *HindIII*. En fragmento de DNA resultante fue insertado en los mismos sitios de restricción del vector pBluescript II KS+, dando lugar a pKS3XFlag. Paralelamente, se amplificó, utilizando DNA genómico de *S. meliloti* 1021 como molde, una región de 1684 pb, formada por el ORF *hfq* de 243 nt y las regiones flanqueantes (714 pb hacia arriba y 727 pb hacia abajo del gen), con una polimerasa corregidora de errores *Pfu*, utilizando los cebadores Hfq\_Fw/Hfq\_Rv (consultar tabla

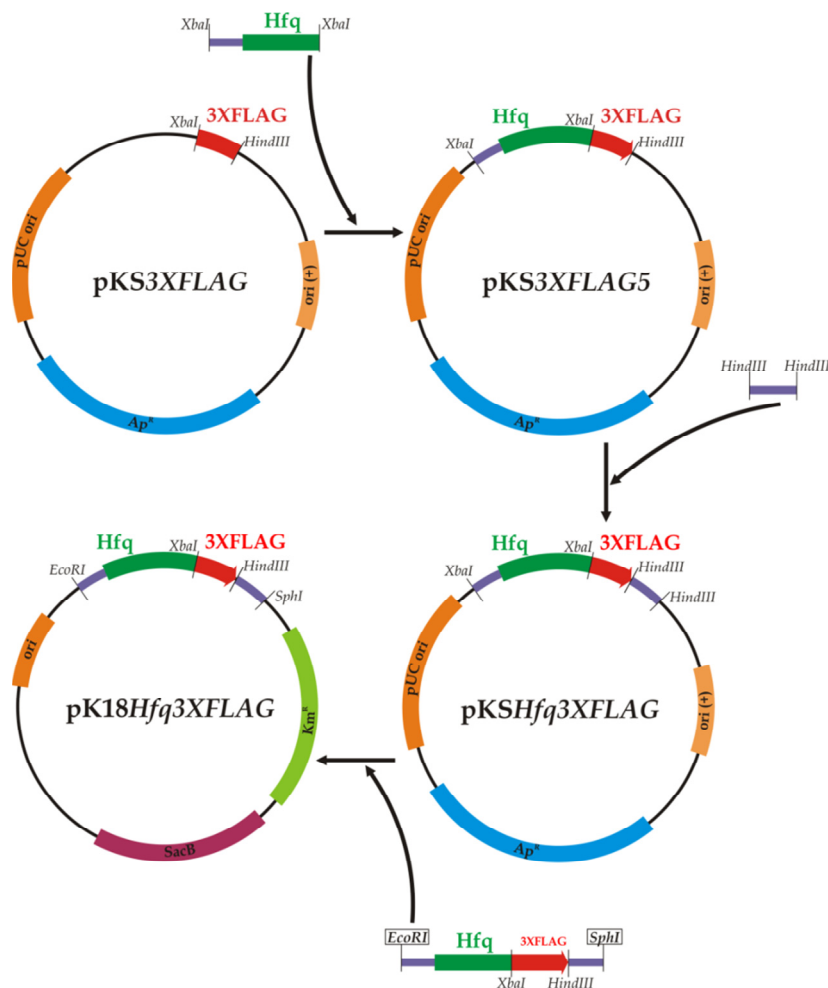


Fig. I-4. Construcciones más relevantes en el proceso de creación del mutante *S. meliloti hfq*<sup>FLAG</sup>

de oligonucleótidos situada en el anexo final de este capítulo). Este fragmento de DNA se insertó en el plásmido pGEM-T Easy, generando pGEM $hfq$ . La secuencia completa de la proteína Hfq (sin el codón de terminación, TGA), además de la región de 655 nt adyacente al codón ATG del gen  $hfq$ , fue amplificada desde el vector pGEM $hfq$ , utilizando la enzima *Phusion*<sup>®</sup> DNA polimerase (NEB), con la pareja de oligonucleótidos 5HfqTag/3HfqTag, ambos con el sitio de restricción *XbaI* en su extremo 5'. El producto de PCR resultante fue clonado en pGEM-T Easy Vector System (Promega), digerido como fragmento *XbaI*, recuperado desde gel y clonado en pKS3XFlag, dando lugar a pKS3XFlag5. Un segundo fragmento de PCR de 873 nt conteniendo el codón de terminación para la traducción de la proteína marcada con el epítipo FLAG, fue generado por PCR de la región 3' del gen  $hfq$  desde la construcción pGEM $hfq$ , utilizando la pareja de oligos 5FlxTag/3FlxTag, que



incorporar el sitio de restricción *HindIII* en ambos extremos del amplicón resultante. El mismo producto fue insertado en pGEM-T Easy recuperado como fragmento *HindIII*, purificado de gel y clonado finalmente en el sitio *HindIII* del plásmido pKS3XFlag5, dando lugar a pKSHfq<sup>3XFLAG</sup>. Este plásmido fue usado como molde para amplificar la región de 1.839 pb con los oligos 5HfqTag/3HfqTag, en donde los sitios *XbaI* y *HindIII* fueron reemplazados por los sitios *EcoRI* y *SphI*, respectivamente. El producto de PCR resultante se clonó en pGEM-T Easy para obtener pGEMHfq<sup>3XFLAG</sup> y posteriormente el fragmento proveniente de la digestión *EcoRI/SphI* del mismo plásmido fue clonado en el polilinker del pK18mobsacB, dando lugar a pK18Hfq<sup>3XFLAG</sup>. Este plásmido fue movilizado mediante cruce triparental a la cepa salvaje *S. meliloti* 2011 B23001 para reemplazar el gen *hfq* por el alelo modificado. Seis de 18 colonias de los transconjugantes finales se analizaron mediante PCR de colonia para, finalmente verificar dos de ellas por Western-Blot con anticuerpos monoclonales Anti-Flag (Sigma-Aldrich).

Todas las construcciones plasmídicas que requirieron amplificaciones previas mediante PCR fueron verificadas por secuenciación. Los correctos reordenamientos genómicos en la cepa de *S. meliloti* derivada se testaron mediante Southern-Blot del DNA genómico digerido con la enzima *HindIII*, con posterior hibridación con una sonda marcada radiactivamente utilizando el protocolo estándar.

## I-2.4. SECUENCIACIÓN MASIVA (RNA-Seq) DE LA SUBPOBLACIÓN DE TRANSCRITOS ASOCIADOS A Hfq<sup>FLAG</sup> (CoIP-RNA)

### I-2.4.1. PREPARACIÓN DEL CoIP-RNA

Las cepas *S. meliloti* 2011 B23001 (control) y su derivada *S. meliloti* 2011 B23001/*hfq*<sup>FLAG</sup> y se cultivaron en 200 ml de medio TY hasta DO 0.6 (fase logarítmica de crecimiento) o 2.4 (fase estacionaria). Las células se recogieron mediante centrifugación 10 min a 6.000xg y a 4°C. El pellet se lavó dos veces con 24 ml de sarcosil 0.1% en TE y posteriormente 2 veces más con 50 ml de PBS frío. Las células se resuspendieron finalmente en 8 ml de Lysis Buffer (kit) + 0,2 mM de fluoruro de fenilmetilsulfonilo (PMSF) y se incubaron con agitación a 4°C. Para aumentar la eficiencia de lisis, las suspensiones bacterianas se sonicaron 3 veces durante 15-20 s en el sonicador *Branson Sonifier*. Los lisados se centrifugaron a 12.000xg durante 10 min a 4°C. El sobrenadante (**fracción celular soluble**) se puso en contacto con 40 µL de la resina proporcionada por el kit (contiene el anticuerpo anti-FLAG) previamente lavada según las instrucciones del fabricante y se incubó en agitación a 4°C durante toda la noche. Al día siguiente, se centrifugaron las muestras durante 1 min a 4°C y 8.200xg. El sobrenadante (**fracción no retenida por el anticuerpo anti-FLAG**) se

retiró para utilizarlo posteriormente para Wester-blot. Se recogió el precipitado (**fracción unida al anticuerpo anti-FLAG**) y se pasó a columnas *Sigma-prep™ spin columns*, previamente enfriadas en hielo (referencia SC1000-1KT). Se centrifugó durante 1 min a 8.200xg y a 4°C. Posteriormente, se lavó la resina retenida en la columna 3 veces con 500 µL de Wash Buffer 1X (kit) centrifugando cada vez 1 min a 8.200xg a 4°C. Para la elución de los complejos RNA-proteína, la resina se resuspendió en 3 µL de solución del péptido a 30 µg/µL y se incubó 30 min a 4°C, agitando suavemente de forma periódica. Finalmente la columna se destapó y se centrifugó 30 s a 8.200xg y a 4°C. El eluido, conteniendo los complejos Hfq<sup>FLAG</sup>-RNA, se trató con 5 U de DNasaI, se sometió a extracción con fenol-cloroformo y el CoIP-RNA se resuspendió en 10 µL de agua y se cuantificó en Nanodrop. El procedimiento fue monitorizado mediante Western-blot de todas las fracciones proteicas (lisado, fracción soluble, no retenida, retenida, etc) utilizando anticuerpos anti-FLAG.

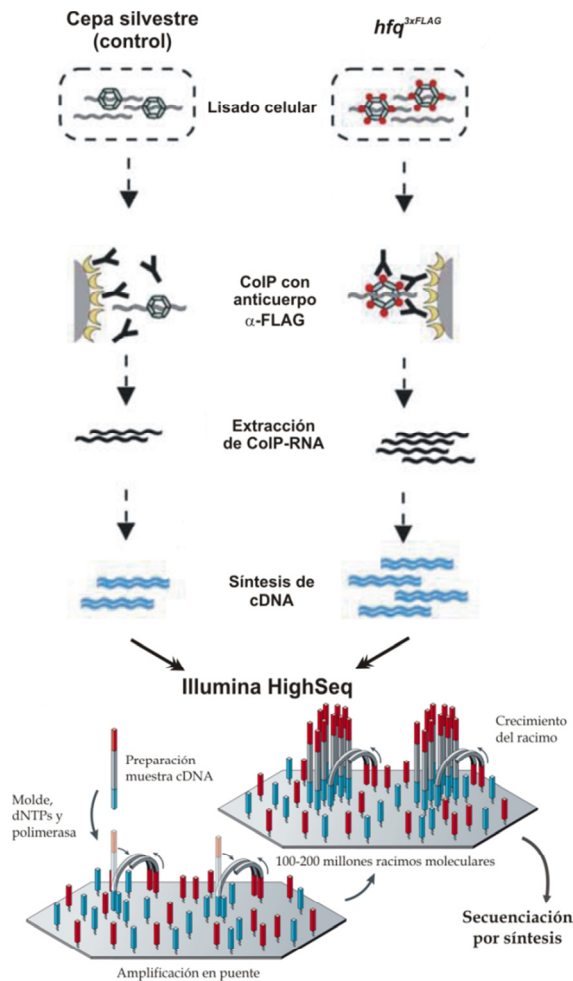
Como primera valoración del enriquecimiento del CoIP-RNA en sRNAs, así como de la presencia o ausencia de RNAs ribosómicos, se tomó 1 µL de la muestra y se sometió a reverso-transcripción (ver apartado de materiales y métodos) utilizando 10 U de SuperScript II (Invitrogen) y posterior amplificación mediante PCR con oligos de genes control (ver tabla de oligos).

#### I-2.4.2. RNA-Seq Y ANÁLISIS INFORMÁTICO

##### ***II.2.4.2.1. Preparación de muestras y secuenciación mediante Illumina/Solexa***

Las muestras de CoIP-RNA se obtuvieron a partir de dos cultivos independientes de la cepa control (*S. meliloti* 2011 B23001) y problema (*S. meliloti* 2011 B23001/hfq<sup>FLAG</sup>) en medio completo TY hasta las fases logarítmica y estacionaria de crecimiento. Previamente a la secuenciación los 4 CoIP-RNA control (2 en fase logarítmica y 2 en fase estacionaria) se mezclaron, al igual que los dos CoIP-RNA problema de cada fase de crecimiento para la generación de 3 bancos de cDNA, control, Hfq-TYLog y Hfq-TYEst.

El proceso general de análisis de la subpoblación de transcritos que se asocian a Hfq (esquemático en Fig. I-5) finaliza con la secuenciación del CoIP-RNA mediante Illumina/Solexa. Este proceso se basa en la generación de multitud de islas moleculares, conocidos como “polonias” (colonias generadas por la polimerasa), que



**Fig. I-5. Diagrama del esquema de trabajo de obtención y secuenciación del CoIP-RNA.** Los bancos de transcritos correspondientes a la cepa control y *hfq*<sup>FLAG</sup> se reverso-transcriben y el cDNA resultante se somete a secuenciación Illumina/Solexa. Ésta se basa en una amplificación en fase sólida, está compuesta por dos pasos básicos: unión inicial de cebadores y extensión de la molécula del cDNA (proveniente de un molde único), amplificación en puente inmediata del molde inmovilizado con oligonucleótidos adyacentes y secuenciación por síntesis.

son secuenciadas simultáneamente. Estas reacciones ocurren en la superficie de un “flujo celular” (básicamente la superficie de un portaobjetos especial hidratado), que provee una gran superficie para millones de reacciones químicas paralelas. Este proceso consta de varios pasos:

**i) Preparación de la muestra:** La muestra de cDNA se fragmenta en un tamaño apropiado (800 nt) utilizando aire comprimido en un sistema conocido como nebulizador. Los extremos del cDNA se ligan a dos adaptadores únicos.

**ii) Generación de las islas por amplificación en puente:** al contrario que los métodos por 454, Illumina se basa en una PCR por emulsión para generar las colonias. Esta técnica utiliza una amplificación única en puente, que ocurre en la superficie del soporte sólido o “célula flujo”. Este soporte está recubierto por adaptadores de cadena sencilla, correspondientes a los adaptadores ligados al cDNA tratado en el paso anterior, y que se unen entre sí en la reacción de extensión de la PCR. Ciclos consecutivos de desnaturalización y extensión generan en las moléculas de cDNA molde un conjunto de amplificadas que provienen de ésta, y que se conocen como islas o racimos.

**iii) Secuenciación por síntesis:** estos millones de racimos únicos se introducen en un secuenciador, que va incorporando nucleótidos marcados fluorescentemente con lectura simultánea mediante un láser, de las moléculas generadas.

#### ***I-2.4.2.2. Análisis bioinformático***

Los ficheros con las lecturas correspondientes a la secuenciación de las condiciones testadas primeramente fueron mapeados con la base de datos GenDB (<http://www.rhizogate.de>) utilizando el software SEGEMEHL (<http://www.bioinf.uni-leipzig.de/Software/segemehl>) con los parámetros: *extension penalty* = 2, *differences* = 1, *accuracy* = 90. Las lecturas que no se identificaron en la base de datos se sometieron a un filtrado (*trimming*) para eliminar secuencias poli-A internas y de baja calidad. Las lecturas generadas después de este paso se volvieron a mapear en el genoma de *S. meliloti* 1021. Del conjunto de lecturas mapeadas se eliminaron las que mapearon en *loci* múltiples (e.g. secuencias repetidas), dando resultado a un fichero que contiene el conjunto de las lecturas que mapearon en sitios únicos del genoma. La normalización de la posición correspondiente a cada nucleótido se calculó como el resultado de la cobertura en número de lecturas que mapean en él, multiplicado por el factor 1.000.000 (número medio de lecturas que mapean en sitio único en todas las condiciones de estudio) y dividido entre el número total de lecturas de esta misma índole. Las regiones identificadas se establecieron con un tamaño mínimo de 20 nt y una cobertura mínima de 10 lecturas. Finalmente, se generaron listas de regiones candidatas con un factor de enriquecimiento (calculado como el cociente entre la cobertura normalizada de la cepa Hfq<sup>FLAG</sup> y la cobertura normalizada de la cepa control, siendo siempre su valor mayor que cero) por encima de 5. El filtrado de las secuencias, la normalización, y la clasificación de las regiones obtenidas se realizaron mediante programas automatizados (*scripts*) desarrollados por Jan Reinkensmeier (Facultad de

Tecnología, Universidad de Bielefeld, Alemania), específicamente para estas muestras.

### I-3. RESULTADOS

#### I-3.1. IDENTIFICACIÓN DE NUEVOS sRNAs EN *S. MELILOTI* MEDIANTE GENÓMICA COMPARADA

##### I-3.1.1. ANÁLISIS BIOINFORMÁTICO

Este trabajo de investigación abordó, en primer lugar, la identificación de nuevos *trans*-sRNAs en el genoma de *S. meliloti* utilizando genómica comparada. Esta aproximación se basó en la combinación de dos algoritmos bioinformáticos, implementados en los programas eQRNA y RNAz, cuyos fundamentos más importantes se han descrito en la introducción de este capítulo.

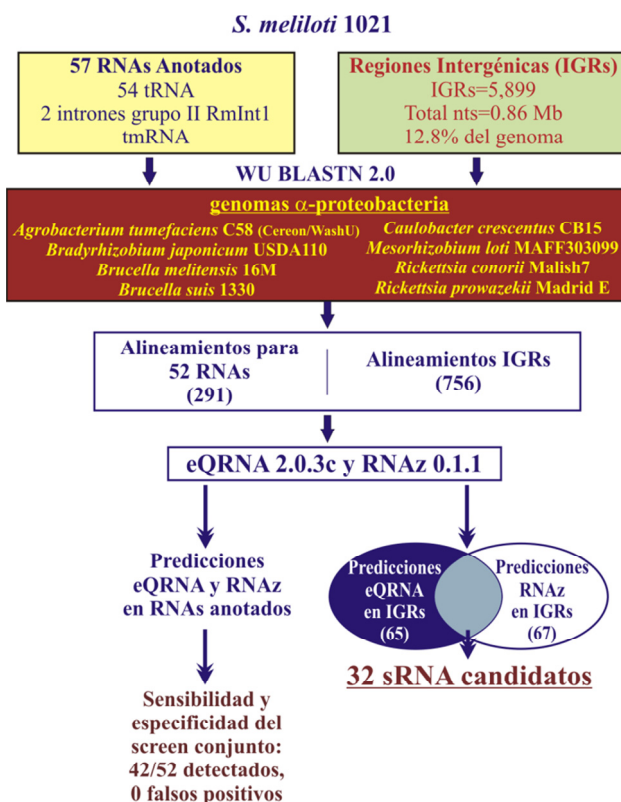
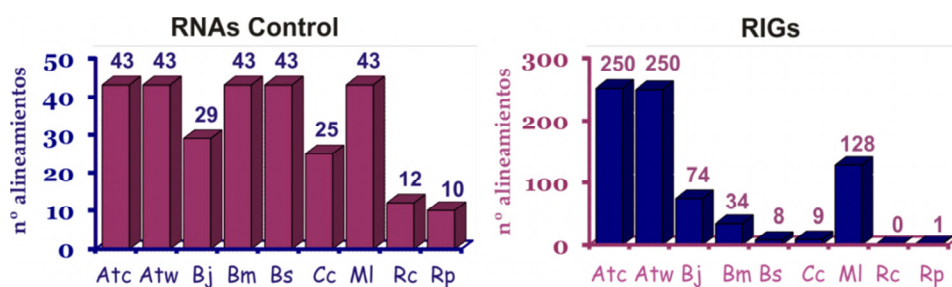


Fig. I-6. Estrategia bioinformática utilizada para la predicción de sRNAs en el genoma de *S. meliloti* 1021.

El *screening* predictivo se realizó en la cepa *S. meliloti* 1021, según la estrategia que se ilustra en la Fig. I-6. Inicialmente se creó un fichero con todas las **RIGs** (secuencias

entre regiones anotadas) del genoma de *S. meliloti* 1021 que contenía 5.899 secuencias, que representan aproximadamente 0,860 Mb, un 12,8% del genoma de esta bacteria. En este fichero no figuran las RIGs menores de 50 nt, por ser inferiores al tamaño esperado para codificar un sRNA de acuerdo a la bibliografía. El tamaño medio de estas RIGs es de 146 nt, siendo la mayor de 2.792 nt. Paralelamente se generó otro fichero, que denominamos de referencia, con secuencias de 57 RNAs no codificantes, ya conocidos en *S. meliloti* antes de comenzar este estudio; 54 tRNAs, 2 copias del intrón del grupo II RmIntI y el tmRNA. Ambos grupos de secuencias (referencia y RIGs) se compararon mediante WU-BLASTN con ocho genomas de  $\alpha$ -proteobacterias de distinta relación filogenética con *S. meliloti*: los simbiontes de leguminosas *Mesorhizobium loti* MAFF303099 y *Bradyrhizobium japonicum* USDA110, el fitopatógeno *Agrobacterium tumefaciens* C58 (Cereum y WashU), los patógenos animales *Brucella melitensis* 16M, *Brucella suis* 1330, *Rickettsia conorii* Malish7 y *Rickettsia prowazekii* Madrid F y la bacteria acuática de vida libre *Caulobacter crescentus* CB15.

Estas comparaciones generaron un total de 756 alineamientos correspondientes a las RIGs y 291 alineamientos que involucraban a 52 de los 57 sRNAs conocidos incluidos en el análisis, todos con valores de  $E \leq 0,00001$ . WU-BLASTN no generó alineamiento alguno para 5 de los 54 tRNAs incluidos en el grupo de secuencias de referencia. La distribución de todos los alineamientos en estos grupos bacterianos se muestra en la Fig. I-7. Los dos grupos de alineamientos así generados fueron analizados simultánea e independientemente con eQRNA y RNAz para la predicción de sRNAs según los criterios y valores de corte descritos en el apartado de Diseño Experimental.



**Fig. I-7. Distribución de los alineamientos** generados por la comparación de las secuencias de los RNAs de referencia (control) y las RIGs con los genomas de: Atc, *A. tumefaciens* (anotación Cereum); Atw, *A. tumefaciens* (anotación WashU); Bj, *B. japonicum*; Bm, *B. melitensis*; Bs, *B. suis*; Cc, *C. crescentus*; Ml, *M. loti*; Rc, *R. conorii*; Rp, *R. prowazekii*.

Los alineamientos con los RNAs de referencia se utilizaron para la estimación de la sensibilidad y especificidad del método (Tabla I-2) así como para establecer el

mejor criterio para la predicción de nuevos sRNAs a partir de los alineamientos generados con las RIGs.

La **sensibilidad** se define como la capacidad del programa para identificar RNAs entre un grupo de secuencias nucleotídicas y fue cuantificada como el porcentaje de RNAs predichos por cada programa entre los del grupo de referencia. En nuestro análisis, eQRNA identificó como RNAs no codificantes 48 de los 52 RNAs conocidos que generaron alineamientos con WU-BLASTN, por tanto, su sensibilidad fue de un 92%. Por su parte, RNAz denominó RNA a 48 de esas 52 secuencias, por lo que la sensibilidad de este algoritmo fue del 83%. Si consideramos las predicciones comunes a ambos programas en este grupo de RNAs de referencia la sensibilidad fue del 81% (42 de 52 secuencias). El término **especificidad** alude a las posibles falsas predicciones positivas (predicciones casuales). Para hacer esta estimación se barajaron aleatoriamente todos los alineamientos correspondientes a los 52 RNAs anotados y se sometieron a un nuevo análisis por los dos programas. Se consideraron falsos positivos en términos estadísticos a aquellas predicciones iniciales que siguen identificándose como RNAs después de la modificación aleatoria de los alineamientos. En este segundo análisis eQRNA volvió a identificar como RNA cuatro secuencias entre las de referencia mientras que RNAz lo hizo con ocho, pero en este caso no hubo predicciones comunes. Concluimos, por tanto, que considerando los candidatos comunes a ambos métodos se mantiene un alto grado de sensibilidad en la predicción de sRNAs (81%) con la máxima especificidad posible (ningún falso positivo estadísticamente hablando).

Tabla I-2. Sensibilidad y especificidad de eQRNA y RNAz.

	BLASTN <sup>a</sup>	eQRNA <sup>b</sup>	RNAz <sup>b</sup>	eQRNA y RNAz <sup>b</sup>
<b>Alineamientos reales (Sensibilidad)</b>	52/57	48/52	43/52	42/52
<b>Alineamientos aleatorios (Especificidad)</b>	52/57	4/52	8/52	0/52

<sup>a</sup> El análisis mediante BLASTN detectó 52 de los 57 RNAs.

<sup>b</sup> Un RNA dado se consideró detectado, si al menos la mitad de su secuencia estaba incluida en la predicción.

El análisis mediante eQRNA de los alineamientos generados por la comparación de las RIGs generó una lista de 65 sRNA candidatos mientras que RNAz predijo 67 (Tablas I-3, I-4). En ambos casos están excluidas todas las secuencias conservadas inmediatamente adyacentes al extremo 5' de ORFs ya anotados, probablemente correspondientes a *riboswitches* u otros elementos *cis*-reguladores, cuya identificación no era el objetivo inicial de este estudio. La comparación de las dos listas de

candidatos reveló 32 predicciones comunes a eQRNA y RNAz (Tabla I-5) que, de acuerdo al análisis previo de los alineamientos generados con los RNAs de referencia, se consideraron como sRNAs más probables.

Tabla I-3. Predicciones de sRNAs en el genoma de *S. meliloti* 1021 correspondientes a eQRNA

Candidato# <sup>a</sup>	Comienzo	Final
<b>C1 (SRP RNA)</b>	<b>259883</b>	<b>260047</b>
C4	732700	732758
C3	734151	734207
C5	954802	954949
C8	1275353	1275454
C18	1315130	1315273
C28	1862341	1862491
C30	1931163	1931294
C31	2157286	2157417
C34	2475520	2475661
C35	2575817	2575937
C36	2921731	2921865
C38	2962478	2962620
C40	2987766	2987933
C46	3130004	3130100
C47	3189168	3189303
C49	3505973	3506085
C50	3522291	3522396
C54	3617197	3617312
B1	163488	163625
B5	743007	743123
B6	1025545	1025666
B7	1107222	1107339
B8	1222801	1222918
B10	1287771	1287820
B11	1402770	1402916
B12	1633802	1633915
A3	291983	292114
A9	1144086	1144232
A11	1322901	1323044

Tabla I-4. Predicciones de sRNAs en el genoma de *S. meliloti* 1021 correspondientes a RNAz

Candidato# <sup>a</sup>	Comienzo	Final
C2	267007	267130
C3	318369	318503
C4	398402	398545
C6	1007497	1007602
C23	1315178	1315359
C25	1449239	1449387
C26	1607646	1607756
C27	1614125	1614272
C28	1712360	1712504
C32	2157322	2157466
C33	2276537	2276680
C37	2922966	2923066
C39	2986376	2986641
C41	2987766	2987933
C42	3044591	3044714
C43	3046671	3046803
C44	3129955	3130146
C48	3234147	3234294
C51	3522263	3522410
C52	3540367	3540511
C53	3542132	3542245
C55	3617197	3617312
B2	194764	194869
B3	486547	486687
B4	551576	551723
B9	1230061	1230250
B13	1653785	1653898
A5	204688	204837
A7	552807	553002
A8	906676	906808
A10	1275503	1275700

a. Las letras antes del número de cada candidato indican la localización genómica: C, Cromosoma; B plásmido simbiótico pSymB; A plásmido simbiótico pSymA

Estas 32 secuencias se compararon mediante BLASTN con todos los genomas bacterianos depositados en las bases de datos en el momento del *screening*. Este análisis reveló que 17 de estas 32 secuencias son elementos repetidos (REP) en el genoma de *S. meliloti* y fueron inicialmente consideradas como no génicas, 5 fueron identificadas como posibles ORFs predichos por el programa *Glimmer 2.0* en algunas bacterias incluidas en la comparación y 2 (B30 y C19), resultaron corresponder a los sRNAs IncA y RNasa P, respectivamente. IncA es un *loci* sRNA identificado después



de iniciado este trabajo entreo los genes *repB* y *repC* de los replicones pSymA y pSymB, que se transcribe en dirección opuesta a éstos y media la incompatibilidad entre plásmidos *repABC* de  $\alpha$ -proteobacterias.

Tabla I-5. Predicciones de sRNAs en el genoma de *S. meliloti* 1021 correspondientes a los programas eQRNA y RNAz.

Candidato# <sup>a</sup>	Comienzo	Final	Tamaño predicho (nt)	Genes Flanqueantes	Hebra <sup>b</sup>	Observaciones <sup>c</sup>
C7	201639	201834	196	<i>polA/SMc02851</i>	< >	posible sRNA
C9	1398397	1398274	124	<i>SMc01933/proS</i>	< >	posible sRNA
C10	1411678	1411808	131	<i>celR2/rpmG</i>	> <	posible ORF
C14	1667641	1667484	158	<i>SMc02051/tig</i>	< >	posible sRNA
C15	1698744	1698610	135	<i>SMc01226/SMc01225</i>	< <	posible sRNA
C16	1699021	1698812	210	<i>SMc01226/SMc01225</i>	< <	posible sRNA
C17	2098405	2098598	194	<i>SMc04270/SMc04273</i>	< <	sec. repetida Sm-2
C19	2357208	2356761	448	<i>SMc01857/SMc01856</i>	< <	<i>RNasaP</i>
C20	2398184	2398328	145	<i>SMc01608/ribH2</i>	> >	posible ORF
C21	2922966	2923066	101	<i>ctrA/SMc00653</i>	> <	sec. repetida Sm-4
C22	2972265	2972118	148	<i>SMc03975/SMc03976</i>	> <	posible sRNA
C24	3074389	3074571	183	<i>pgm/glgX1</i>	> >	sec. repetida Sm-2
C45	3105374	3105169	206	<i>SMc02983/SMc02984</i>	< >	posible sRNA
B29	24848	25053	206	<i>SMb20017/SMb20018</i>	< <	sec. repetida Sm-4
B30	56401	56555	155	<i>repC1/repB1</i>	< <	<i>Inca</i>
B31	66313	66506	194	<i>SMb20055/SMb20056</i>	> >	sec. repetida
B32	73482	73615	134	<i>SMb20064/SMb20065</i>	< >	sec. repetida Sm-4
B33	231462	231581	120	<i>SMb20223/smc22-1</i>	> <	sec. repetida Sm-4
B34	334401	334598	198	<i>thuB/SMb20331</i>	> <	sec. repetida Sm-5
B35	577732	577875	144	<i>SMb20551/SMb20552</i>	< >	posible sRNA
B36	783541	783640	100	<i>SMb21220/SMb21221</i>	< <	sec. repetida Sm-4
B37	908039	908163	125	<i>SMb21162/hutU</i>	> <	sec. repetida Sm-4
B38	1116680	1116860	181	<i>SMb21577/atcU2</i>	< >	posible ORF
B39	1230061	1230250	190	<i>SMb20993/SMb20994</i>	> <	sec. repetida Sm-4
B41	1249055	1249166	112	<i>xdhB2/SMb21676</i>	> <	sec. repetida
B42	1373274	1373420	147	<i>SMb21444/SMb21445</i>	> <	sec. repetida Sm-4
B43	1525569	1525724	156	<i>SMb20720/SMb20721</i>	> <	sec. repetida
B44	1669957	1670129	173	<i>paaX/SMb21642</i>	< <	sec. repetida
A1	143338	143475	138	<i>SMA0255/SMA0257</i>	> <	sec. repetida Sm-5
A2	512130	512373	244	<i>SMA0922/traG</i>	> <	posible ORF
A4	1304125	1304268	144	<i>kdpA/SMA2335</i>	< <	sec. repetida Sm-5
A6	1328169	1328367	199	<i>SMA2355/SMA2357</i>	> >	posible ORF

<sup>a</sup> Las letras antes del número de cada candidato indican la localización genómica: C, Cromosoma; B plásmido simbiótico pSymB; A plásmido simbiótico pSymA. Los candidatos sobreados son los elegidos finalmente para la verificación experimental.

<sup>b</sup> Orientación de los genes flanqueantes. Los genes presentes en la cadena cuya secuencia aparece en las bases de datos se indican mediante (>) y los de la cadena complementaria por (<).

<sup>c</sup> Observaciones e información adicional obtenida por comparación mediante BLASTN de todos los candidatos.

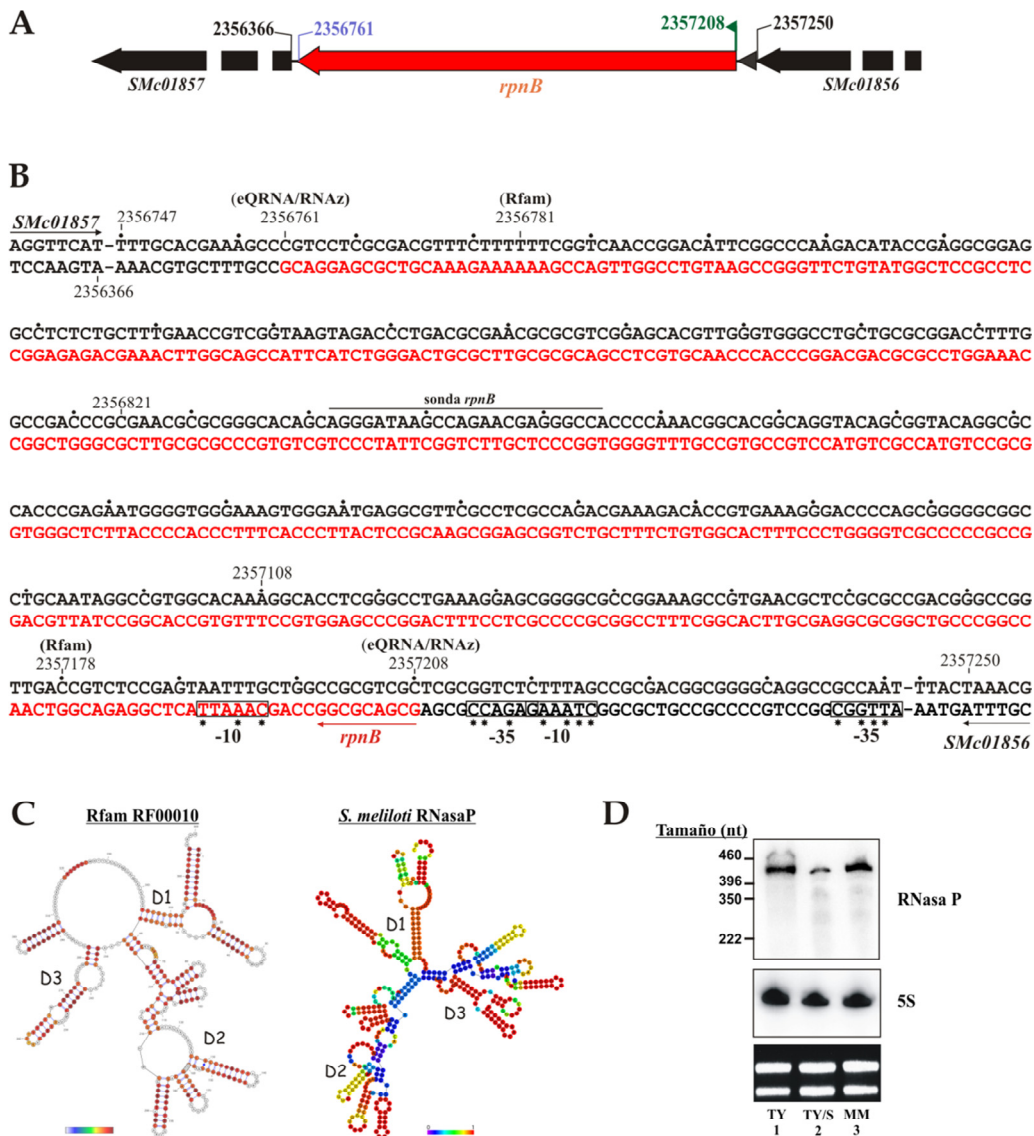
Las 8 secuencias restantes, 7 localizadas en el Cromosoma y una en el plásmido simbiótico pSymB, se consideraron nuevos *loci* sRNA más probables siendo denominados *smr*, en alusión a *S. meliloti* RNA.

### I-3.2. sRNAs DE FUNCIÓN CONOCIDA

Entre los candidatos resultantes de nuestro *screening* e, aparecieron los ortólogos de RNasa P, SRP y 6S en *S. meliloti*, todos ellos RNAs no codificantes de función bien caracterizada en *E. coli* y ampliamente conservados en bacterias. Sin embargo, sus respectivos *loci* no fueron anotados tras el primer análisis de la secuencia genómica de *S. meliloti* 1021. A continuación se describen los hallazgos relevantes para la identificación y anotación de los mismos como genes RNA.

#### I-3.2.1. RNasa P

eQRNA y RNAz identificaron como posible RNA la secuencia de 448 nt comprendida entre las coordenadas G 2.357.208 y G 2.356.761 del cromosoma de *S. meliloti* 1021 (candidato C19 en Tabla I-5). Este *locus* se localiza en la RIG delimitada por los ORFs *mraZ* (SMc1857) y SMc01856 que podrían codificar un posible regulador transcripcional y una sarcosina oxidasa, respectivamente (Fig. I-8A). BLASTN reveló que esta secuencia nucleotídica estaba conservada en un gran número de bacterias, con porcentajes de identidad superiores al 65% en la mayoría de los casos, apareciendo anotada en muchas de ellas como correspondiente al gen *rpnB* que codifica el componente RNA (M1) de la ribonucleasa RNasa P. En Rfam aparecen catalogadas como codificantes del RNA de la RNasa P de la clase bacteriana A un total de 4.081 *loci* cromosómicos, entre ellos el de *S. meliloti*, anotado en esta base de datos entre las posiciones genómicas G 2.357.178 y A 2.356.781 (428 nt) y por tanto, contenido también en nuestra predicción (Fig. I-8B). El modelo de co-varianza para esta familia de ribozimas ha sido inferido del alineamiento de 306 de estas secuencias (modelo Rfam RF00010). Del mismo se deduce una estructura secundaria consenso en la que aparecen al menos tres subdominios (D1-D3), identificables también en el transcrito de *S. meliloti* 1021 cualquiera que sea la secuencia utilizada (predicción de eQRNA/RNAz o anotación de Rfam) para el plegamiento *in silico* de la molécula (Fig. I-8C). Un oligonucleótido de 25 residuos que hibrida en la región central del gen *rnpB* reconoce un único transcrito del tamaño esperado para la RNasa P (428-448 nt) en RNA total de *S. meliloti* 1021 extraído de bacterias en vida libre cultivadas en medio completo TY o medio mínimo, confirmando así la expresión de este RNA esencial en esta bacteria (Fig. I-8D). Esta molécula de RNA podría originarse por transcripción desde cualquiera de los dos promotores  $\sigma^{70}$  predichos en su región 5', por lo que la identificación precisa de su extremo 5' primario debe hacerse experimentalmente (Fig. I-8B).



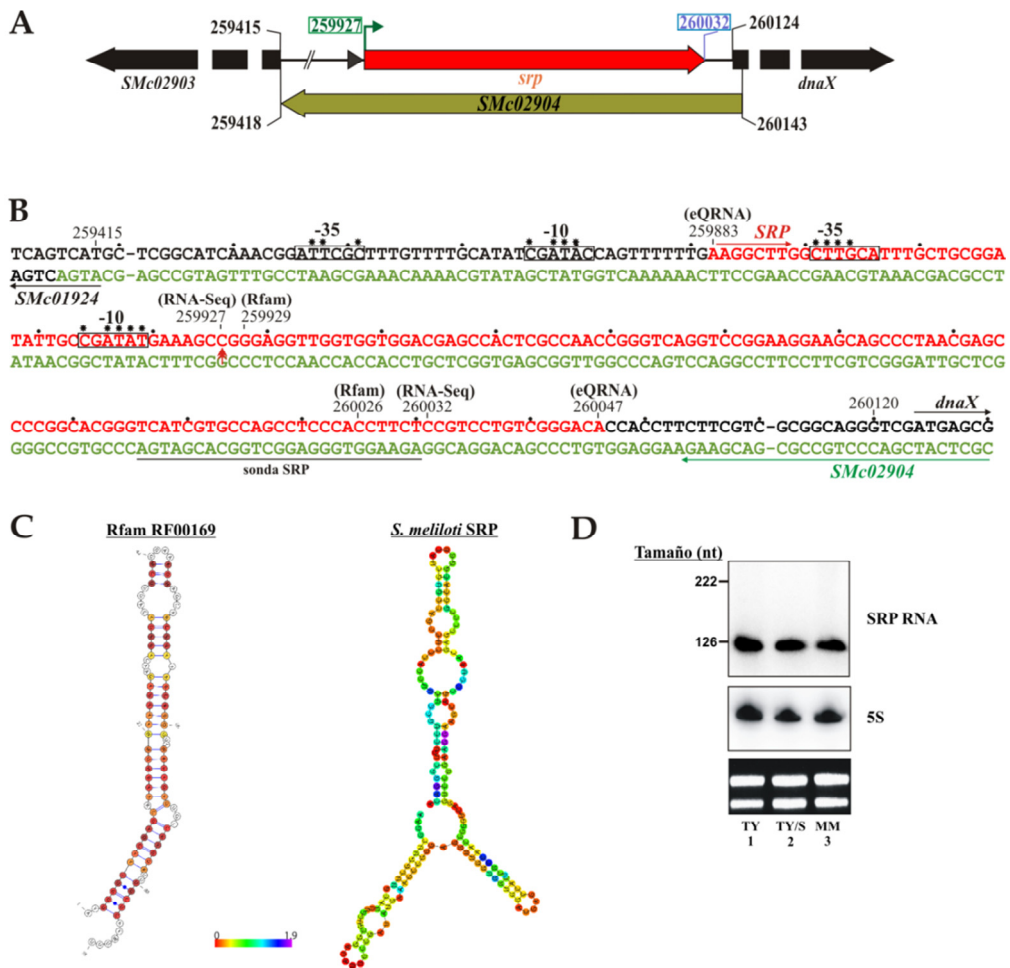
**Fig. I-8. Anotación del gen *rpnB* (RNasa P) en *S. meliloti* 1021.** A) Diagrama correspondiente a la región genómica donde se localiza el gen de la RNasa P B) Secuencia nucleotídica de la región genómica del gen *rpnB*, con indicación de las coordenadas predichas para el mismo por eQRNA/RNAz, de la anotación de Rfam, de posibles cajas promotoras en su región 5' y de la sonda empleada para la detección de su expresión. C) Representación de la estructura secundaria consenso de la familia RNasa P identificada en Rfam (izquierda) y de la RNasa P de *S. meliloti* predicha por eQRNA (derecha). D1, D2 y D3 corresponden a los dominios conservados del transcrito. D) Análisis de la expresión del RNA de la RNasa P mediante northern blot, utilizando una sonda específica del transcrito en tres muestras de RNA total de *S. meliloti* (TY: medio rico fase logarítmica; TY/S: medio rico fase estacionaria; MM: medio mínimo fase logarítmica).

### I-3.2.2. RNA SRP O 4.5S

La secuencia de 165 nt comprendida entre las posiciones A 259.883 y A 260.047 del cromosoma de *S. meliloti* 1021 fue identificada como RNA por el programa

eQRNA pero no por RNAz, por lo que no aparece en nuestra lista priorizada de 32 candidatos. Sin embargo, cuando se interrogaron con la misma los genomas bacterianos mediante BLASTN aparecieron regiones con alto grado de homología (porcentajes de identidad superiores al 75% en todos los casos) en un gran número de bacterias de relación filogenética diversa. En el momento de realizar nuestro *screening* esta secuencia mapeaba en la RIG definida por los ORFs *dnaX* y *SMc02904* que codifican la subunidad  $\tau$  de la DNA polimerasa III y una proteína hipotética, respectivamente, definiendo una región genómica de alta sintenia en bacterias. Con posterioridad esta región ha sido reanotada en *S. meliloti* 1021 de forma que en su anotación actual el gen *SMc02904* solapa completamente con el supuesto *locus* RNA y su hebra codificante es complementaria a éste (Fig. I-9A). En la base de datos Rfam la secuencia de 98 nt entre las posiciones G 259.929 y C 260.026 del cromosoma de *S. meliloti* 1021, y por tanto comprendida en la predicción de eQRNA (Fig. I-9B), aparece anotada como la codificante de la molécula de RNA de pequeño tamaño que forma parte de la ribonucleoproteína SRP. En Rfam aparecen recogidas como SRP RNA bacteriano un total de 2.940 secuencias, todas ellas de localización cromosómica. Su estructura secundaria consenso ha sido inferida del alineamiento de 271 de ellas (modelo de co-varianza Rfam RF00169) y a la misma se ajustaría discretamente el transcrito predicho en *S. meliloti* 1021 (Fig. I-9C).

Un oligonucleótido complementario a una secuencia de 25 nt de la región 3' del RNA identificado por eQRNA produce una fuerte y única señal de hibridación correspondiente a un transcrito de tamaño ligeramente inferior a 126 nt que se acumula constitutivamente en bacterias cultivadas en vida libre (Fig. I-9D). Con posterioridad a estos experimentos la secuenciación masiva (RNA-Seq) de la fracción de RNAs de pequeño tamaño de *S. meliloti* 2011 reveló la expresión de un transcrito de 107 nt denominado SmelC035, cuyos extremos 5' y 3' mapean en las posiciones cromosómicas C 259.927 y C 260.032 (Fig. I-9B), respectivamente, de la cepa de referencia *S. meliloti* 1021, en gran correlación con la anotación de Rfam para el RNA de SRP y con la hibridación Northern (Schlüter *et al.*, 2010). La expresión de este sRNA sería compatible con la transcripción desde el promotor  $\sigma^{70}$  predicho inmediatamente antes del extremo 5' identificado mediante RNA-Seq (Fig. I-9B).



**Fig. I-9. Anotación del RNA 4.5S o SRP de *S. meliloti* 1021.** A) Diagrama correspondiente a la región genómica donde se localiza el gen *srp* B) Secuencia nucleotídica de la región genómica del gen *srp*, con indicación de las coordenadas predichas para el mismo por eQRNA/RNAz, de la anotación de Rfam, de posibles cajas promotoras en su región 5' y de la sonda empleada para la detección de su expresión. C) Representación de la estructura secundaria consenso de la familia SRP identificada en Rfam (izquierda) y del SRP de *S. meliloti* predicho por eQRNA (derecha). D) Análisis de la expresión del RNA 4.5S mediante northern blot, utilizando una sonda específica del transcrito en tres muestras de RNA total de *S. meliloti* (TY: medio rico fase logarítmica; TY/S: medio rico fase estacionaria; MM: medio mínimo fase logarítmica).

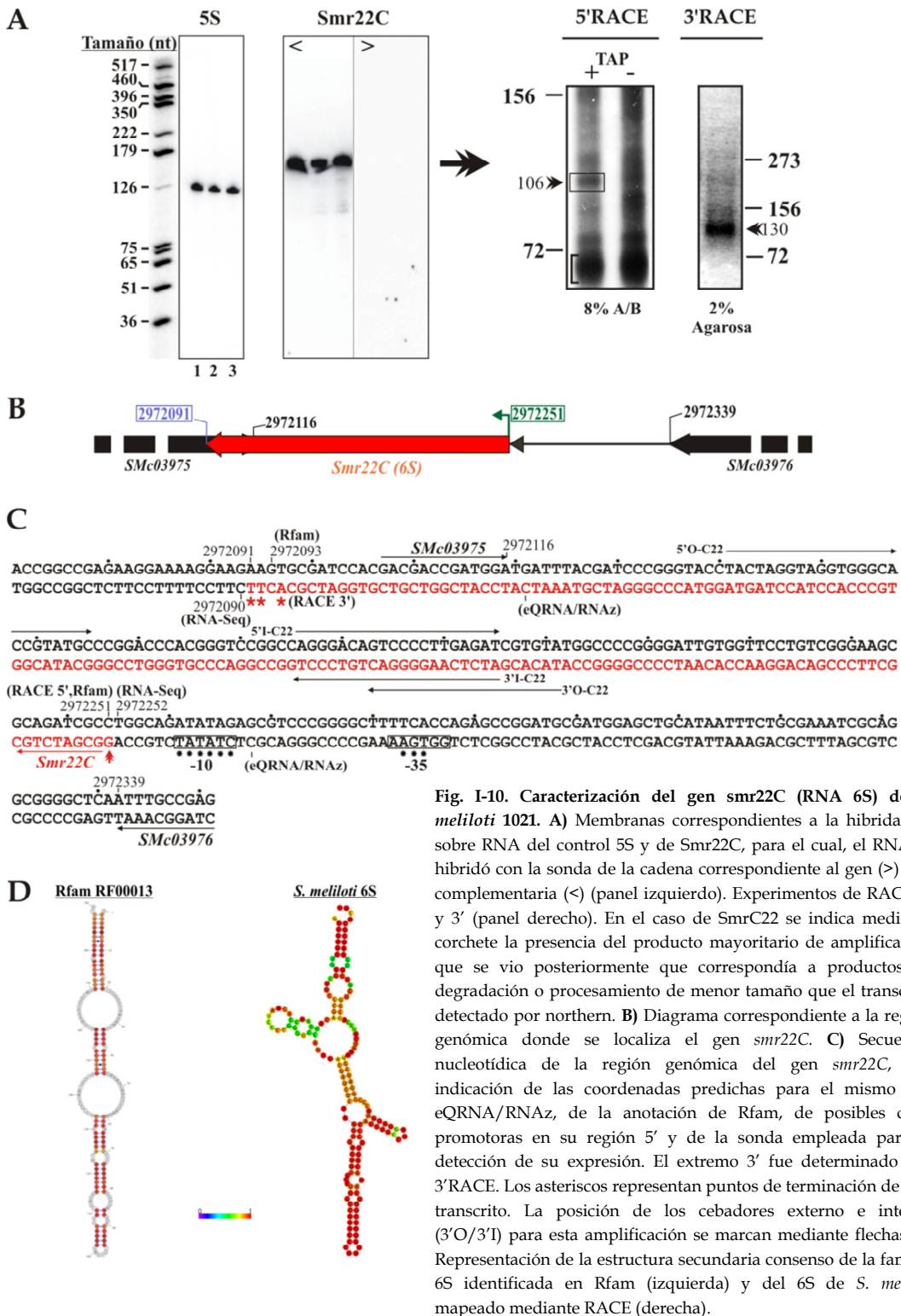
### I-3.2.3. RNA 6S

En la lista de predicciones comunes a eQRNA y RNAz se encuentra la secuencia de 148 nt comprendida entre las coordenadas C 2.972.118 y C 2.972.265 del cromosoma de *S. meliloti* 1021 (C22; Tabla I-5), en la RIG definida por los genes *SMc03975* y *SMc03976*, que codifican un regulador transcripcional del tipo LysR y una proteína hipotética, respectivamente. BLASTN puso de manifiesto una gran conservación de esta secuencia nucleotídica en bacterias pero no reveló identidad

alguna para este *locus*. En el momento del *screening* tampoco aparecía ningún sRNA anotado en Rfam en esta región del genoma de *S. meliloti* por lo que inicialmente C22 fue considerado como un sRNA desconocido hasta entonces en bacterias, que denominamos Smr22C, y fue sometido a verificación experimental mediante hibridación Northern y mapeo RACE.

Dado que ni eQRNA ni RNAz predicen la hebra codificante del sRNA, para asignar la polaridad del gen, el RNA total de la bacteria se hibridó de forma independiente con dos oligonucleótidos totalmente complementarios, cada uno diseñado de acuerdo a la secuencia nucleotídica de cada cadena. La hibridación reveló la presencia de un transcrito discreto, de tamaño compatible con la predicción bioinformática (126-179 nt) cuando se utilizó una de las sondas, mientras que con la complementaria no se obtuvo señal alguna de hibridación (Fig. I-10A, panel izquierdo). La acumulación de Smr22C fue aparentemente constitutiva en condiciones de crecimiento en medio rico (fase logarítmica y fase estacionaria), así como en medio mínimo (fase logarítmica).

Para la identificación del extremo 5' del transcrito detectado en los geles se utilizó la tecnología RACE sobre RNA total obtenido de bacterias crecidas en medio TY hasta fase logarítmica (Fig. I-10A, panel derecho). Este experimento reveló una serie de productos mayoritarios de tamaño inferior a 72 nt como resultado de la amplificación del cDNA de Smr22C (identificadas por un corchete en Fig. I-10A), tanto en las muestras de RNA tratadas con TAP como en las control. La secuenciación de los mismos identificó extremos 5' muy diversos, probablemente correspondientes a productos de procesamiento o degradación de este transcrito, que no son detectables en los Northern. Sin embargo, también fue evidente un producto RACE minoritario, aunque del tamaño esperado y obtenido específicamente en muestras de RNA tratadas con TAP. Su secuenciación identificó un extremo 5' para Smr22C que mapea en la coordenada G 2.972.251 del cromosoma de *S. meliloti* 1021, que fue considerado como el sitio de inicio de transcripción más probable para este sRNA (Fig. I-10B y C). El análisis de la secuencia nucleotídica de Smr22C desde su extremo 5' primario hasta la región 3' del gen predicha por eQRNA/RNAz no evidenció terminadores de transcripción reconocibles, por lo que el extremo 3' de este transcrito también fue determinado mediante RACE. En este caso se obtuvo un producto de amplificación del cDNA mayoritario, aparentemente único y del tamaño esperado. La secuenciación de los insertos de hasta diez plásmidos obtenidos tras el clonaje de este fragmento de DNA reveló tres extremos 3' para este transcrito que, no



**Fig. I-10. Caracterización del gen *smr22C* (RNA 6S) de *S. meliloti* 1021. A) Membranas correspondientes a la hibridación sobre RNA del control 5S y de *Smr22C*, para el cual, el RNA se hibridó con la sonda de la cadena correspondiente al gen (>) y la complementaria (<) (panel izquierdo). Experimentos de RACE 5' y 3' (panel derecho). En el caso de *Smr22C* se indica mediante corchete la presencia del producto mayoritario de amplificación que se vio posteriormente que correspondía a productos de degradación o procesamiento de menor tamaño que el transcrito detectado por northern. B) Diagrama correspondiente a la región genómica donde se localiza el gen *smr22C*. C) Secuencia nucleotídica de la región genómica del gen *smr22C*, con indicación de las coordenadas predichas para el mismo por eQRNA/RNAz, de la anotación de Rfam, de posibles cajas promotoras en su región 5' y de la sonda empleada para la detección de su expresión. El extremo 3' fue determinado por 3'RACE. Los asteriscos representan puntos de terminación de este transcrito. La posición de los cebadores externo e interno (3'O/3'I) para esta amplificación se marcan mediante flechas D) Representación de la estructura secundaria consenso de la familia 6S identificada en Rfam (izquierda) y del 6S de *S. meliloti* mapeado mediante RACE (derecha).**

obstante, mapean en posiciones prácticamente contiguas en el cromosoma de *S. meliloti* 1021; T 2.972.091, T 2.972.092 y A 2.972.094, solapando con la región 3' del gen *SMc03975*, que se transcribe en dirección opuesta a Smr22C (Fig. I-10B y C). El tamaño de Smr22C calculado considerando su sitio de inicio de transcripción y el extremo 3' más distante a éste (nucleótido A 2.972.091) es de 161 nt, en muy buena correlación con el del transcrito detectado mediante Northern. El análisis de la región promotora de *smr22C* identificó cajas -10/-35 con la secuencia consenso reconocida por  $\sigma^{70}$  cuya localización es compatible con el inicio de la transcripción desde la posición determinada mediante RACE 5' (Fig. I-10C).

Con posterioridad se anotó en Rfam como correspondiente al sRNA 6S la secuencia de 159 nt comprendida entre las posiciones G 2.972.251 y T 2.972.092 del genoma de *S. meliloti* 1021, coincidiendo casi exactamente con las determinadas experimentalmente en este trabajo. En esta base de datos aparecen 1.700 secuencias de esta familia de sRNAs, cuyos miembros han sido identificados en casi todos los genomas bacterianos secuenciados hasta la fecha. La estructura secundaria consenso de estos transcritos se ajusta al modelo de co-varianza RF00013 inferido del alineamiento de 154 secuencias. Ésta consiste en una larga horquilla interrumpida por varios bucles de nucleótidos desapareados y el plegamiento predicho *in silico* para la secuencia del RNA 6S de *S. meliloti* determinada en este trabajo se puede considerar ajustado efectivamente a este modelo.

### I-3.3. IDENTIFICACIÓN DE NUEVOS sRNAs

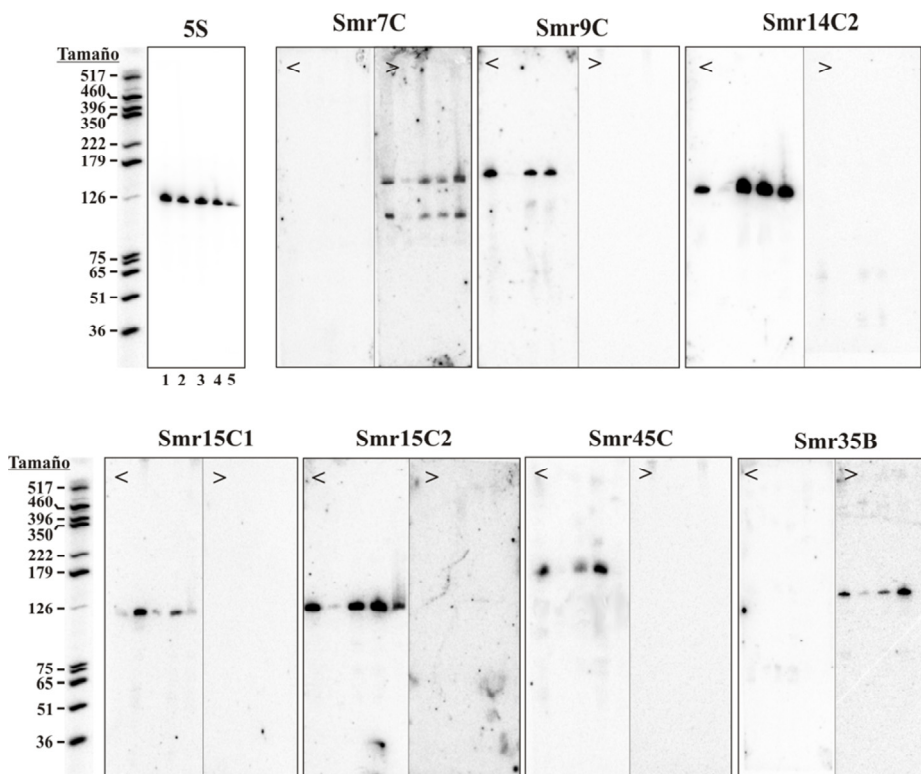
Las siete secuencias restantes de nuestra lista priorizada de ocho candidatos nunca aparecieron anotadas durante el desarrollo de este trabajo como sRNAs en la base de datos Rfam y la búsqueda de homologías mediante BLASTN reveló que las mismas estaban conservadas en un número más limitado de bacterias que los tres sRNAs descritos en el epígrafe anterior. Estos datos anticipaban que se trataba de nuevos sRNAs bacterianos, que denominamos Smr7C, Smr9C, Smr14C, Smr15C1, Smr15C2, Smr35B y Smr45C, cuya primera caracterización molecular se llevó a cabo según se describe a continuación.

#### I-3.3.1 POLARIDAD Y EXPRESIÓN DIFERENCIAL DE LOS GENES *smr*

En primer lugar, para verificar la expresión de estos siete sRNAs en *S. meliloti* 1021 se planteó la hibridación de RNA total de la bacteria con oligonucleótidos de 25 residuos diseñados de acuerdo a la predicción bioinformática. Al igual que para el RNA 6S (Smr22C) se llevaron a cabo dos hibridaciones independientes con sondas totalmente complementarias para determinar la dirección de la transcripción en cada caso. En esta serie de experimentos el RNA total para realizar las hibridaciones se



obtuvo en cinco condiciones biológicas diferentes: crecimiento en medio completo TY (fase logarítmica y estacionaria) y medio mínimo (MM; fase logarítmica) para estudiar la expresión en vida libre, y crecimiento en MM suplementado con 20  $\mu$ M de luteolina y nódulos maduros inducidos en alfalfa por esta bacteria como condiciones relevantes en la simbiosis.

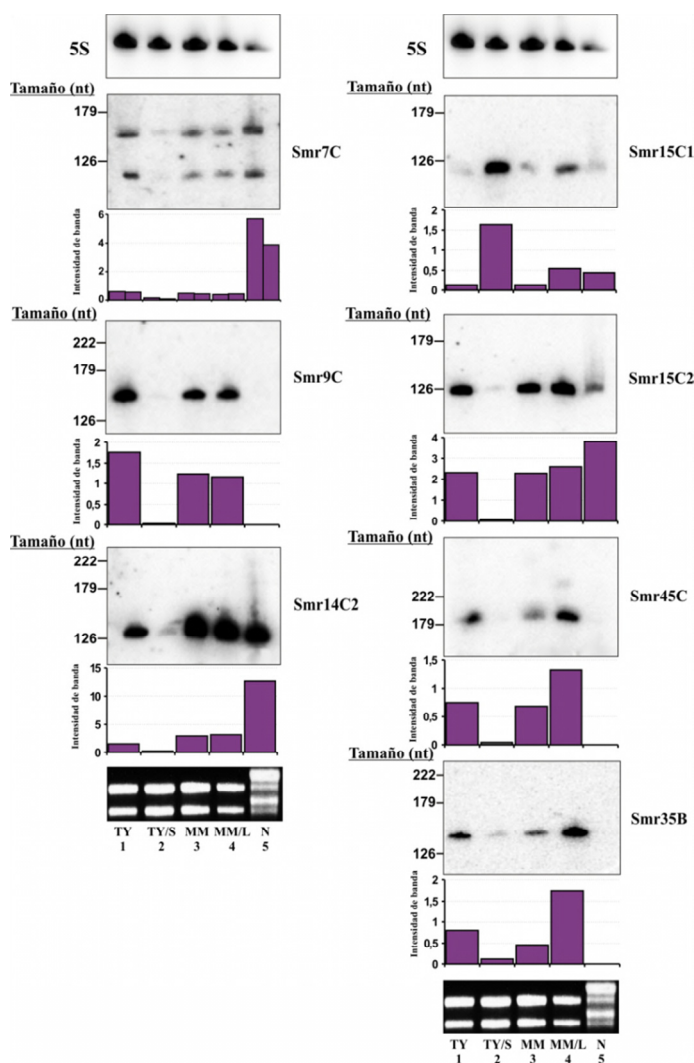


**Fig. I-11. Asignación de la polaridad de los genes *smr*.** Las membranas fueron hibridadas con oligonucleótidos complementarios (ver anexo) para ambas cadenas de los *loci* en estudio, la cadena anotada en la base de datos general de *S. meliloti* 1021 (>) y la complementaria (<). Los tiempos de exposición se optimizaron para cada sRNA, por lo que la intensidad de las bandas no es comparable entre cada transcrito. El marcador pGEM, marcado en 5', se resolvió en el mismo gel en donde se realizó la electroforesis y posteriormente se cortó de la membrana para estimar aproximadamente el tamaño de cada sRNA. Los números situados en la parte inferior de cada mebrana identifican las condiciones en las que se extrajo en RNA total: 1, TY f. logarítmica; 2, TY f. estacionaria; 3, MM logarítmica; 4, MM inducido por luteolina; 5, nódulo.

Como se observa en la Fig. I-11, en todos los casos se obtuvieron señales de hibridación correspondientes a transcritos cortos (<200 nt), en buena correlación con las predicciones bioinformáticas. Estas señales se observan en cada caso únicamente con un oligonucleótido pero no con la sonda complementaria, lo que permitió la asignación inequívoca de la cadena codificante de cada transcrito. Cada sonda detectó una única molécula de RNA salvo la específica de Smr7C que reveló dos

señales de hibridación de intensidad similar que, como se describe con detalle en las siguientes secciones, corresponden al transcrito primario y a su forma procesada.

La acumulación de los transcritos Smr en cada una de las condiciones biológicas analizadas se valoró por cuantificación de cada señal de hibridación mediante el programa *Quantity One*® (Bio-rad) en relación a la del RNA 5S, que se utilizó como control de expresión constitutiva (Fig. I-12). No obstante hay que tener en cuenta que algunos datos transcriptómicos han revelado que el RNA 5S se acumula del orden de 8 veces más en bacteroides que en bacterias cultivadas en medio TY (Barnett *et al.*, 2004), por lo que la expresión de los sRNAs en tejido nodular podría estar incluso subestimada en nuestros ensayos.



**Fig. I-12. Patrones de expresión de los genes *smr*.** Los números situados en la parte inferior de cada columna de la Fig. indican las condiciones en las que se extrajo el RNA total: 1, TY f. logarítmica; 2, TY f. estacionaria; 3, MM f. logarítmica; 4, MM con luteolina; 5, nódulo. Los tamaños correspondientes al marcador se muestran a la derecha de cada membrana. Las señales de hibridación ha sido cuantificadas con *Quantity One*® y normalizadas con la señal del 5S RNA, cuya hibridación se muestra en la parte superior de cada serie. Los resultados se representan en los histogramas situados en la parte inferior de cada membrana. Las unidades de intensidad de banda son unidades arbitrarias. En la parte inferior de cada serie se muestran los controles de carga en las electroforesis en gel desnaturizante de agarosa al 1,75% y teñidas con bromuro de etidio.

Los siete sRNAs identificados presentaron perfiles de expresión diferencial. En todos los casos esta expresión depende de la fase de crecimiento de las bacterias, siendo la mayoría reprimidos en fase estacionaria. La acumulación en fase estacionaria tardía sólo se observó en el caso de Smr15C1.

Todos los genes sRNA se expresaron en el cultivo en MM donde la luteolina estimuló moderadamente la acumulación de los transcritos Smr35B, Smr45C y Smr15C1, aproximadamente 4, 2 y 3 veces respecto a su expresión en MM no suplementado, respectivamente. La expresión de Smr7C y SmrC14 se encontró incrementada en el nódulo del orden de 13 y 5 veces, respectivamente, en relación a las bacterias en vida libre, mientras que la inducción de Smr15C1 y Smr15C2 fue más moderada en tejido nodular (del orden de 2 veces respecto a las bacterias cultivadas). Sin embargo, los transcritos Smr9C, Smr35B y Smr45C no fueron detectados en bacterias endosimbióticas.

### I-3.3.2. CARACTERIZACIÓN DE LAS UNIDADES TRANSCRIPCIONALES DE LOS GENES *smr*

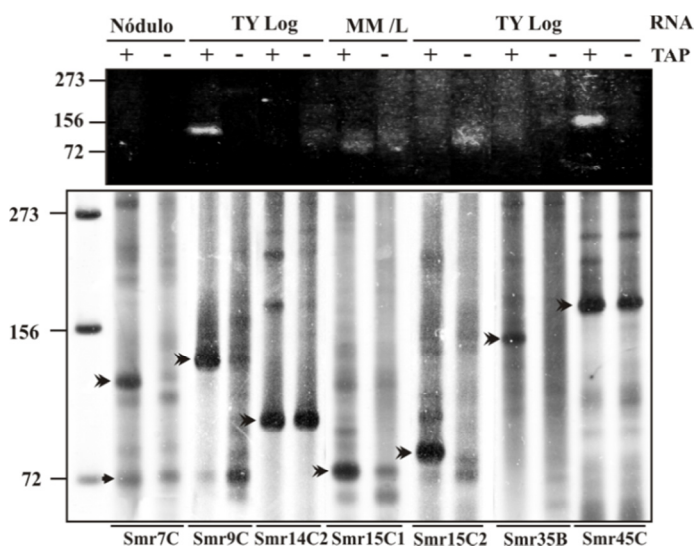
Confirmada la expresión de los 7 sRNAs candidatos, se procedió a la caracterización de las unidades transcripcionales de éstos. Para ello se mapearon sus extremos 5' mediante RACE, y se realizó un análisis bioinformático de las regiones promotoras y terminadoras. Los resultados de estas determinaciones se presentan en las Fig I-13, I-14 y I-15.

#### ***I-3.3.2.1. Identificación de los sitios de inicio de transcripción***

Las imágenes de los geles de agarosa y poliacrilamida en los que se resolvieron los productos de amplificación de los cDNAs correspondientes a cada transcrito se muestran en la Fig. I-13.

Dado que la interpretación de los resultados obtenidos de este análisis resulta compleja, para una mejor comprensión se describen caso por caso los correspondientes a cada sRNA.

**-Smr7C:** en el gel de poliacrilamida se observan varios productos RACE específicos de este transcrito que, sin embargo, no se detectan en agarosa. Cinco de seis secuencias independientes del fragmento de 118 pb que se obtiene únicamente en muestras de RNA tratadas con TAP situaron el inicio de transcripción de este sRNA en el residuo G 201.681 del Cromosoma de *S. meliloti* 1021, mientras que la secuencia restante lo hizo en la posición G 201.679. Además, aparece un segundo producto de 72 pb tanto en muestras tratadas con TAP como en los controles por lo



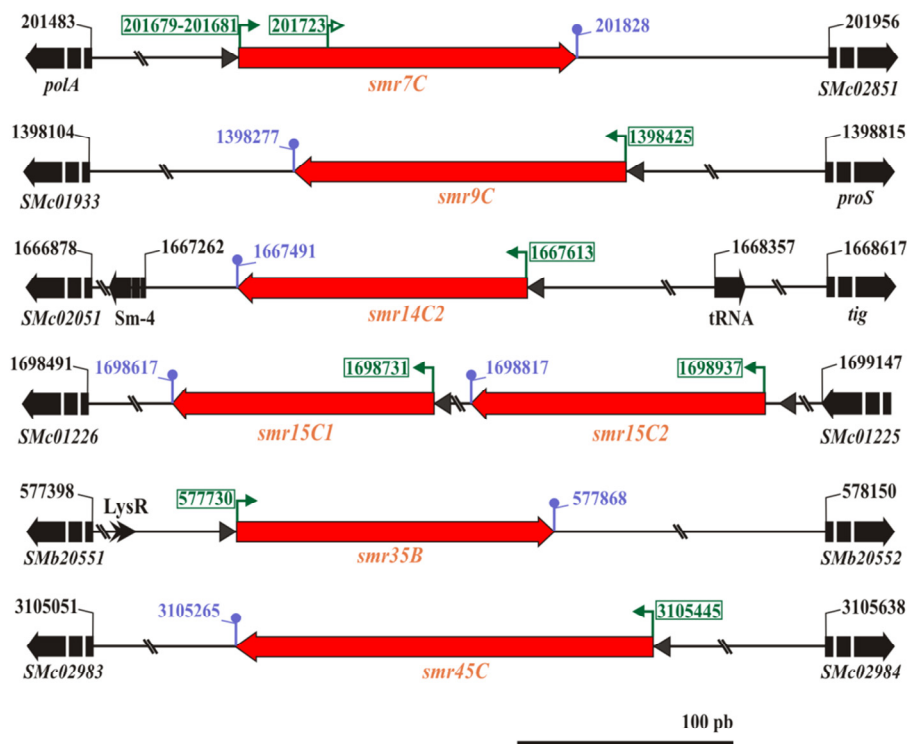
**Fig. I-13. RACE 5' de los transcritos Smr.** Gel de poliacrilamida al 10% teñido con nitrato de plata correspondiente a los productos de amplificación de los tratamientos con TAP (+) y los no tratados con TAP (-) de las muestras de RNA. En la parte superior se muestran la procedencia de las muestras de RNA total. El marcador de peso molecular utilizado fue el  $\Phi$ 29. La doble flecha señala el producto de amplificación incrementado en el tratamiento con TAP. La flecha simple señala el producto correspondiente al punto de procesamiento, únicamente detectado en el caso de Smr7C.

que probablemente corresponde a la amplificación de la forma procesada de Smr7C. La secuenciación de este fragmento de PCR localizó el sitio de procesamiento de este transcrito en el residuo C 201.723 del cromosoma.

**-Smr9C:** la amplificación del cDNA de este transcrito originó un producto mayoritario específico de 138 pb, únicamente después del tratamiento del RNA con TAP, que es resuelto en ambos tipos de geles. Varias secuencias independientes del mismo mapearon inequívocamente el extremo 5' primario de Smr9C en una única posición del Cromosoma de *S. meliloti* 1021, el residuo A 1.398.425.

**-Smr14C2:** el gel de poliacrilamida revela la presencia de un producto RACE mayoritario de 108 nt, apenas visible en agarosa, que es igualmente abundante en muestras tratadas con TAP y control. Sin embargo, ninguna de las 10 secuencias analizadas del fragmento obtenido en muestras control correspondía específicamente al transcrito Smr14C2, por lo que se consideró un artefacto de PCR o amplificación inespecífica. Por el contrario, 7 de 12 secuencias correspondientes al fragmento obtenido a partir de RNA tratado con TAP fueron específicas de este sRNA mapeando la posición del inicio de transcripción en el residuo A 1.667.613 del cromosoma.

**-Smr15C1 y Smr15C2:** estos dos transcritos tienen la particularidad de estar codificados en la misma RIG y sus secuencias nucleotídicas primarias muestran una alta homología (identidad del 84%). El gel de poliacrilamida resolvió (no así la agarosa) productos RACE específicos de 80 pb y 92 pb para Smr15C1 y Smr15C2, respectivamente, sólo visibles después del tratamiento del RNA de partida con TAP. La secuenciación de varios de estos fragmentos de PCR localizó el sitio de inicio de transcripción de Smr15C1 en el nucleótido A 1.638.731 y el de Smr15C2 en la posición A 1.638.937 del cromosoma.



**Fig. I-14. Regiones genómicas de los genes sRNA identificados en *S. meliloti* 1021.** Este esquema resume las predicciones bioinformáticas y los resultados del mapeo experimental. Los genes *smr* están representados por flechas rojas, siendo la dirección de la flecha el sentido de la transcripción del gen. Los ORFs que flaquean a cada región intergénica se muestran con flechas discontinuas marcadas en negro. Los números indican la posición en el genoma de *S. meliloti* 1021. Los extremos 5' y 3' determinados experimentalmente mediante RACE se muestran encuadrados, siendo de color verde los correspondientes al extremo 5' y en azul a los del 3'. El extremo 3' de los sRNAs diferencialmente expresados se asignaron en la última Uridina de la serie de Us de cada terminador Rho-independiente, representado por un círculo azul encima de una línea vertical del mismo color. La flecha de punta blanca indica el punto de procesamiento interno de Smr7C. Las flechas simples muestran los posibles promotores  $\sigma^{70}$  y el posible sitio de unión de factores de transcripción tipo LysR en el promotor de *smr35B* se indica con una flecha de doble punta.

**-Smr35B:** el gel de poliacrilamida evidenció la presencia de un producto RACE mayoritario de 153 pb específico de este transcrito que no es visible tras la amplificación del cDNA obtenido de RNA control y que no se resuelve en agarosa. El análisis de varias secuencias independientes del mismo sitúa el inicio de transcripción en la posición A 577.730 del megaplásmido pSymB.

**-Smr45C:** la amplificación específica del cDNA de este transcrito, obtenido tanto de RNA tratado con TAP como del control, genera un producto mayoritario de 184 pb que, no obstante, es menos abundante cuando se utiliza RNA no defosforilado. La diferencia en la intensidad de ambas bandas es más evidente cuando los productos de PCR se resolvieron en geles de agarosa. La secuencia de este fragmento localizó el inicio de transcripción de Smr45C en el residuo cromosómico G 3.105.445.

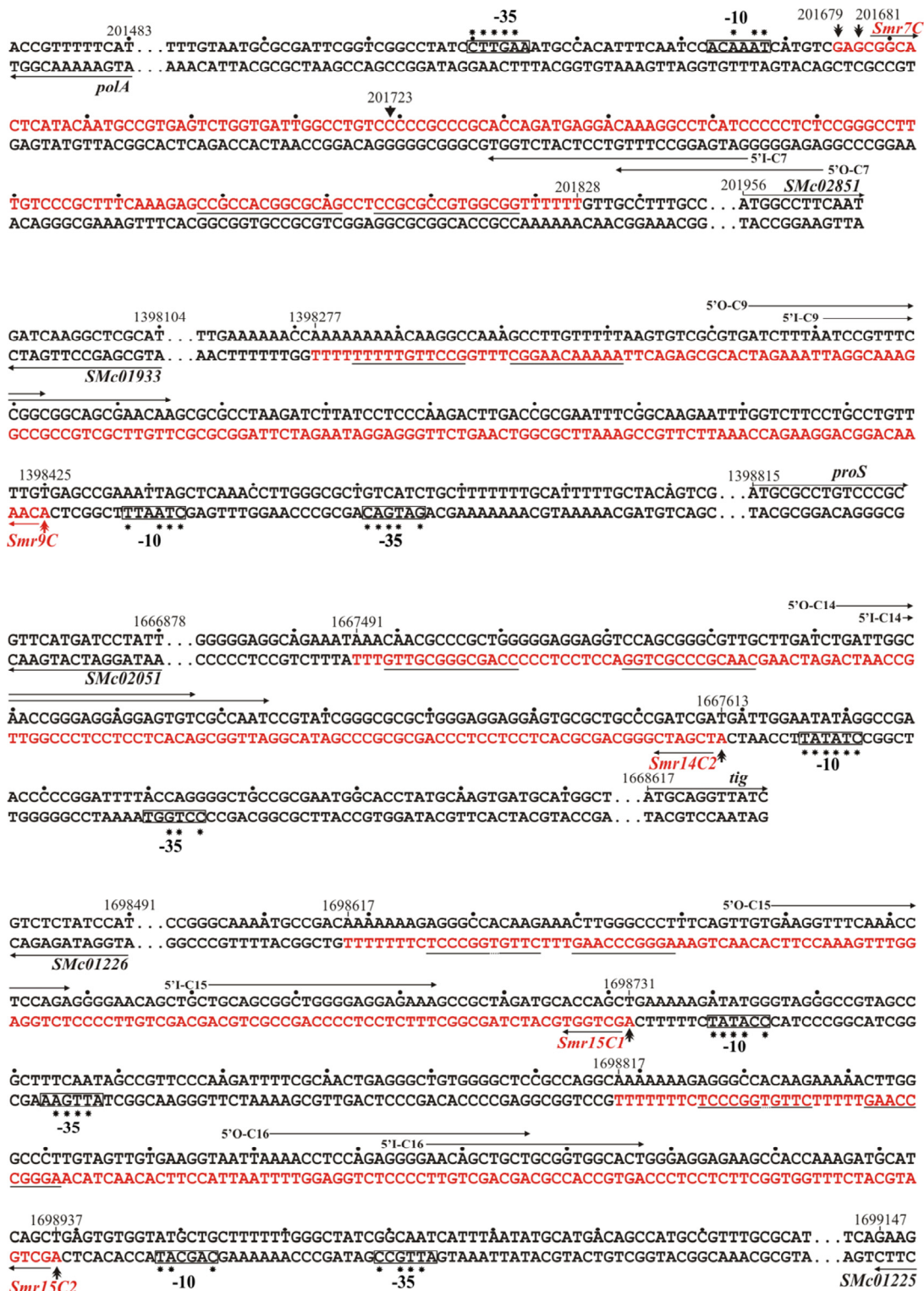
Las regiones genómicas de estos 7 genes, junto con sus características más relevantes se representan en la Fig. I-14.

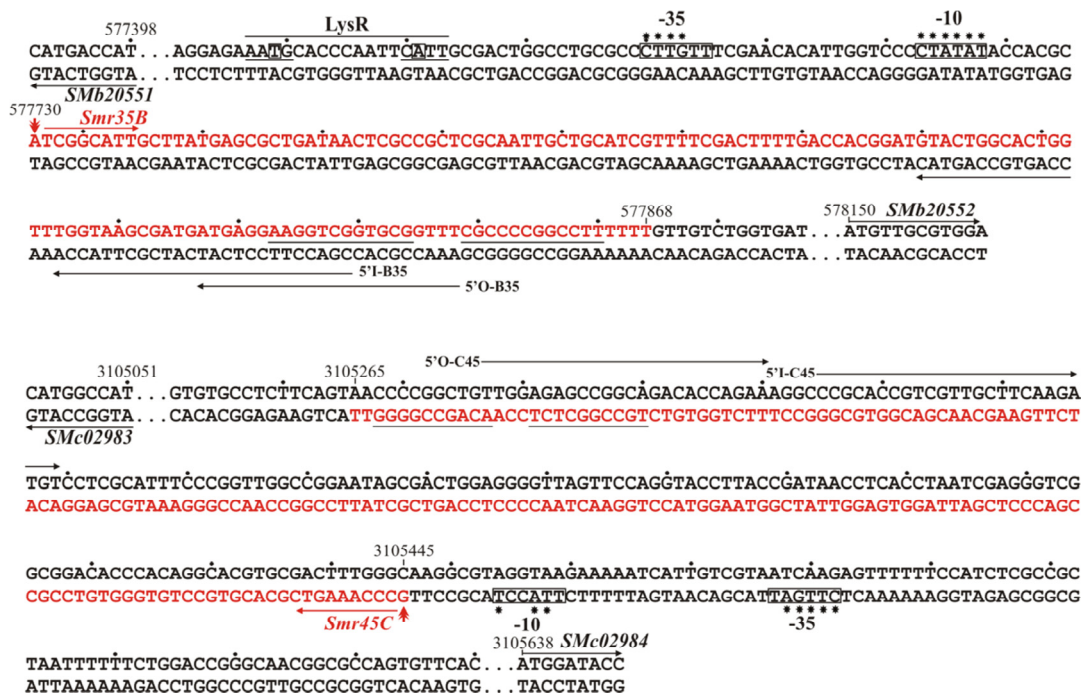
### I-3.3.3. PREDICCIÓN DE PROMOTORES Y TERMINADORES DE TRANSCRIPCIÓN

Las regiones genómicas que preceden a los extremos 5' de los transcritos Smr fueron inicialmente inspeccionadas en busca de la secuencia consenso **CTTGAC-N<sub>17</sub>-CTATAT** reconocida por el factor  $\sigma^{70}$  de la RNA polimerasa en *S. meliloti* y otras  $\alpha$ -proteobacterias filogenéticamente próximas (MacLellan *et al.*, 2006). Este análisis localizó cajas -10/-35 de este tipo adecuadamente espaciadas en las regiones promotoras de los siete genes *smr* identificados, todas en posiciones compatibles con el inicio de la transcripción desde los residuos nucleotídicos mapeados experimentalmente (Fig. I-14, I-15).

Además se buscaron específicamente motivos de 47 pb relacionados con los sitios consenso de unión del activador transcripcional NodD (cajas *nod*), en las regiones promotoras de los genes aparentemente inducidos por luteolina *smr15C1*, *smr35B* y *smr45C*. Aunque no se identificaron en ningún caso las típicas cajas *nod* conservadas, se encontró una repetición invertida situada alrededor del motivo **T-N11-A**, localizado 55 nt por encima del inicio de transcripción de Smr35B (Fig. I-15). Esta secuencia característica ha sido propuesta como posible sitio de unión de los activadores transcripcionales tipo LysR, como NodD (Goethals *et al.*, 1992).

La terminación de la transcripción de la gran mayoría de los sRNAs mejor caracterizados codificados en RIGs es independiente de la proteína Rho. Estos terminadores consisten en horquillas (estructuras tallo-lazo) sustentadas por el apareamiento de secuencias palindrómicas, que terminan con una serie de uridinas,





**Fig. I-15. Secuencias de los genes *smr*.** La secuencia nucleotídica de los genes *smr* se muestra en color rojo. Los genes flanqueantes y las coordenadas en el genoma de *S. meliloti* se indican en negro. Mediante flechas dobles en color rojo se muestra el inicio de transcripción de cada gen, determinado mediante 5' RACE. La flecha simple indica el sitio de procesamiento, del transcrito SmrC7. La posición de los cebadores externos e internos (5'0/5'1) para la amplificación por PCR anidada de los cDNAs en el RACE 5' se indican mediante flechas. Los extremos 3' se han asignado a la última uridina situada en el último tándem de uridinas situado después del tallo y bucle del terminador Rho-independiente. Las secuencias invertidas de cada terminador aparecen subrayadas. Las cajas promotoras -10 y -35 del factor  $\sigma^{70}$  se muestran encuadradas. El posible sitio de unión de un factor de transcripción LysR se indica en la región promotora del gen *smr35B*. Las líneas situadas bajo la secuencia de reconocimiento de LysR representan las secuencias invertidas alrededor del motivo T-N11-A.

que finalmente provocan el desacoplamiento de la RNA polimerasa del transcrito sintetizado. El programa TransTerm (Kingsford *et al.*, 2007) identificó terminadores Rho-independientes típicos para los siete transcritos Smr objeto de estudio, todos a una distancia de los respectivos sitios de inicio de transcripción en buena correlación con los tamaños aproximados inferidos de los experimentos de Northern (Fig. I-15). No obstante, la predicción del terminador de transcripción de Smr45C fue la de menor fiabilidad. Asumiendo el final de la transcripción de estos genes en el último residuo de la serie de Us típica de estos terminadores la longitud de éstos sRNAs sería: **Smr7C**, 148-150 nt (transcrito primario) y 106 nt (transcrito procesado); **Smr9C**, 149 nt; **Smr14C2**, 123 nt; **Smr15C1**, 115 nt; **Smr15C2**, 121 nt; **Smr35B**, 139 nt; y **Smr45C**, 181 nt.



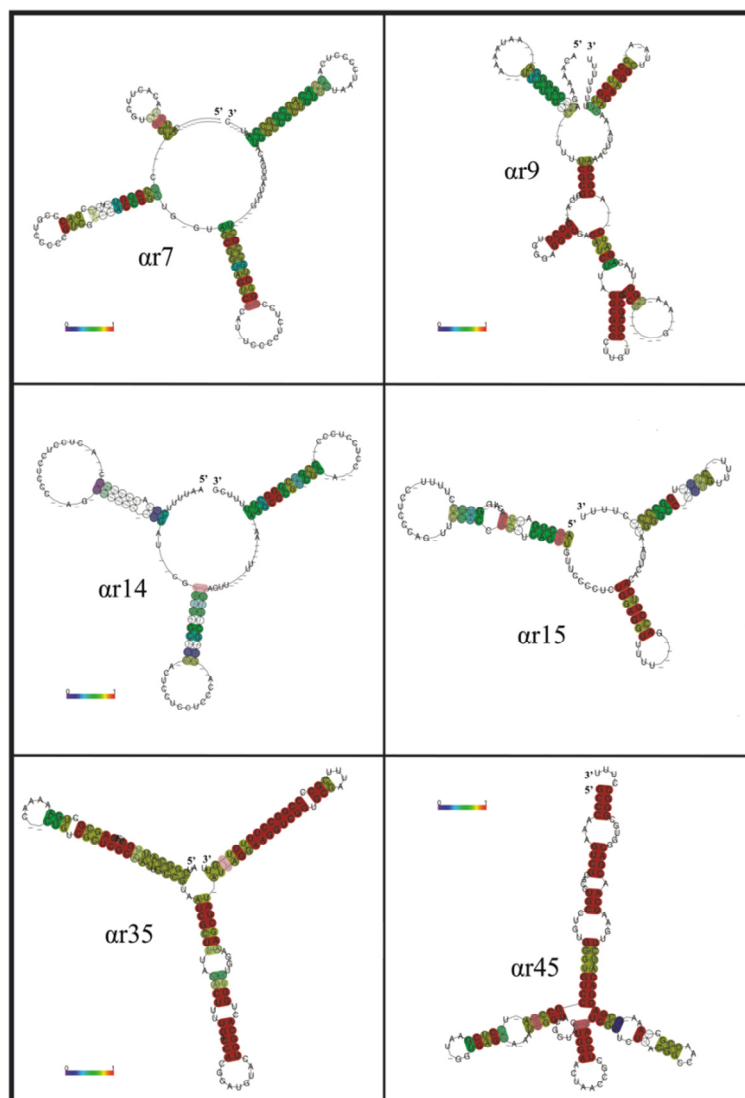
### I-3.3.4. CONSERVACIÓN DE LOS sRNAs Smr EN BACTERIAS

#### ***I-3.3.4.1. Las familias $\alpha$ de sRNAs: modelos de co-varianza y estructuras secundarias consenso.***

Las secuencias nucleotídicas de los siete transcritos Smr, determinadas experimentalmente en este trabajo, fueron inicialmente utilizadas para interrogar la base de datos Rfam con el programa Infernal (<http://www.sanger.ac.uk/Software/Rfam>). Esta búsqueda reveló homología estructural parcial de Smr15C1/Smr15C2 y Smr45C con los sRNAs denominados SuhB (modelo Rfam RF00519) y SpeF (RF00518), respectivamente, mientras que los restantes transcritos no aparecieron representados en esta base de datos. Concluimos, por tanto, que estas siete secuencias representan familias de sRNAs no descritas hasta ahora en bacterias y fueron comparadas mediante BLASTN con todos los genomas bacterianos depositados a fecha 20 de Abril de 2011 en bases de datos públicas (1.615 secuencias; <http://www.ncbi.nlm.nih.gov>). Las regiones genómicas que mostraron un 78%-89% de identidad en su secuencia nucleotídica primaria con los transcritos Smr fueron extraídas para generar alineamientos que se curaron manualmente para definir modelos de co-varianza (MC) de cada familia de sRNAs, con los que inferir las respectivas estructuras secundarias consenso. Como era esperado por su alta homología este análisis agrupó los transcritos Smr15C1 y Smr15C2 en la misma familia de sRNAs. La utilización de los seis MC para la identificación de nuevos miembros de cada familia de sRNAs en bacterias reveló que éstas sólo estaban representadas en  $\alpha$ -proteobacterias filogenéticamente próximas a *S. meliloti* y más específicamente en especies del orden Rhizobiales. Por consistencia con la nomenclatura de los sRNAs identificados en *S. meliloti*, las familias definidas por éstos se denominaron  $\alpha 7$ ,  $\alpha 9$ ,  $\alpha 14$ ,  $\alpha 15$ ,  $\alpha 35$  y  $\alpha 45$ , en alusión a  $\alpha$ -proteobacteria RNA.

Las estructuras secundarias consenso de cada una de las seis familias de sRNAs  $\alpha$  se muestran en la Fig. I-16. Todas presentan una disposición típica en subdominios en horquilla (tallo-lazo), generalmente interrumpidos por tallos internos y bucles de nucleótidos desapareados. Estas estructuras están sustentadas por un grado variable de covarianza, que es particularmente alto en el caso de los tres subdominios de las familias  $\alpha 7$ ,  $\alpha 14$  y  $\alpha 15$ , y los dominios en posición 5' de las familias  $\alpha 9$  y  $\alpha 35$ . Confirmando predicciones anteriores, en la mayoría de los casos, el dominio 3' está constituido por una horquilla, en la que el tallo es rico en apareamientos G-C, seguida de series largas de residuos de Uridina que en general están desapareadas, ajustándose así a las características estructurales de los

terminadores de transcripción independientes de la proteína Rho. La excepción sería la familia  $\alpha$ 45, en la que el tallo del dominio 3' está sustentado por apareamiento de nucleótidos muy conservados a nivel de secuencia primaria, pero sin embargo, no parece ser un terminador Rho-independiente típico.



**Fig. I-16. Estructura consenso de las familias de sRNA  $\alpha$ .** El código de color de los apareamientos de bases se corresponde al formato Vienna RNA. (Gruber *et al.*, 2008).

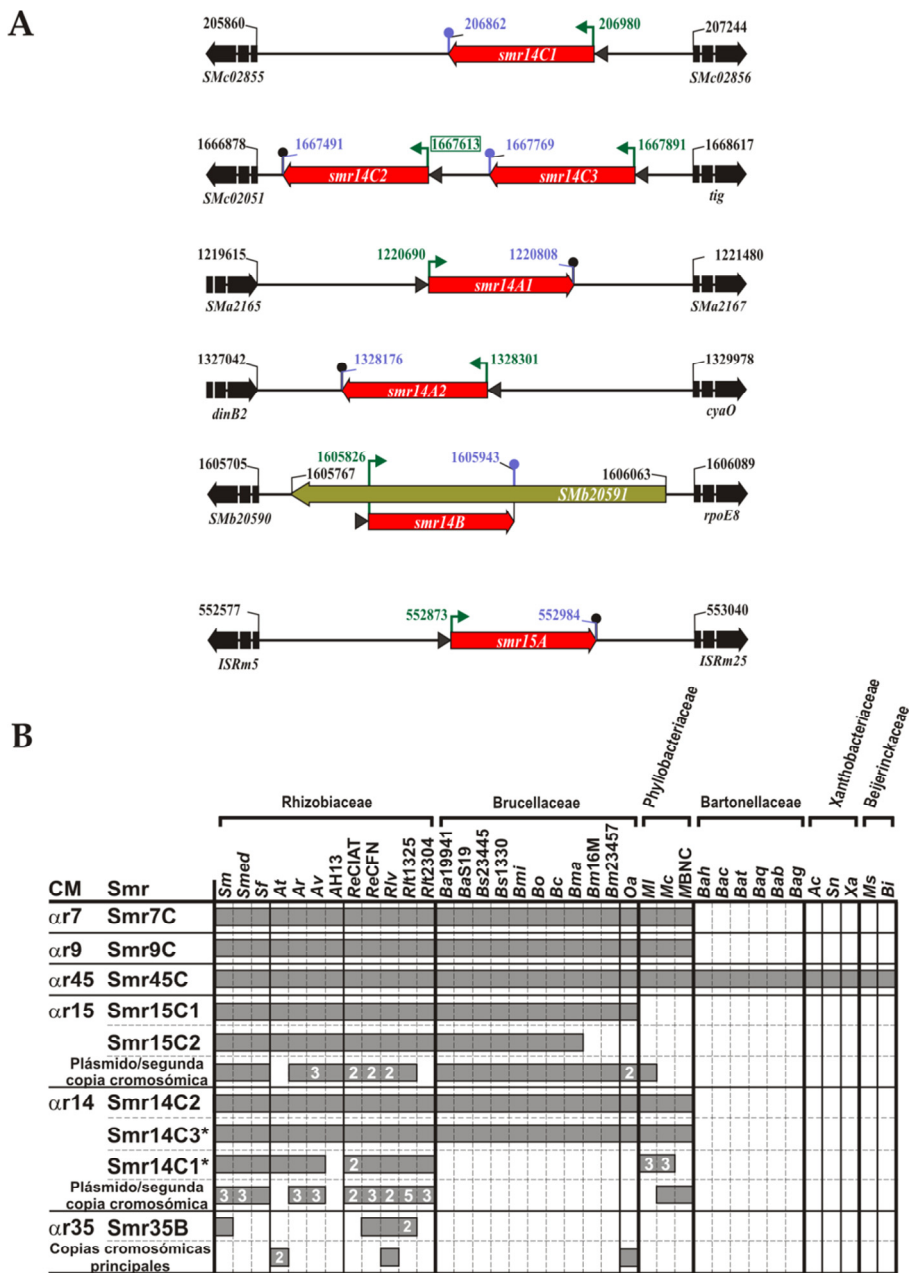
#### I-3.3.4.2. Distribución de las familias $\alpha$ de sRNAs en el orden Rhizobiales

Los MC que definen las familias  $\alpha$  fueron utilizados para investigar con detalle la presencia de estos sRNAs en especies bacterianas del orden Rhizobiales. Los resultados de esta comparativa se resumen en la Fig. I-17.

En primer lugar los modelos  $\alpha$ 14 y  $\alpha$ 15 identificaron nuevas copias de estos genes en el genoma de *S. meliloti* 1021, además de *smr15C1/smr15C2* y *smr14C2*, predichos por eQRNA/RNAz en nuestro *screening* inicial (Fig. I-17A). Una tercera copia del gen que codifica Smr15C se localizó en el megaplásmido pSymA (Smr15A) mientras que el modelo  $\alpha$ 14 identificó cinco copias adicionales de Smr14C2; dos codificadas en el Cromosoma (Smr14C1 y Smr14C3), dos en el plásmido pSymA (Smr14A1 y Smr14A2) y la quinta en pSymB (Smr14B). Al igual que Smr15C1/Smr15C2, los sRNAs Smr14C2 y Smr14C3 están codificados por genes dispuestos en tándem en la misma RIG. Todos los nuevos sRNAs predichos por los MC de estas dos familias están también codificados en RIGs del genoma de *S. meliloti* 1021 con la excepción de Smr14B, cuyo *locus* es complementario (antisentido) a una parte interna de la secuencia codificante de la proteína SMb20591.

En bacterias del orden Rhizobiales (Fig. I-17B) el patrón de distribución dominante está representado por las familias de sRNAs  $\alpha$ 7,  $\alpha$ 9 y  $\alpha$ 14 cuyos MC identificaron miembros en las tres familias de bacterias que incluyen las especies de mayor parentesco con *S. meliloti*; Rhizobiaceae, Brucellaceae y Phyllobacteriaceae. La familia  $\alpha$ 15 está también muy distribuida en este orden pero carece de representantes codificados en el cromosoma de especies del género *Mesorhizobium* (familia Phyllobacteriaceae). La distribución más amplia correspondió a la familia de sRNAs  $\alpha$ 45 cuya presencia se extiende a especies de otras tres familias bacterianas de relación filogenética más distante con *S. meliloti*; Bartonellaceae, Xanthobacteriaceae y Beijerinckaceae. Todos los miembros de  $\alpha$ 7,  $\alpha$ 9 y  $\alpha$ 45 están codificados por genes presentes en copia única en los respectivos cromosomas.

Sin embargo, al igual que se ha descrito en *S. meliloti*, los genes que codifican representantes de las familias de sRNAs  $\alpha$ 14 y  $\alpha$ 15 aparecen en número de copias variable en cada genoma individual, muchos de ellos localizados en replicones extracromosómicos, como los plásmidos accesorios en bacterias de las familias Rhizobiaceae y Phyllobacteriaceae o segundos cromosomas en especies del género *Brucella*. Los miembros de la familia  $\alpha$ 15 aparecen en dos copias cromosómicas en 19 bacterias pertenecientes a las familias Rhizobiaceae y Brucellaceae. Estos dos genes son contiguos en RIGs de los genomas de bacterias de la familia Rhizobiaceae



**Fig. I-17. Conservación de los sRNAs Smr de *S. meliloti* 1021 en el orden Rhizobiales. A) Mapas de las nuevas copias identificadas en el genoma de *S. meliloti* por los modelos  $\alpha$ r14 y  $\alpha$ r15. B) Los modelos de covarianza se generaron utilizando como entrada la secuencia de los sRNA de *S. meliloti* listado a la izquierda. Las nuevas copias cromosómicas del gen *smr14C* se indican con asteriscos. Todas las especies bacterianas con representantes de las familias  $\alpha$ r RNA se indican en la parte superior de la Fig., agrupados por familias, i.e. Rhizobiaceae, Brucellaceae, Phyllobacteriaceae, Bartonellaceae, Xanthobacteriaceae y Beijerinckaceae. La identificación de las especies es como sigue: Sm, *S. meliloti* 1021; Smed, *S. medicae* WSM419; Sf, *S. fredii* NGR234; At, *Agrobacterium tumefaciens* C58; Ar, *A. radiobacter* K84; Av, *A. vitis* S4; AH13, *A. sp.* H13-3; ReCIAT, *Rhizobium etli* CIAT652; ReCFN, *R. etli* CFN42; Rlv, *R. leguminosarum* bv. viceae 3841; Rlt1325, *R. leguminosarum* bv. trifolii WSM1325; Rlt2304, *R. leguminosarum* bv. trifolii WSM2304; Ba19941, *Brucella abortus* bv. 1 9-941; BaS19, *B. abortus* S19; Bs23445, *B. suis* ATCC23445; Bs1330, *B. suis***

1330; *Bmi*, *B. microti* CCM4915; *Bo*, *B. ovis* ATCC25840; *Bc*, *B. canis* ATCC 23365; *Bma*, *B. melitensis* bv. abortus 2308; *Bmi*16M, *B. melitensis* bv. 1 16M; *Bm*23457, *B. melitensis* ATCC23457; *Oa*, *Ochrobactrum anthropi* ATCC49188; *MI*, *Mesorhizobium loti* MAFF303099; *Mc*, *M. ciceri* bv. biserrulae WSM1271; *MBNC*, *M. sp.* BNC1; *Bah*, *Bartonella henselae* Houston-1; *Bac*, *B. clarridgeiae* 73; *Bat*, *B. tribocorum* CIP105476; *Baq*, *B. quintana* Toulouse; *Bab*, *B. bacilliformis* KC583; *Bag*, *B. grahamii* as4aup; *Ac*, *Azorhizobium caulinodans* ORS571; *Sn*, *Starkeya novella* DSM506; *Xa*, *Xanthobacter autotrophicus* Py2; *Ms*, *Methylocella silvestris* BL2, *Bi*, *Beijerinckia indica* subsp. *indica* ATCC9039. Las barras indican la distribución de cada familia de sRNAs en estas especies bacterianas. En el caso de la existencia de más de una copia de cada gen sRNA, el número de copias cromosómicas y extracromosómicas también se indica.

mientras que en las especies del género *Brucella* en las que coexisten se localizan en posiciones distantes en el cromosoma principal (cromosoma I) de estas bacterias. No apareció segunda copia cromosómica de estos genes en los genomas de *B. melitensis* bv. abortus 2308, *B. melitensis* bv. 1 16M y *Ochrobactrum anthropi* ATCC49188. Sin embargo, con la excepción de *A. tumefaciens* C58 y *R. leguminosarum* bv. trifolii 2304, al menos una copia adicional del gen que codifica sRNAs de la familia  $\alpha$ 15 fue identificada en elementos extracromosómicos de todas estas bacterias.

La familia de sRNAs  $\alpha$ 14 presentó un patrón de distribución incluso más complejo que  $\alpha$ 15 en el orden Rhizobiales. Al igual que en *S. meliloti* 1021 se identificaron dos copias de genes  $\alpha$ 14 dispuestas en tándem en RIGs de todas las especies y cepas del género *Mesorhizobium* y las demás especies del género *Sinorhizobium*. Sin embargo, en *O. anthropi* ATCC49188 y en especies de los géneros *Agrobacterium* y *Brucella* las segundas copias cromosómicas de este gen se localizan en posiciones alejadas de la primera. En la mayoría de especies pertenecientes a las familias Rhizobiaceae y Phyllobacteriaceae se identificaron copias adicionales de estos genes en número muy variable (hasta seis copias más en *R. leguminosarum* bv. trifolii WSM1325), tanto en el cromosoma principal como en plásmidos accesorios.

Finalmente, la familia de sRNAs  $\alpha$ 35 evidenció una representación más limitada y dispersa no sólo a nivel de especies bacterianas sino incluso de cepas. Sólo siete genes sRNA fueron identificados por el MC  $\alpha$ 35 como genuinos representantes de esta familia. Tres de ellos se localizan en los cromosomas de *A. tumefaciens* C58, *O. anthropi* ATCC49188 y *R. leguminosarum* bv. viceae 3841, mientras que los cuatro restantes son extracromosómicos y fueron identificados en los plásmidos p42f, pRL11 y pR1132502/pR1132504 de *R. etli* CFN42, *R. leguminosarum* bv. viceae 3841 y *R. leguminosarum* bv. trifolii 1325, respectivamente.

### I-3.4. IDENTIFICACIÓN DE NUEVOS sRNAs EN *S. MELILOTI*, MEDIANTE SECUENCIACIÓN MASIVA

En esta segunda parte del capítulo se presenta la co-inmunoprecipitación con la chaperona de RNA Hfq como protocolo genérico para aislar subpoblaciones de

transcritos enriquecidas en sRNAs y la subsiguiente secuenciación masiva (RNA-Seq) de los mismos como aproximación experimental de alto rendimiento para la generación de un inventario de sRNAs representativo del RNoma bacteriano. No obstante, más allá de la mera catalogación de nuevos sRNAs, algunos hallazgos de este estudio tienen evidentes implicaciones en la biología de esta proteína y en la de los propios sRNAs que interactúan con la misma, por lo que serán presentados y discutidos con detalle en los Capítulos 2 y 3.

#### I-3.4.1. GENERACIÓN DE LA CEPA *S.MELILOTI* 2011 B23001 $hfq^{FLAG}$

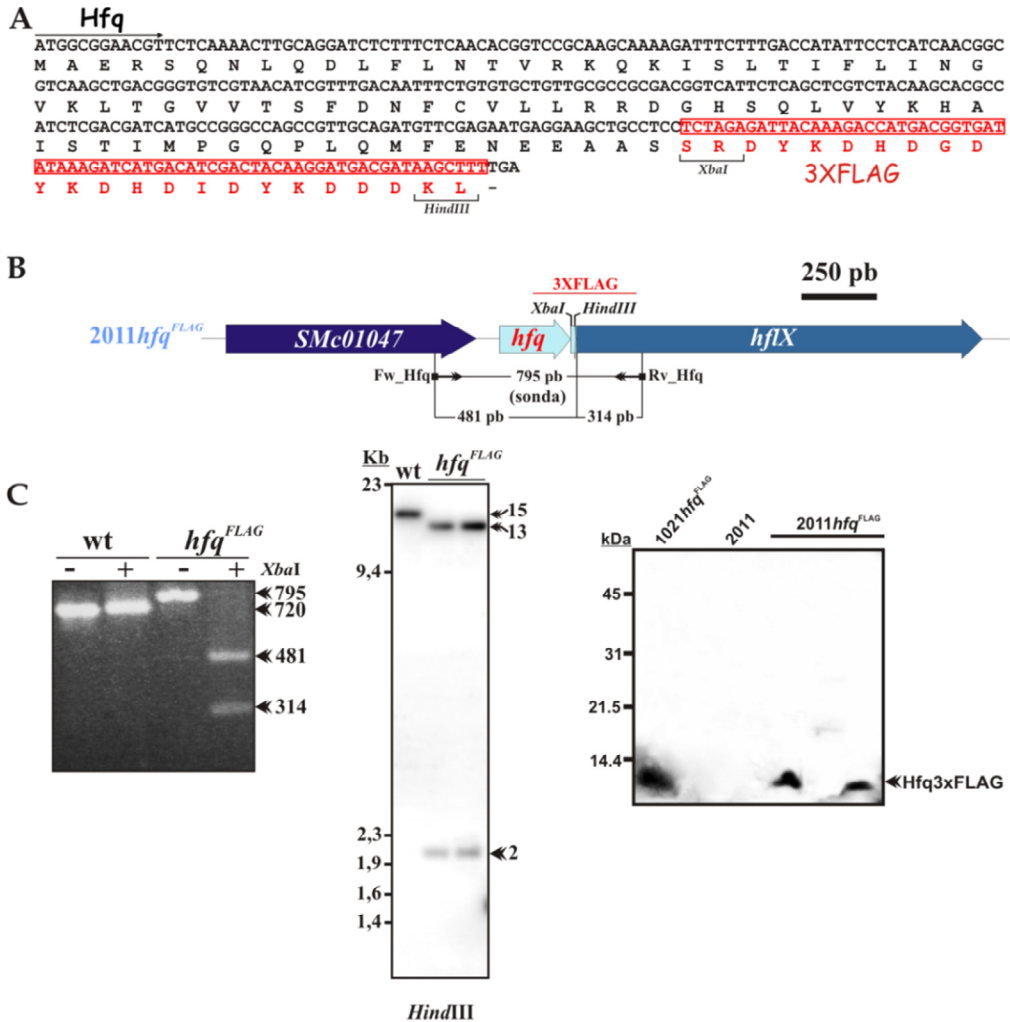
Para el aislamiento de la fracción de RNAs que se unen a la proteína Hfq, se construyó la cepa *S. meliloti* 2011B23001 $hfq^{3XFLAG}$  (en adelante *S. meliloti* 2011 $hfq^{FLAG}$  o 2011 $hfq^{FLAG}$ ), como derivada de la silvestre *S. meliloti* 2011B23001, en la que se modificó el gen cromosómico que codifica esta proteína para expresar una versión de la misma marcada con tres unidades del epítipo FLAG en su extremo C-terminal (Fig. I-18A, I-18B).

La correcta modificación del gen *hfq* en la cepa 2011 $hfq^{FLAG}$  se verificó mediante PCR, hibridación Southern de DNA genómico y Western blot (Fig. I-18C). La amplificación de la región adyacente al gen *hfq* en la cepa silvestre *S. meliloti* 1021B23001, con la pareja de cebadores Fw\_Hfq/Rv\_Hfq, generó un fragmento de DNA de 720 nt que en la cepa derivada fue de 795 nt en perfecta correlación con la extensión de 75 nt introducida en el extremo 3' del gen. El producto de amplificación del alelo modificado fue sensible a la digestión con *XbaI*, cuyo sitio de reconocimiento, junto con el de *HindIII*, fue introducido en el proceso de generación de 2011 $hfq^{FLAG}$ . La hibridación de DNA total de ambas cepas digerido con *HindIII* y utilizando este mismo amplicón de la cepa derivada como sonda reveló el perfil esperado con una banda única de 15 kb correspondiente a la cepa silvestre y dos, de 13 y 2 kb correspondientes a la marcada con el FLAG. Finalmente, se verificó mediante Western-blot que los anticuerpos anti-FLAG reconocían una proteína recombinante de aproximadamente 12kDa en lisados celulares de la cepa 2011 $hfq^{FLAG}$ , pero no en los de la silvestre.

#### I-3.4.2. ANÁLISIS FENOTÍPICO DE LA CEPA *S. MELILOTI* 2011 $hfq^{FLAG}$

Normalmente se asume que la adición del FLAG al extremo C-terminal de Hfq no altera fenotipos esenciales de las bacterias ni la capacidad de la proteína para unir RNA (Pfeiffer *et al.*, 2007; Sittka *et al.*, 2008; Berghoff *et al.*, 2011). No obstante y dado que Hfq es un regulador global de la expresión génica, se realizó un análisis fenotípico básico de la cepa *S. meliloti* 2011 $hfq^{FLAG}$ , para verificar que esta

modificación no comprometía el crecimiento de *S. meliloti* en medio TY ni su rendimiento simbiótico.



**Fig. I-18. Verificación de la modificación introducida en el gen cromosómico *hfq* en la cepa *S. meliloti* 2011 *hfq*<sup>FLAG</sup>.** A) Secuencia de la proteína recombinante Hfq<sup>3XFLAG</sup>. B) Diagrama de la región genómica correspondiente al gen *hfq*, con indicación de la modificación del extremo C-terminal de la proteína (3XFLAG) e información relevante para la interpretación de los paneles inferiores. C) Aproximaciones experimentales realizadas para la verificación de la mutación: gel de agarosa correspondiente a la electroforesis de los productos de PCR (izquierda); southern-blot del DNA genómico digerido con *XbaI* de la cepa salvaje y dos colonias de la cepa *hfq*<sup>FLAG</sup> (centro); western-blot con anticuerpos anti-FLAG, correspondiente a los extractos proteicos de las mismas cepas sometidas a southern-blot. Se incluye además en el primer carril un control positivo, correspondiente a la cepa *hfq*<sup>FLAG</sup> en *S. meliloti* 1021.

En primer lugar se monitorizó el crecimiento de la cepa salvaje *S. meliloti* 2011 y la cepa modificada, en medio completo TY durante 30 horas, en agitación y a 28°C, tomándose medidas de la densidad óptica (D.O.) a 600 nm cada dos horas. Las

respectivas curvas de crecimiento, representadas en la Fig. I-19A, no muestran diferencias evidentes entre ambas cepas.

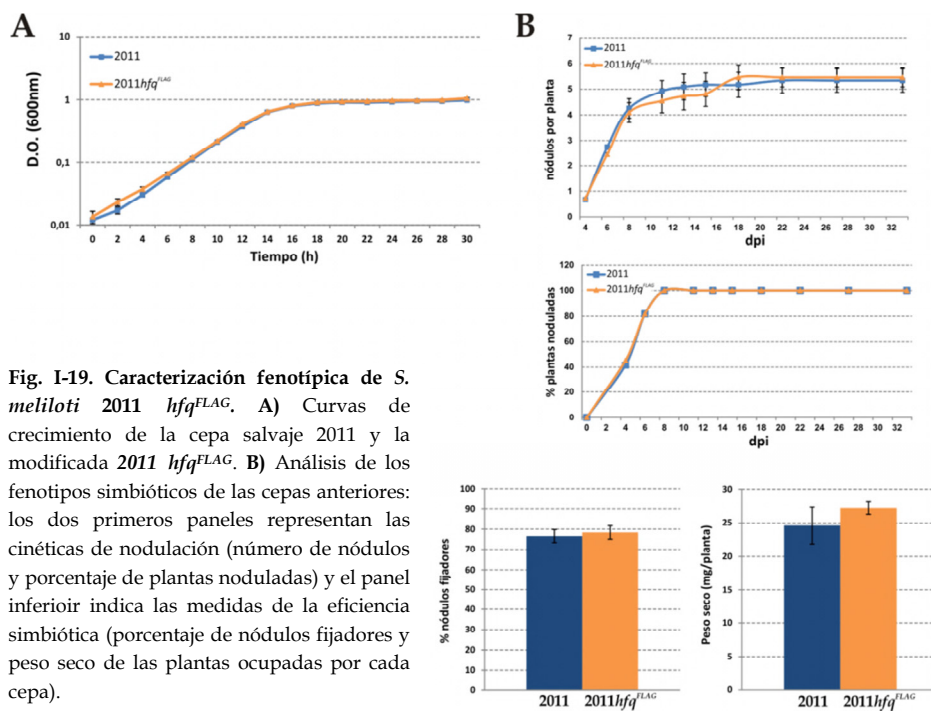


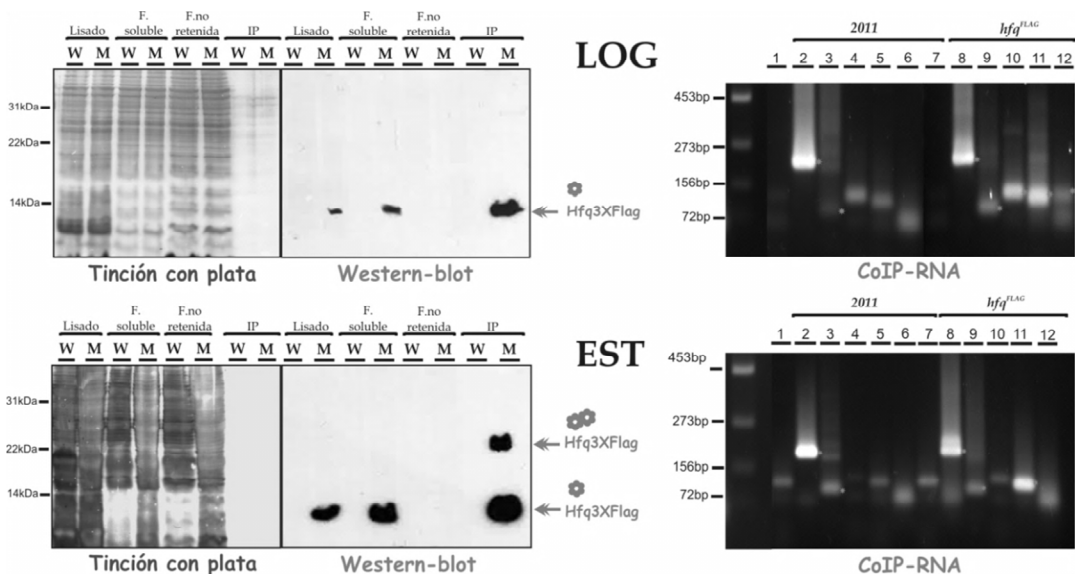
Fig. I-19. Caracterización fenotípica de *S. meliloti* 2011 *hfq*<sup>FLAG</sup>. A) Curvas de crecimiento de la cepa salvaje 2011 y la modificada 2011 *hfq*<sup>FLAG</sup>. B) Análisis de los fenotipos simbióticos de las cepas anteriores: los dos primeros paneles representan las cinéticas de nodulación (número de nódulos y porcentaje de plantas noduladas) y el panel inferior indica las medidas de la eficiencia simbiótica (porcentaje de nódulos fijadores y peso seco de las plantas ocupadas por cada cepa).

Respecto al fenotipo simbiótico (Fig. I-19B), se determinaron inicialmente la infectividad y cinética de nodulación de cada cepa en *M. sativa* (alfalfa). Para ello se inocularon dos sets de 12 plantas crecidas en cultivo hidropónico y en ausencia de fuente de nitrógeno, uno con la cepa silvestre y otro con la cepa que contiene el alelo modificado, y se monitorizaron el número de nódulos por planta (infectividad) y el porcentaje de plantas noduladas (cinética de nodulación) durante 32 días. No se detectaron diferencias significativas entre las dos cepas respecto a ambos parámetros. Además, como estimación indirecta de la fijación biológica de nitrógeno en cada caso se contó el número de nódulos rosáceos (fijadores) y se determinó el peso seco de las plantas al final del experimento. Tampoco se observaron diferencias entre las dos cepas, obteniéndose un porcentaje de nódulos fijadores cercano al 80% y un peso seco similar de aproximadamente 25 mg/planta, en ambos casos.

#### I-3.4.3. CO-IP DE LA FRACCION DE RNAS ASOCIADOS A Hfq<sup>FLAG</sup>

La inmunoprecipitación (IP) de Hfq<sup>FLAG</sup> se llevó a cabo en lisados celulares de la cepa *S. meliloti* 2011*hfq*<sup>FLAG</sup>, cultivada en medio completo TY hasta las fases logarítmica y estacionaria, utilizando anticuerpos monoclonales anti-FLAG. La cepa





**Fig. I-20. Monitorización de la IP de Hfq<sup>FLAG</sup>.** Fase logarítmica (LOG) arriba, fase estacionaria (EST) abajo. Tinción con plata y Western-blot de las distintas fracciones recogidas a lo largo del experimento (paneles sección izquierda). Las distintas fracciones se separaron en dos geles equivalentes SDS-PAGE al 15%, tiñéndose con plata uno (izquierda) y transfiriéndose el otro a una membrana de polivinilo (Sigma) (derecha). Las bandas observadas fueron obtenidas después de 5 minutos de exposición en una película Kodak. W=cepa control; M= cepa problema. Primera valoración del CoIP RNA sometido a RNA-Seq (paneles sección izquierda). Geles de agarosa al 2%, correspondientes a la RT-PCR de las muestras del CoIP RNA de las dos cepas (2011=cepa control; *Hfq<sup>FLAG</sup>*= cepa problema). Los asteriscos en el gel muestran las bandas de amplificación específicas. 1=control sin RT; 2, 8=RNA 16S; 3, 9=RNA 5S; 4, 10= Smr7C; 5, 11=Smr15C1; 6, 12=Smr15C2.

silvestre 2011 se sometió al mismo protocolo como control de la especificidad del proceso. La IP se monitorizó mediante Western-blot, con el mismo anticuerpo, de distintas fracciones proteicas (Fig. I-20, sección izquierda); lisado bacteriano, o extracto proteico total, fracción de proteínas solubles obtenida por centrifugación del extracto anterior (la sometida a inmunoprecipitación), fracción de proteínas no retenidas por el anticuerpo y proteína(s) inmunoprecipitada(s).

Los Western-blots (uno por cada condición de cultivo) revelaron en todos los casos la señal esperable específica de la proteína Hfq<sup>FLAG</sup> en las fracciones proteicas correspondientes a la cepa modificada. No se observó señal alguna en los concentrados de las proteínas no unidas al anticuerpo, lo que sugiere que la totalidad de la proteína recombinante presente en los extractos de partida quedó retenida por la resina y fue finalmente recuperada (último carril de cada Western). El nitrato de plata, sin embargo, no fue suficientemente sensible para detectar proteína alguna en las fracciones resultantes de la IP. Por otra parte, el anticuerpo anti-FLAG reconoció una segunda proteína de aproximadamente 24 kDa en esta fracción cuando

se obtuvo de bacterias cultivadas hasta fase estacionaria que, no obstante, podría corresponder a un dímero de Hfq<sup>FLAG</sup>.

El CoIP RNA que se sometió a RNA-Seq se obtuvo tanto de la cepa silvestre (control de especificidad) como de su derivada mediante extracción con fenol-cloroformo a partir de los complejos RNA-proteína obtenidos mediante IP de Hfq<sup>FLAG</sup> en lisados de bacterias en fase logarítmica y exponencial de crecimiento en medio TY. De cada cepa y condición de cultivo se hicieron dos réplicas biológicas obteniéndose, por tanto, un total de 8 preparaciones de CoIP-RNA cuya concentración se valoró individualmente en NanoDrop (Tabla I-6).

Antes de enviar las muestras para su secuenciación se hizo una primera valoración del contenido en sRNAs. Para ello, una pequeña fracción de cada muestra se sometió a reverso-transcripción (RT) y el cDNA se amplificó con oligonucleótidos de varios RNAs control (ver apartado de oligos); los ribosómicos 16S y 5S, y los sRNAs Smr7C, Smr15C1 y Smr15C2 (Fig. I-20, sección derecha). En ningún caso las amplificaciones del CoIP RNA no sometido a RT revelaron contaminación de las muestras por DNA. Además, se detectó señal para los RNAs ribosómicos en ambas cepas y en las dos fases de crecimiento. Esto podría sugerir una asociación inespecífica de los RNAs 16S y 5S, posiblemente por su gran abundancia, a la resina de la IP. Por el contrario, sí se observa una diferencia de intensidades en las bandas correspondientes a los sRNAs entre la cepa control y la problema, y en algún caso, entre las dos fases de crecimiento.

Tabla I-6. CoIP RNAs obtenidos.

Muestra <sup>1</sup>	Concentración (ng/μl)	Volumen (μL)	RNA total (ng)
Control LOG_1	28,91	10	289,1
Control LOG_2	30,68	10	306,8
Problema LOG_1	47,87	10	478,7
Problema LOG_2	90,45	10	904,5
Control EST_1	19,06	10	190,6
Control EST_2	27,72	10	277,2
Problema EST_1	35,58	10	355,8
Problema EST_2	91,61	10	916,1

<sup>1</sup> El número en cada muestra indica la réplica biológica de los cultivos en fase logarítmica (LOG) y estacionaria (EST).

#### I-3.4.4. ANÁLISIS DE LOS DATOS DE SECUENCIACIÓN

Las cuatro preparaciones de Co-IP RNA de la cepa control *S. meliloti* 2011 se mezclaron para obtener una única muestra identificada como **Control\_2011**, mientras

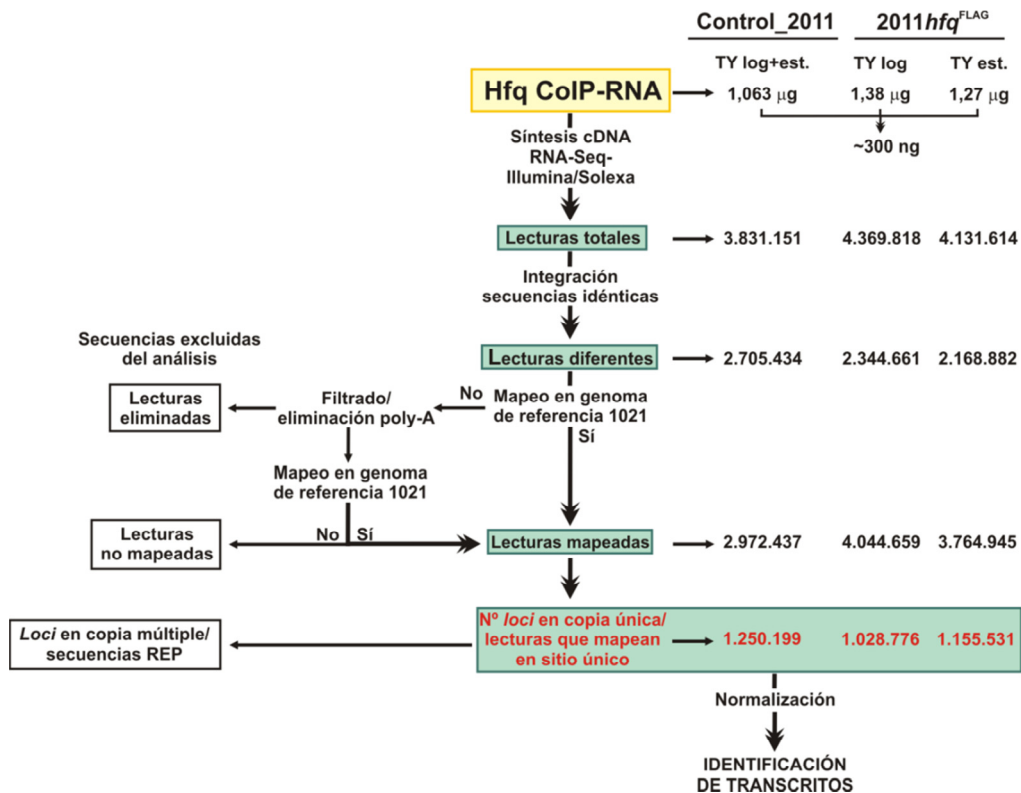
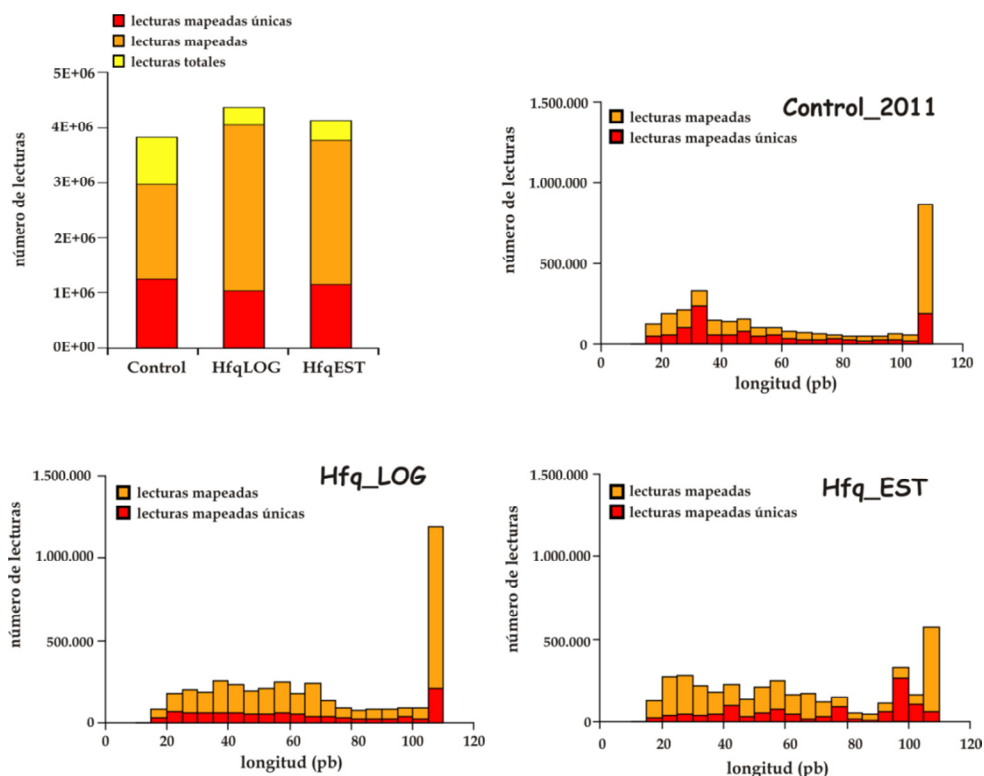


Fig. I-21. Diagrama resumen de las diferentes etapas del procesamiento de los datos de secuenciación del CoIP RNA. Explicación en el texto.

que las muestras de CoIP-RNA denominadas **Hfq\_Log** y **Hfq\_Est** resultaron de la mezcla de las preparaciones correspondientes a los dos cultivos independientes de 2011 $hfq^{FLAG}$  en cada condición (Fig. X). La cantidad de RNA en cada preparación fue de 1,063  $\mu$ g (Control\_2011), 1,38 (Hfq\_Log) y 1,27  $\mu$ g (Hfq\_Est), de los que 300 ng en cada caso fueron sometidos a reverso transcripción y secuenciación en plataforma Illumina/Solexa (RNA-Seq). La secuenciación generó, por tanto, tres bancos de lecturas, que para la muestra Control\_2011 fueron 3.831.151, 4.369.818 para Hfq\_Log y 4.131.614 para Hfq\_Est. La integración de las lecturas con secuencias idénticas (i.e. reducción a una única lectura) generó ficheros con 2.705.434 lecturas para el control, 2.344.661 para Hfq\_Log y 2.168.882 para Hfq\_Log. Estas secuencias se utilizaron para interrogar el genoma de *S. meliloti* 1021 de forma que después de esta primera comparación las que no mapearon se sometieron a un proceso de filtrado (*trimming*) para la eliminación de secuencias poli-A y otros artefactos reconocibles y se volvieron a mapear en el genoma de la bacteria, generando un nuevo conjunto de lecturas mapeadas, que se integró con las anteriores para obtener los ficheros de las "lecturas mapeadas" totales (i.e. total de loci representados en el CoIP-RNA), que resultaron

ser 2.972.437 para la condición control, 4.044.659 para Hfq<sup>LOG</sup> y 3.764.945 para Hfq<sup>EST</sup>. Dado que en muchos casos una misma secuencia mapeó en múltiples *loci* (i.e. secuencias repetidas o *repeats* de difícil interpretación), cada fichero se sometió a un filtrado para la eliminación de este tipo de secuencias, quedando finalmente para su posterior análisis el conjunto de lecturas de sitio o mapeo único, que fueron 1.250.199, 1.028.776 y 1.155.531 para el control, Hfq\_Log y Hfq\_Est, respectivamente (Fig. I-21).

La longitud de estas lecturas estuvo comprendida entre los 20 y los 108 nt, correspondiendo el porcentaje mayoritario de ellas a las de mayor tamaño (108 nt) en los bancos control y Hfq\_Log, mientras que en Hfq\_Est la distribución fue más heterogénea, con una mayoría de lecturas de tamaño ligeramente inferior a los 100 nt (Fig. I-22).



**Fig. I-22. Análisis estadístico de las lecturas obtenidas mediante Illumina/Solexa.** Representación de la proporción de las diferentes categorías de lecturas mapeadas en los tres bancos de cDNA (panel superior izquierdo). Distribución de los rangos de tamaño de las lecturas mapeadas y mapeadas únicas en las tres librerías (resto de paneles).

### **Clasificación de las regiones identificadas por secuenciación**

Las regiones con una cobertura y longitud mínimas de 10 lecturas y 20 nt, respectivamente, y factor de enriquecimiento 5 o superior, respecto a su representación en la librería control fueron catalogadas como RNAs asociados a Hfq. tRNAs y rRNAs aparecieron representados en cada librería individual, pero el enriquecimiento calculado para los mismos en CoIP-RNA de 2011hfq<sup>FLAG</sup> respecto al control fue siempre inferior a cinco. Las regiones asociadas a Hfq fueron clasificadas en seis categorías de acuerdo a su posición relativa respecto a los ORFs y loci RNA

anotados, utilizando como referencia la última anotación del genoma de *S. meliloti* 1021, que ya incluye los 1.125 sRNAs identificados y catalogados en Schlüter *et al.* (2010) mediante RNA-Seq (Fig. I-22):

(1) **Trans-sRNAs**: Regiones localizadas entre ORFs distantes más de 60 pb de la ATG anotada y más de 20 pb respecto al codón de terminación.

(2) **Regiones 5'/3'UTR**: agrupan las regiones que se sitúan total o parcialmente en las primeras 60 pb por encima de la ATG (5'-UTR) o 20 pb por debajo del codón de terminación (3'-UTR) del ORF de referencia, sin solapar con éste y transcritas en su misma dirección.

(3) **Cis-mRNA líder**: esta categoría corresponde a cualquier región que se encuentre identificada como tal en la base de datos GenDB. Abarcan todo o parte de la región 5'-UTR del gen anotado, solapan con el ORF y se transcriben en su misma dirección. Corresponden a sRNAs generados por atenuación o parada de la transcripción del mRNA, transcripción desde promotores alternativos o procesamiento del mRNA (Schlüter *et al.* (2010).

(4) **mRNAs sentido**: cualquier región que, comenzando desde la ATG anotada o por debajo de ésta, solapa total o parcialmente con la secuencia del ORF y se transcribe en la dirección de éste.

(5) **Regiones 5'/3'-UTR antisentido**: son las de idéntica localización a las 5'/3'-UTRs pero transcritas en dirección opuesta al ORF de referencia.

(6) **Cis-sRNAs antisentido**: regiones que solapan en total o parcialmente con el ORF de referencia y se transcriben en dirección opuesta a éste.

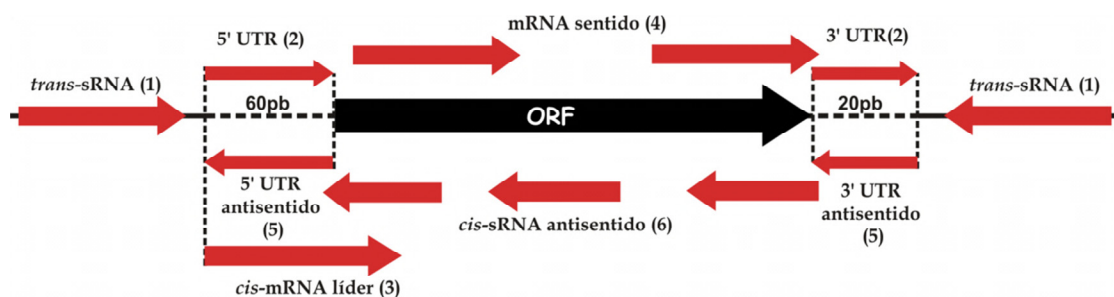


Fig. I-23. Clasificación de las categorías correspondientes a las distintas regiones identificadas mediante RNA-Seq. Representación de la proporción de las diferentes categorías de lecturas mapeadas en los tres bancos de cDNA (panel superior izquierdo). Distribución de los rangos de tamaño de las lecturas mapeadas y mapeadas únicas en las tres librerías (resto de paneles).

La distribución de estas regiones en los bancos de CoIP-RNA preparados a partir de lisados de *S. meliloti* 2011hfq<sup>FLAG</sup> en las fases exponencial y estacionaria de crecimiento se ilustra en la Fig. X. En el banco Hfq\_Log se identificaron un total de 3.149 regiones, de las que el 86% corresponden a mRNAs sentido (2.710 regiones) (Fig. I-24, sección izquierda). El 14% restante se catalogaron mayoritariamente como *trans*-sRNAs (162) y *cis*-sRNAs antisentido (151), seguidos de las regiones 5'/3'-UTR (66) y los *cis*-mRNA líder (50), mientras que las regiones 5'/3'-UTR antisentido fueron minoritarias en el banco (10). Por su parte, en el banco Hfq\_Est aparecieron representadas un menor número de regiones asociadas a Hfq (1.493), siendo también los mRNA sentido la categoría predominante, representando un 82% del total (1.224 regiones) (Fig. I-24, sección derecha). El 18% restante se distribuyó de la siguiente forma: 99 *cis*-sRNAs antisentido, 78 *trans*-sRNAs, 43 regiones 5'/3'-UTR, 33 *cis*-mRNA líder y 9 regiones 5'/3'-UTR antisentido.

Estas dos listas fueron comparadas para identificar aquéllas regiones representadas en ambos bancos, que fueron finalmente integradas para generar un catálogo único de regiones de *S. meliloti* 1021 asociadas a Hfq, expresadas en fase logarítmica y/o exponencial de crecimiento. El número y distribución por categorías de estas regiones fueron: mRNAs sentido (3.805 regiones), *cis*-sRNAs antisentido (234), *trans*-sRNAs (226), 5'/3'-UTRs (108), *cis*-mRNA líder (73) y 3'/5'-UTRs antisentido (16).

La caracterización precisa de cualquiera de las regiones mapeadas en mRNAs anotados (mRNAs sentido) y su catalogación como transcritos independientes o procesados del mensajero es imposible mediante este abordaje y requeriría de aproximaciones experimentales complementarias. Estas regiones en la mayoría de los casos podrían corresponder a sitios de los mRNAs con especial afinidad por Hfq. El

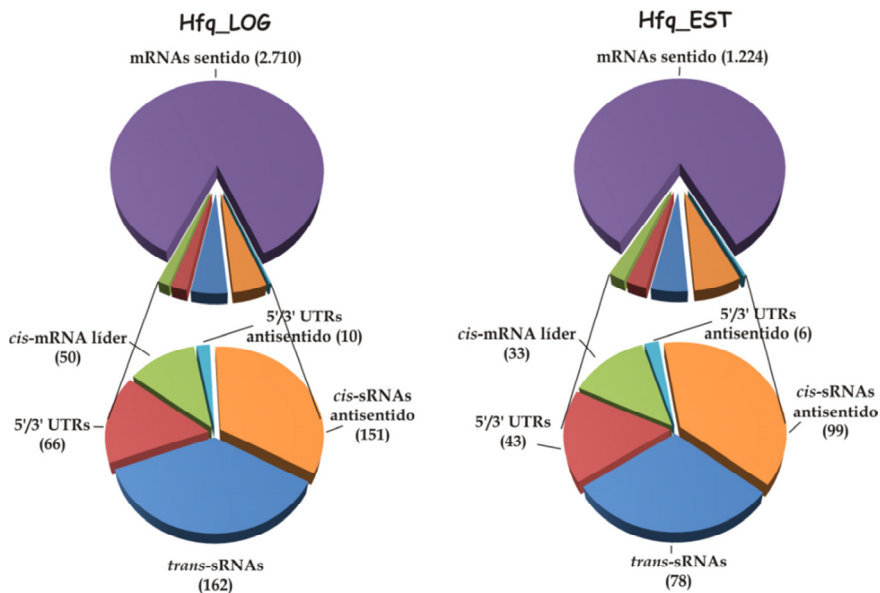
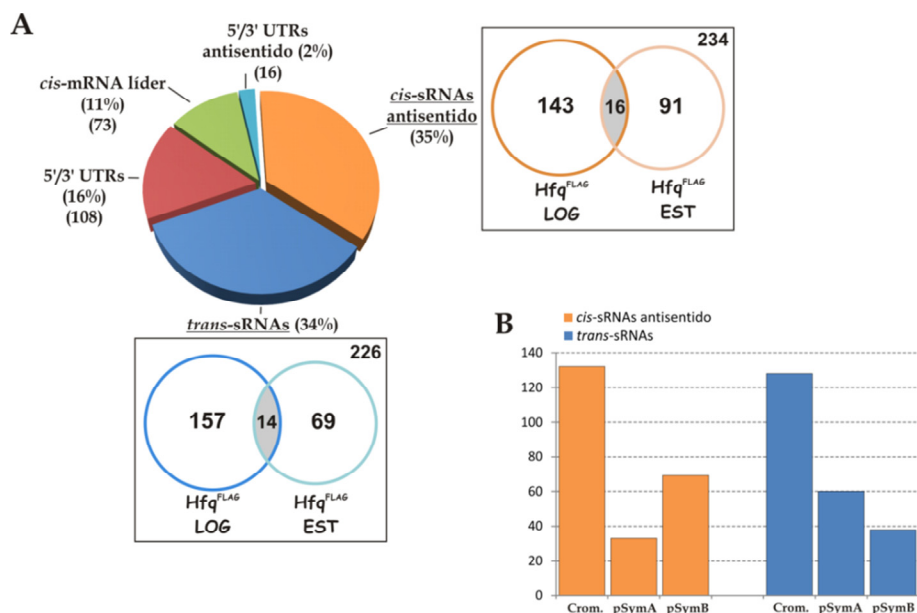


Fig. I-24. Distribución de las categorías de RNAs en las librerías correspondientes a fase logarítmica y estacionaria. Los diagramas de sectores superiores indican la distribución total de las 6 categorías catalogadas. Los inferiores indican la distribución de las categorías de RNAs excluyendo a los mRNAs sentido. El número de regiones identificadas correspondiente a cada categoría se indica entre paréntesis.

resto de regiones, sin embargo, sí son catalogables como RNAs no codificantes y su representación (en porcentaje relativo) en CoIP-RNA (ambos bancos conjuntamente) se muestra en la Fig. I-25A. No obstante, la adecuada caracterización de las regiones 5'/3'-UTR y su potencial regulador (e.g. como *riboswitches* u otros motivos reguladores) tampoco es posible mediante este análisis y, al igual que las mapeadas en la región codificante de los mRNAs, en algún caso podrían ser simples sitios de asociación del mRNA con Hfq. Por el contrario, todas las regiones asignadas al grupo de los *cis*-mRNA líder en nuestro estudio sí representan verdaderos sRNAs y corresponden a 73 de los 379 sRNAs catalogados como tales por Schlüter *et al.* (2010). Las regiones identificadas como *trans*-sRNAs y *cis*-sRNAs antisentido son las mayoritarias en el CoIP-RNA, representando en conjunto 460 *loci* (un 69% del total asociado a Hfq) que se relacionan en la Tabla I-7, presentada en el anexo de este capítulo, para su consulta.

El mapeo en el genoma de *S. meliloti* 1021 de estas regiones reveló la siguiente distribución por replicones (Fig. I-25B): 132 de los 234 *cis*-sRNAs antisentido se localizan en el cromosoma, 33 en pSymA y 69 en pSymB, mientras que de los 226 *trans*-sRNAs 128 lo hacen en el cromosoma, 60 en pSymA y 38 en pSymB. Por otra parte, ambos grupos de transcritos parecen ser bastante específicos de la condición

biológica en la que se preparó el CoIP-RNA pues sólo 14 *trans*- y 16 *cis*-sRNAs aparecieron en ambos bancos, Hfq\_Log y Hfq\_Est.

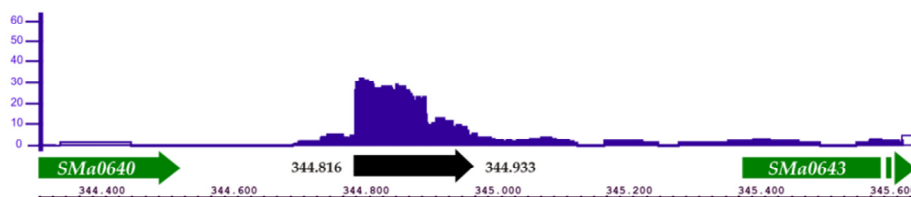


**Fig. I-25. Análisis conjunto de las librerías correspondientes a fase logarítmica y estacionaria. A)** Distribución global de las categorías de RNAs catalogadas (excluyendo los mRNAs sentido), indicando entre paréntesis el porcentaje que representan frente al total. Para las categorías de *trans*-sRNAs y los *cis*-sRNAs antisentido, se muestran diagramas Venn correspondientes a la distribución de los RNAs en cada librería. **B)** Diagrama que indica la distribución de los transcritos correspondientes a *trans*-sRNAs y los *cis*-sRNAs antisentido en los tres replicones del genoma de *S.meliloti*.

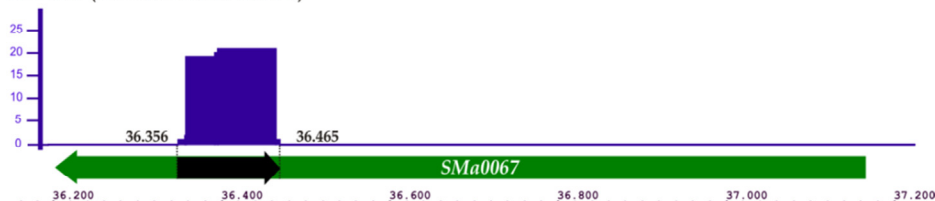
Finalmente, este análisis reveló también que 196 *trans*-sRNAs y 225 *cis*-sRNAs no habían sido anotados en *S. meliloti* 1021, por lo que representan RNAs no codificantes de nuevo descubrimiento. La Fig. I-26 muestra tres ejemplos de transcritos identificados que no corresponden a ninguna región anotada previamente según el mapeo de las lecturas visualizado con IGB; un *trans*-sRNA típico, REG70, situado entre dos genes transcritos en la misma dirección, y dos *cis*-sRNAs antisentido, REG137 y REG1358, el primero de los cuales solapa exclusivamente en la región codificante del transportador de monosacáridos *SMa0067*, y el segundo con la secuencia de dos genes contiguos que codifican proteínas hipotéticas *SMb21450* y *SMb21452*.



REG70 (*trans*-sRNA)



REG137 (*cis*-sRNA antisentido)



REG1358 (*cis*-sRNA antisentido)

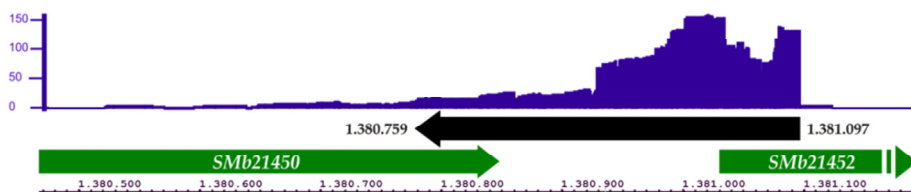


Fig. I-26. Ejemplos de RNAs no anotados en el genoma de *S. meliloti* 1021 identificados en nuestro estudio. Diagramas obtenidos de IGB en donde se representa el enriquecimiento de las librerías Hfq<sup>FLAG</sup> frente al control 2011 para cada transcrito. La anotación y localización de las regiones anotadas (en verde) y sin anotar (negro) se indican en la parte inferior de cada panel, además de las coordenadas de los RNAs identificados. Las flechas discontinuas indican una extensión mayor del gen anotado.

## I-4. DISCUSIÓN

Cada vez es más evidente el impacto de la actividad de los RNAs no codificantes en la fisiología bacteriana, lo que hace necesaria la anotación de sus genes en los genomas secuenciados. El primer análisis de un genoma bacteriano se limita esencialmente a la anotación de ORFs, rRNAs y tRNAs. Sin embargo, el catálogo de los genes que codifican sRNAs no es todavía lo suficientemente extenso como para inferir características diferenciales de los mismos con las que desarrollar modelos estadísticos implementables en algoritmos informáticos para la predicción de sus secuencias en análisis de genomas individuales (*machine learning approaches*). Por ello las primeras estrategias para la identificación de estos genes en bacterias se han basado en la combinación de genómica comparada y técnicas moleculares clásicas para la validación individual de cada predicción, en una labor tediosa y dilatada en el

tiempo. Sin embargo, el advenimiento y desarrollo recientes de las nuevas técnicas de secuenciación masiva (Next Generation Sequencing Technologies; NGS) han facilitado enormemente esta tarea, posibilitando la identificación directa, la primera caracterización molecular y la catalogación de los sRNAs expresados por las bacterias, en poco tiempo y a costes razonables. En este Capítulo hemos presentado un estudio pionero que integra ambos abordajes para la identificación de los componentes de una fracción representativa del RNoma del endosimbionte diazotrófico *S. meliloti*.

### ***Predicción e identificación de genes sRNA en el genoma de S. meliloti***

Nuestra estrategia inicial para la identificación de sRNAs en *S. meliloti* combina predicciones bioinformáticas fundamentadas en la conservación de la estructura termodinámicamente favorecida de los sRNAs con la confirmación experimental de la expresión de éstos mediante hibridación Northern y determinación RACE de sus inicios de transcripción. La fracción del genoma comparada fue la no anotada en la cepa de referencia *S. meliloti* 1021 (Galibert *et al.*, 2001), correspondiente a las RIGs de más de 50 nt cuyas secuencias fueron utilizadas para interrogar bajo parámetros estrictos los genomas de siete  $\alpha$ -proteobacterias filogenéticamente relacionadas. Este abordaje resultó en la identificación de un total de diez sRNAs; tres de ellos homólogos reconocibles de los RNAs de función conocida, RNasa P, 4.5S (SRP) y 6S, y los siete restantes sin homología aparente en las bases de datos (e.g. Rfam) que fueron considerados, por tanto, nuevos sRNAs bacterianos que denominamos Smr. Con posterioridad a este estudio estrategias predictivas basadas esencialmente en la conservación de secuencias nucleotídicas primarias flanqueadas por elementos de transcripción predecibles (i.e. promotores y terminadores de transcripción), combinadas con hibridación Northern y/o sobre *microarrays*, elevaron a algo más de una treintena los *loci* sRNA identificados en esta bacteria (Ulvé *et al.*, 2007, Valverde *et al.*, 2008). Llama la atención el reducido número de sRNAs identificados por nuestro *screening* en comparación con los resultantes de los estudios citados o con otros, que con características similares, han sido documentados previamente en diversos modelos bacterianos ((Carter *et al.*, 2001; Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Chen *et al.*, 2002; Livny *et al.*, 2005; Livny *et al.*, 2006). Sin embargo, este resultado era esperable y puede considerarse consecuencia lógica de la restricción impuesta a la búsqueda. En primer lugar, para este análisis sólo fueron tenidas en cuenta las secuencias de las RIGs derivadas de la primera anotación del genoma de *S. meliloti* 1021 y entre ellas, las mayores de 50 nt no adyacentes a los codones de inicio de traducción predichos para los ORFs flanqueantes. Este criterio, unido al carácter comparativo de la búsqueda, impide la identificación de sRNAs de tamaño inferior a

50 nt, los específicos de *S. meliloti* (i.e. no conservados) así como de cualquier categoría de *cis*-sRNAs (sentido y antisentido) asociados a los ORFs anotados u originados por procesamiento de o co-transcripción con éstos. De esta forma nuestro *screening* tiene un sesgo evidente hacia la identificación de posibles *trans*-sRNAs que, no obstante, es común a todas las búsquedas equivalentes en otros genomas bacterianos. En segundo lugar, nuestra metodología tiene en consideración el problema del alto índice de falsas predicciones positivas, muy común y generalizado en el campo de la biología computacional de los RNAs no codificantes. Por ello se combinaron dos herramientas bioinformáticas, eQRNA y RNAz, con criterios predictivos complementarios, realizando posteriormente una reducción de los teóricos falsos positivos (i.e. en términos puramente estadísticos) derivados de la aplicación de ambos algoritmos. Dado que la casi totalidad de los RNAs no codificantes de pequeño tamaño anotados en el genoma de *S. meliloti* correspondían a los tRNAs, la validación de la sensibilidad de nuestro *screening* se hizo necesariamente utilizando esta población de transcritos como referencia de RNAs de estructura conservada y termodinámicamente estable (i.e. *control positivo* de las predicciones). Aunque esta validación podría ser no generalizable para la predicción de sRNAs diferentes a los tRNAs, ésta resultó en un *screening* con altos valores de sensibilidad y especificidad en la predicción de nuevos sRNAs en el genoma de esta bacteria. Por tanto, a pesar de sus limitaciones, nuestros resultados experimentales apoyan la combinación de eQRNA y RNAz como estrategia fiable para la predicción de sRNAs en genomas bacterianos con un bajo índice de falsos positivos.

### ***Los sRNAs Smr exhiben características típicas de los trans riboreguladores***

Además de estar codificados en unidades transcripcionales independientes localizadas en RIGs, otras características diferenciales de los *trans*-sRNAs mejor caracterizados son su expresión diferencial, su localización cromosómica y su conservación en especies bacterianas filogenéticamente próximas. ..

La primera validación de la expresión de los sRNAs identificados en nuestro *screening* se realizó mediante hibridación Northern de RNA total de *S. meliloti* 1021 obtenido en diversas condiciones biológicas relevantes en los estadios de vida libre y endosimbiótica de esta bacteria. Como era esperado por su función esencial o de mantenimiento celular, los sRNAs RNAsa P, 4.5S y 6S mostraron altos niveles de expresión en todas las condiciones, con perfiles aparentemente constitutivos en todos los casos. Los siete transcritos Smr, sin embargo, evidenciaron una expresión diferencial. El mapeo de los extremos 5' de las especies de RNA detectadas en los geles reveló que se trataba de transcritos primarios codificados en unidades

transcripcionales autónomas con promotores  $\sigma^{70}$  reconocibles y terminadores de transcripción independientes de Rho típicos de los *trans*-sRNAs bacterianos conocidos (MacLellan *et al.*, 2006; Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Rivas *et al.*, 2001). El primer análisis de las secuencias promotoras de estos genes no reveló la presencia de secuencias de DNA potencialmente reconocibles por factores sigma alternativos o factores de transcripción conocidos. La excepción fue *smr35B*, localizado en el plásmido pSymB, en cuyo promotor fue identificada la secuencia consenso de reconocimiento de los reguladores transcripcionales de la familia LysR, a la que pertenece el activador de los genes de nodulación NodD (Goethals *et al.*, 1992; Salgado *et al.*, 2004). Si bien no se trata de una caja Nod típica este hallazgo podría justificar la discreta inducción de este gen en presencia del flavonoide luteolina.

En línea a lo descrito para otros *trans*-sRNAs bacterianos la acumulación dependiente de la etapa de crecimiento fue también evidente en los transcritos Smr. Sin embargo, mientras que la mayoría de sRNAs mejor caracterizados se acumulan cuando las bacterias entran en fase estacionaria tardía de crecimiento (Argaman *et al.*, 2001; Wassarman *et al.*, 2001), los sRNAs Smr lo hacen durante el crecimiento logarítmico, con la excepción de Smr15C1. *S. meliloti* carece de un homólogo reconocible del gen *rpoS*, que en *E. coli* y otras bacterias codifica el factor sigma de fase estacionaria o de respuesta a estrés,  $\sigma^{38}$  (Repoila *et al.*, 2003). Por tanto, nuestros resultados anticipan diferencias sustanciales en los mecanismos de regulación subyacentes a las respuestas al estrés, dependientes e independientes de RpoS, en las distintas especies bacterianas.

Los estudios transcriptómicos y proteómicos han revelado profundos cambios en la expresión génica en *S. meliloti* durante el proceso de diferenciación morfológica de los bacteroides (Djordjevic *et al.*, 2003; Barnett *et al.*, 2004; Becker *et al.*, 2004). Este primer estudio de la expresión de los sRNAs Smr reveló que dos de ellos, Smr7C y Smr14C, son inducidos en bacterias endosimbióticas por lo que su posible contribución a la regulación de la infección y/o diferenciación del bacteroide debe ser explorada en el futuro.

La identificación rigurosa de homólogos de los sRNAs está basada en la conservación estructural más que en la de la secuencia nucleotídica primaria, que puede ser más divergente entre transcritos. Las estructuras secundarias similares están sustentadas por diversos grados de co-varianza de los nucleótidos complementarios que conforman los tallos de las diferentes horquillas. Los modelos de co-varianza (MC) inferidos de los alineamientos de secuencias nucleotídicas homólogas a los sRNAs Smr fueron, por tanto, utilizados para interrogar todos los

genomas bacterianos en busca de nuevos miembros de cada familia. Este análisis reveló que los transcritos Smr identificados en *S. meliloti* sólo están conservados en especies del orden Rhizobiales en el grupo de las  $\alpha$ -proteobacterias, lo que definió seis nuevas familias de sRNAs (Smr15C1 y Smr15C2 satisfacen el mismo MC) que denominamos  $\alpha r$  ( $\alpha$ -proteobacteria RNA). Las familias  $\alpha r7$ ,  $\alpha r9$  y  $\alpha r45$  representan nuevos ejemplos de sRNAs bacterianos codificados en genes cromosómicos de copia única localizados en regiones de gran sintenia entre los respectivos genomas. Los genes de copia única, sin embargo, apenas representan el 58% del total de los anotados en *S. meliloti* 1021 (Galibert *et al.*, 2001). En general los genomas de bacterias que establecen interacciones de diversa naturaleza con plantas evidencian altos niveles de paralogía, lo que sugiere que su expansión mediante duplicaciones génicas no ha estado contenida durante la evolución, facilitando así la adquisición de nuevas funciones para la vida en el suelo y en el interior de las células vegetales (Batut *et al.*, 2004; Galibert *et al.*, 2001). Los miembros de las familias  $\alpha r14$  y  $\alpha r15$  están presentes en múltiples copias en los genomas individuales, situación que no es excepcional en bacterias (Waters and Storz, 2009), por lo que la significación biológica de estas reiteraciones es otro aspecto a tener en cuenta en la caracterización funcional de estos sRNAs. En general, la distribución de las familias  $\alpha r$  de sRNAs en especies del orden Rhizobiales sugiere que tanto la herencia vertical como las numerosas duplicaciones ancestrales han contribuido a la evolución de los sRNAs en bacterias que establecen residencias intracelulares crónicas en eucariotas superiores.

### **Identificación de sRNAs en *S. meliloti* mediante secuenciación masiva (RNA-Seq)**

La secuenciación masiva de librerías de cDNA (RNA-Seq) no sólo permite la identificación de sRNAs sino también la determinación de su secuencia nucleotídica exacta y de la abundancia relativa de cada transcrito en las muestras de RNA, soslayando así muchas de las limitaciones de la genómica comparada y de la hibridación de *microarrays* (Sharma y Vogel, 2009). En *S. meliloti* un estudio pionero muy reciente basado en la secuenciación en plataforma 454 e Illumina/Solexa de la fracción de transcritos de pequeño tamaño (i.e. 50-350 nt) aislada en ocho condiciones biológicas diferentes identificó 1.125 posibles sRNAs, que incluyen todos los identificados mediante genómica comparada y que fueron clasificados como *trans*-sRNAs (173) *cis*-sRNAs antisentido (117), transcritos de regiones líder de mRNAs (379) y *cis*-sRNAs sentido (456) (Schlüter *et al.*, 2010). Este catálogo de transcritos puede considerarse, por tanto, como RNoma de referencia en esta especie bacteriana.

Es bien sabido que casi la totalidad de *trans*-sRNAs bacterianos de los que se tiene información funcional se asocian en la célula con la chaperona de RNA Hfq (Aiba,

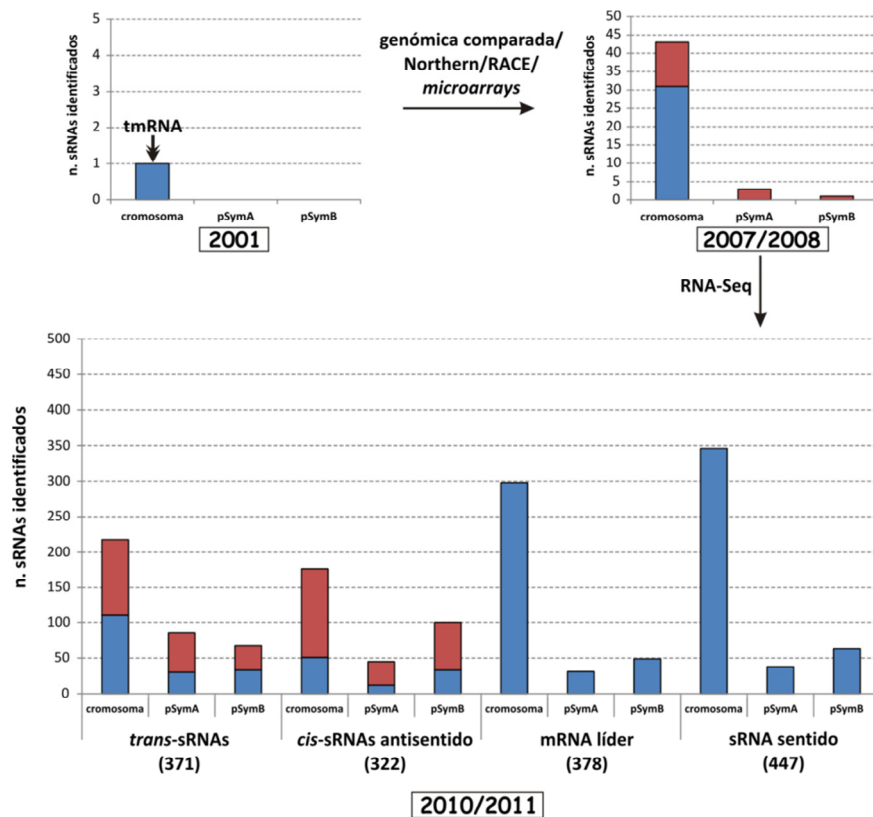
2007; Valentin-Hansen *et al.*, 2004). Estudios más recientes, sin embargo, han revelado que la capacidad de interacción de Hfq con RNAs no codificantes se extiende también a la clase de los *cis*-sRNAs antisentido (Lorenz *et al.*, 2010). Mientras que la catalogación de los sRNAs asociados a los mensajeros y transcritos en el mismo sentido de éstos (i.e. *cis*-sRNAs sentido, regiones líder de mRNAs y 5'/3'-UTRs) como sRNAs con potencial regulador es tremendamente imprecisa en ausencia de evidencias experimentales adicionales, los *trans*-sRNAs y los *cis*-RNAs antisentido representan inequívocamente verdaderos transcritos independientes de los mRNAs y, en consecuencia, una fracción representativa del RNoma bacteriano. La secuenciación masiva de la subpoblación de transcritos aislada por co-inmunoprecipitación (CoIP) con una proteína Hfq marcada en su región C-terminal con un péptido (FLAG) reconocido por anticuerpos comerciales ha sido utilizada con éxito para la identificación de gran parte de los sRNAs conocidos en *Salmonella* y *Rhodobacter sphaeroides* (Sittka *et al.*, 2008; Berghoff *et al.*, 2011). Esta es una estrategia genérica, potencialmente aplicable a cualquier modelo bacteriano genéticamente manipulable que nosotros hemos implementado también en *S. meliloti* con el mismo fin. Esta modificación en el gen cromosómico *hfq* no altera fenotipos esenciales de *S. meliloti*, lo que sugiere que esta variante de la proteína no está comprometida en su capacidad para unir RNA, tal como ha sido descrito en las bacterias citadas.

El análisis de las secuencias de cDNA de CoIP-RNA obtenido en las fases exponencial y estacionaria de crecimiento y la catalogación de las regiones asociadas a Hfq (i.e. enriquecidas en cada caso respecto al CoIP-RNA control) reveló la presencia mayoritaria de transcritos derivados de mRNAs. Entre las regiones asignables a cualquier categoría de sRNAs fueron mayoritarias las correspondientes a *trans*-sRNAs (226) y *cis*-sRNAs antisentido (234), representando en conjunto el 69% del total de sRNAs identificados mediante esta estrategia. Las regiones representadas en ambas librerías de CoIP-RNA fueron minoritarias respecto al total (i.e. 14 *trans*-sRNAs y 16 *cis*-sRNAs antisentido) lo que sugiere que la mayoría de estos transcritos se expresa diferencialmente como es esperable de los ribo-reguladores. Por otra parte, sólo el 15,6% de los *trans*-sRNAs y el 8% de los *cis*-sRNAs antisentido identificados por Schlüter *et al.* (2010) aparecieron en nuestras librerías de CoIP-RNA, mientras que estos porcentajes aumentaron al 36% y 21%, respectivamente, si consideramos también las librerías de CoIP-RNA obtenido de bacterias sometidas a shock térmico y salino (A. Becker y J.-P. Schlüter, datos no publicados). El resto de transcritos representados en las librerías de CoIP-RNA derivadas de fase logarítmica y estacionaria de crecimiento (i.e. 163 *trans*-sRNAs y 206 *cis*-sRNAs antisentido) pueden considerarse de nuevo descubrimiento. La Fig. I-27 muestra la evolución de la caracterización del RNoma de *S. meliloti* y nuestra contribución a la misma desde la

publicación de la secuencia genómica de la cepa de referencia 1021 hasta la finalización de este trabajo de Tesis Doctoral.

La escasa coincidencia entre los catálogos de sRNAs generados por el estudio de Schlüter *et al.* 2010 y nuestro trabajo de investigación puede tener varias causas: *i*) los ribo-reguladores (*trans*-sRNAs y *cis*-sRNAs antisentido) suelen acumularse en la célula en condiciones biológicas específicas, por lo que es esperable que el análisis de un mayor número de condiciones resulte en la identificación de más sRNAs, entre los que probablemente estaría también alguno de los identificados por Schlüter *et al.* (2010) que no aparecen representados en las librerías de CoIP-RNA analizadas en este trabajo; *ii*) la subpoblación de transcritos asociados a Hfq muy probablemente represente una fracción específica del RNoma de *S. meliloti*, de forma que nuestros resultados anticipan que otros sRNAs expresados por la bacteria no se asocian a Hfq o lo hacen a otra u otras proteínas no identificadas; *iii*) la CoIP con Hfq resulta en un enriquecimiento de las preparaciones de RNA en transcritos que muy probablemente no pudieran ser detectados en el conjunto del RNA total de la bacteria.

En definitiva, nuestros resultados confirman la eficacia de la RNA-Seq para la identificación de sRNAs en bacterias aunque la saturación de la búsqueda de estos transcritos requerirá, al menos en *S. meliloti*, de la secuenciación de nuevas poblaciones de RNAs obtenidas en un mayor número de condiciones biológicas y/o por CoIP con otras proteínas de relevancia en la actividad de los ribo-reguladores.



**Fig. I-27. Evolución de la identificación de sRNAs en el genoma de *S. meliloti*.** Histogramas que representan el número de sRNAs anotados en *S. meliloti* en 2011 (primer panel), 2007/2008 (segundo panel) y 2010/2011 (panel inferior). En rojo se muestra la contribución de este trabajo a la identificación de sRNAs. En azul se indica la contribución de otros trabajos: 2007/2008 (Ulve *et al.*, 2007; Valverde *et al.*, 2008); 2011/2012 (Schlüter *et al.*, 2010). En el primer panel además se indica el tmRNA como único representante anotado.





# **ANEXO CAPÍTULO 1**



Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio

Clase <sup>1</sup>	Inicio	Final	Hebra	ID	Condición	Replicón	Longitud (nt)	Enriquecimiento	Región GenDB <sup>2</sup>
<i>cis</i> -sRNA antisentido	160956	160988	-	REG1613	LOG	Cromosoma	33	10.38	SMc04113
<i>cis</i> -sRNA antisentido	161323	161356	-	REG1614	LOG	Cromosoma	34	6.37	SMc04113
<i>cis</i> -sRNA antisentido	161435	161457	-	REG1615	LOG	Cromosoma	23	8.45	SMc04113
<i>trans</i> -sRNA	203994	204033	+	REG573	LOG	Cromosoma	40	8.85	
<i>trans</i> -sRNA	228742	228766	-	REG1632	LOG	Cromosoma	25	6.80	
<i>cis</i> -sRNA antisentido	257230	257293	+	REG593	LOG	Cromosoma	64	10.46	SMc02901
<i>cis</i> -sRNA antisentido	262035	262088	+	REG594	LOG	Cromosoma	54	9.37	SMc04634
<i>cis</i> -sRNA antisentido	325569	326193	-	REG591	EST	Cromosoma	625	49.06	SmelC052
<i>trans</i> -sRNA	398479	398621	+	REG643	LOG	Cromosoma	143	10.12	
<i>trans</i> -sRNA	398573	398655	-	REG1668	LOG	Cromosoma	83	7.45	
<i>trans</i> -sRNA	399027	399049	+	REG644	LOG	Cromosoma	23	7.85	
<i>cis</i> -sRNA antisentido	409487	409519	+	REG651	LOG	Cromosoma	33	7.80	AL591688_ ; 476
<i>trans</i> -sRNA	416859	416881	+	REG656	LOG	Cromosoma	23	6.50	
<i>trans</i> -sRNA	424107	424270	+	REG260	EST	Cromosoma	164	33.27	SmelC073
<i>trans</i> -sRNA	445945	446041	+	REG662	LOG	Cromosoma	97	7.68	
<i>trans</i> -sRNA	485217	485236	+	REG261	EST	Cromosoma	20	6.56	
<i>trans</i> -sRNA	491157	491250	+	REG265	EST	Cromosoma	94	7.90	
<i>trans</i> -sRNA	523554	523610	-	REG605	EST	Cromosoma	57	9.73	SmelC088
<i>trans</i> -sRNA	523648	523697	-	REG606	EST	Cromosoma	50	8.26	SmelC088
<i>trans</i> -sRNA	542996	543036	+	REG675	LOG	Cromosoma	41	9.53	
<i>cis</i> -sRNA antisentido	549009	549054	+	REG683	LOG	Cromosoma	46	6.82	SMc02155
<i>cis</i> -sRNA antisentido	564065	564084	-	REG1734	LOG	Cromosoma	20	6.08	SMc02149
<i>cis</i> -sRNA antisentido	586585	586619	-	REG1740	LOG	Cromosoma	35	8.22	AL591688_ ; 654
<i>cis</i> -sRNA antisentido	586644	586699	-	REG1741	LOG	Cromosoma	56	8.15	AL591688_ ; 655
<i>trans</i> -sRNA	734125	734145	+	REG722	LOG	Cromosoma	21	7.64	
<i>trans</i> -sRNA	757553	757651	+	REG734	LOG	Cromosoma	99	9.17	
<i>cis</i> -sRNA antisentido	779296	779316	-	REG615	EST	Cromosoma	21	5.19	SMc00766
<i>cis</i> -sRNA antisentido	779353	779383	-	REG616	EST	Cromosoma	31	5.64	SMc00766

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	779402	779424	-	REG617	EST	Cromosoma	23	6.39	SMc00766
<i>trans</i> -sRNA	798023	798044	+	REG755	LOG	Cromosoma	22	7.64	
<i>cis</i> -sRNA antisentido	805113	805147	-	REG621	EST	Cromosoma	35	6.75	SMc00783
<i>cis</i> -sRNA antisentido	815849	816232	-	REG628	EST	Cromosoma	384	28.05	SmelC137
<i>trans</i> -sRNA	816245	816705	-	REG629	EST	Cromosoma	461	17.44	SmelC139
<i>cis</i> -sRNA antisentido	819161	819193	+	REG289	EST	Cromosoma	33	5.41	SMc00799
<i>cis</i> -sRNA antisentido	819280	819327	+	REG290	EST	Cromosoma	48	9.89	SMc00799
<i>trans</i> -sRNA	843429	843451	+	REG772	LOG	Cromosoma	23	17.50	SmelC151
<i>trans</i> -sRNA	843475	843535	+	REG773	LOG	Cromosoma	61	7.24	SmelC151
<i>trans</i> -sRNA	923650	923747	-	REG638	EST	Cromosoma	98	10.38	SmelC181
<i>cis</i> -sRNA antisentido	947900	947934	-	REG1813	LOG	Cromosoma	35	7.36	SMc00988
<i>trans</i> -sRNA	954879	954925	+	REG794	LOG	Cromosoma	47	8.03	
<i>trans</i> -sRNA	954929	954950	+	REG795	LOG	Cromosoma	22	5.68	
<i>cis</i> -sRNA antisentido	1010600	1010863	-	REG1827	LOG	Cromosoma	264	20.10	SMc00020
<i>cis</i> -sRNA antisentido	1010971	1011109	-	REG1828	LOG	Cromosoma	139	10.09	SMc00020
<i>cis</i> -sRNA antisentido	1011284	1011327	-	REG1829	LOG	Cromosoma	44	9.21	SmelC203
<i>trans</i> -sRNA	1029150	1029177	+	REG310	EST	Cromosoma	28	26.62	
<i>cis</i> -sRNA antisentido	1032875	1032919	-	REG1834	LOG	Cromosoma	45	7.53	SMc00028
<i>cis</i> -sRNA antisentido	1058120	1058139	+	REG313	EST	Cromosoma	20	7.23	SMc00099
<i>cis</i> -sRNA antisentido	1058149	1058195	+	REG819	LOG	Cromosoma	47	6.08	SmelC218
<i>cis</i> -sRNA antisentido	1060747	1060777	+	REG820	LOG	Cromosoma	31	9.20	SMc00101
<i>trans</i> -sRNA	1069859	1069890	+	REG824	LOG	Cromosoma	32	9.86	
<i>trans</i> -sRNA	1099165	1099200	+	REG321	EST	Cromosoma	36	11.14	
<i>trans</i> -sRNA	1127193	1127232	-	REG1866	LOG	Cromosoma	40	17.24	
<i>cis</i> -sRNA antisentido	1134618	1134660	+	REG836	LOG	Cromosoma	43	8.08	SMc02380
<i>cis</i> -sRNA antisentido	1134662	1134916	+	REG837	LOG	Cromosoma	255	9.46	SMc02380
<i>cis</i> -sRNA antisentido	1144305	1144330	+	REG842	LOG	Cromosoma	26	6.24	SMc02391
<i>cis</i> -sRNA antisentido	1144332	1144359	+	REG843	LOG	Cromosoma	28	6.35	SMc02391
<i>trans</i> -sRNA	1145007	1145036	+	REG844	LOG	Cromosoma	30	6.83	
<i>cis</i> -sRNA antisentido	1145525	1145695	+	REG847	LOG	Cromosoma	171	13.12	SMc02393
<i>cis</i> -sRNA antisentido	1145753	1145790	+	REG848	LOG	Cromosoma	38	8.76	SMc02393
<i>cis</i> -sRNA antisentido	1145796	1145996	+	REG849	LOG	Cromosoma	201	22.09	SMc02393

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	1146066	1146135	+	REG850	LOG	Cromosoma	70	8.26	SMc02393
<i>cis</i> -sRNA antisentido	1146273	1146300	-	REG851	LOG	Cromosoma	28	8.02	SMc02393
<i>cis</i> -sRNA antisentido	1146982	1147022	-	REG1873	LOG	Cromosoma	41	6.08	SMc02409
<i>cis</i> -sRNA antisentido	1149756	1149787	+	REG857	LOG	Cromosoma	32	8.67	SMc02397
<i>cis</i> -sRNA antisentido	1149790	1149818	+	REG858	LOG	Cromosoma	29	7.51	SMc02397
<i>cis</i> -sRNA antisentido	1149849	1149877	+	REG859	LOG	Cromosoma	29	6.01	SMc02397
<i>trans</i> -sRNA	1150789	1150847	+	REG326	EST	Cromosoma	59	8.26	
<i>trans</i> -sRNA	1157601	1157622	+	REG864	LOG	Cromosoma	22	6.14	
<i>cis</i> -sRNA antisentido	1157849	1157905	+	REG865	LOG	Cromosoma	57	7.81	SMc04541
<i>trans</i> -sRNA	1167539	1167581	+	REG867	LOG	Cromosoma	43	9.05	
<i>trans</i> -sRNA	1176554	1176579	+	REG329	EST	Cromosoma	26	8.51	
<i>cis</i> -sRNA antisentido	1208078	1208112	-	REG1879	LOG	Cromosoma	35	6.67	SMc02553
<i>cis</i> -sRNA antisentido	1208174	1208196	-	REG1880	LOG	Cromosoma	23	8.03	SMc02553
<i>cis</i> -sRNA antisentido	1208382	1208410	-	REG1881	LOG	Cromosoma	29	6.02	SMc02553
<i>cis</i> -sRNA antisentido	1208432	1208458	-	REG1882	LOG	Cromosoma	27	9.88	SMc02553
<i>cis</i> -sRNA antisentido	1235641	1235673	-	REG1888	LOG	Cromosoma	33	6.80	SMc00564
<i>cis</i> -sRNA antisentido	1235769	1235870	-	REG1889	LOG	Cromosoma	102	11.70	SMc00564
<i>trans</i> -sRNA	1243125	1243179	+	REG888	LOG	Cromosoma	55	11.30	
<i>trans</i> -sRNA	1263981	1264002	+	REG895	LOG	Cromosoma	22	6.63	
<i>cis</i> -sRNA antisentido	1277411	1277439	+	REG336	EST	Cromosoma	29	6.49	SMc00606
<i>cis</i> -sRNA antisentido	1332364	1332430	-	REG661	EST	Cromosoma	67	7.79	SMc01787
<i>cis</i> -sRNA antisentido	1332433	1332637	-	REG662	EST	Cromosoma	205	107.18	SmelC271
<i>trans</i> -sRNA	1333054	1333081	+	REG910	LOG	Cromosoma	28	5.83	
<i>trans</i> -sRNA	1333333	1333353	-	REG1915	LOG	Cromosoma	21	6.39	
<i>trans</i> -sRNA	1333650	1333694	-	REG1916	LOG	Cromosoma	45	7.88	
<i>cis</i> -sRNA antisentido	1352926	1352974	+	REG911	LOG	Cromosoma	49	6.23	SMc01807
<i>trans</i> -sRNA	1398185	1398211	-	REG1920	LOG	Cromosoma	27	6.79	
<i>trans</i> -sRNA	1398225	1398265	-	REG1921	LOG	Cromosoma	41	6.51	
<i>trans</i> -sRNA	1398255	1398318	-	REG665	EST	Cromosoma	64	23.17	SmelC289
<i>trans</i> -sRNA	1398321	1398368	-	REG1922	LOG	Cromosoma	48	6.56	SmelC289
<i>trans</i> -sRNA	1411536	1411561	-	REG1924	LOG	Cromosoma	26	7.29	
<i>cis</i> -sRNA antisentido	1426216	1426271	+	REG936	LOG	Cromosoma	56	8.99	SMc01358

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	1426304	1426410	+	REG345	EST	Cromosoma	107	6.78	SMc01358
<i>cis</i> -sRNA antisentido	1426580	1426601	+	REG937	LOG	Cromosoma	22	5.83	SMc01358
<i>cis</i> -sRNA antisentido	1426603	1426642	+	REG938	LOG	Cromosoma	40	6.27	SMc01358
<i>trans</i> -sRNA	1450835	1450858	+	REG946	LOG	Cromosoma	24	8.05	
<i>cis</i> -sRNA antisentido	1452540	1452580	-	REG1954	LOG	Cromosoma	41	7.20	SMc01377
<i>trans</i> -sRNA	1452606	1452633	-	REG1955	LOG	Cromosoma	28	5.86	
<i>trans</i> -sRNA	1491225	1491247	+	REG1018	LOG	Cromosoma	23	5.30	
<i>cis</i> -sRNA antisentido	1505764	1505834	+	REG347	EST	Cromosoma	71	9.55	SMc01268
<i>trans</i> -sRNA	1507052	1507083	+	REG1030	LOG	Cromosoma	32	8.74	
<i>trans</i> -sRNA	1507110	1507150	+	REG1031	LOG	Cromosoma	41	16.20	
<i>cis</i> -sRNA antisentido	1541044	1541083	-	REG1970	LOG	Cromosoma	40	6.08	SMc01015
<i>trans</i> -sRNA	1555872	1555920	+	REG1046	LOG	Cromosoma	49	6.23	
<i>cis</i> -sRNA antisentido	1577111	1577150	-	REG1972	LOG	Cromosoma	40	8.95	SMc01048
<i>cis</i> -sRNA antisentido	1577157	1577185	-	REG1973	LOG	Cromosoma	29	6.25	SMc01048
<i>trans</i> -sRNA	1587275	1587313	+	REG1077	LOG	Cromosoma	39	8.25	
<i>trans</i> -sRNA	1608616	1608671	+	REG1100	LOG	Cromosoma	56	10.96	SmelC369
<i>trans</i> -sRNA	1635207	1635295	-	REG1995	LOG	Cromosoma	89	7.60	
<i>cis</i> -sRNA antisentido	1638249	1638313	+	REG1130	LOG	Cromosoma	65	13.06	SMc02077
<i>trans</i> -sRNA	1668500	1668540	+	REG1134	LOG	Cromosoma	41	9.06	
<i>trans</i> -sRNA	1674828	1674857	-	REG678	EST	Cromosoma	30	6.17	
<i>trans</i> -sRNA	1674862	1674882	-	REG679	EST	Cromosoma	21	8.18	
<i>cis</i> -sRNA antisentido	1689583	1689627	+	REG1143	LOG	Cromosoma	45	6.50	SmelC407
<i>trans</i> -sRNA	1698604	1698731	-	REG681	EST	Cromosoma	128	44.33	SmelC411
<i>trans</i> -sRNA	1698799	1699136	-	REG2014	LOG	Cromosoma	338	21.29	SmelC412
<i>trans</i> -sRNA	1718731	1719025	-	REG684	EST	Cromosoma	295	26.77	SmelC416
<i>trans</i> -sRNA	1727991	1728013	-	REG2031	LOG	Cromosoma	23	6.31	
<i>cis</i> -sRNA antisentido	1730188	1730226	+	REG1160	LOG	Cromosoma	39	6.82	SMc02814
<i>trans</i> -sRNA	1750253	1750272	+	REG1162	LOG	Cromosoma	20	17.33	
<i>cis</i> -sRNA antisentido	1757576	1757601	+	REG1169	LOG	Cromosoma	26	8.16	SMc04529
<i>cis</i> -sRNA antisentido	1757609	1757630	+	REG1170	LOG	Cromosoma	22	6.38	SMc04529
<i>cis</i> -sRNA antisentido	1757634	1757659	+	REG1171	LOG	Cromosoma	26	10.26	SMc04529
<i>cis</i> -sRNA antisentido	1759748	1759776	-	REG2039	LOG	Cromosoma	29	7.42	SmelC427

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>trans</i> -sRNA	1820131	1820157	+	REG1185	LOG	Cromosoma	27	5.83	
<i>trans</i> -sRNA	1820367	1820389	+	REG1186	LOG	Cromosoma	23	6.36	
<i>trans</i> -sRNA	1820394	1820536	+	REG1187	LOG	Cromosoma	143	13.99	
<i>trans</i> -sRNA	1822907	1823094	+	REG391	EST	Cromosoma	188	48.00	
<i>trans</i> -sRNA	1823148	1823184	+	REG1191	LOG	Cromosoma	37	5.46	SmelC435
<i>trans</i> -sRNA	1823211	1823237	+	REG1192	LOG	Cromosoma	27	6.32	SmelC435
<i>trans</i> -sRNA	1825887	1826004	+	REG394	EST	Cromosoma	118	48.18	
<i>trans</i> -sRNA	1851395	1851414	+	REG1194	LOG	Cromosoma	20	6.08	
<i>cis</i> -sRNA antisentido	1851962	1851982	+	REG395	EST	Cromosoma	21	9.72	SMc00282
<i>trans</i> -sRNA	1862637	1862666	+	REG396	EST	Cromosoma	30	6.67	
<i>trans</i> -sRNA	1862645	1862666	+	REG1196	LOG	Cromosoma	22	8.23	
<i>trans</i> -sRNA	1923656	1923709	+	REG1214	LOG	Cromosoma	54	9.47	
<i>trans</i> -sRNA	1923683	1923708	+	REG400	EST	Cromosoma	26	7.82	
<i>trans</i> -sRNA	1923816	1923851	+	REG1215	LOG	Cromosoma	36	10.92	
<i>trans</i> -sRNA	1923869	1923935	+	REG1216	LOG	Cromosoma	67	9.80	
<i>trans</i> -sRNA	1923877	1923933	+	REG401	EST	Cromosoma	57	6.89	
<i>trans</i> -sRNA	1923956	1924005	+	REG1217	LOG	Cromosoma	50	12.79	
<i>trans</i> -sRNA	1925999	1926020	+	REG1220	LOG	Cromosoma	22	6.62	
<i>cis</i> -sRNA antisentido	1944938	1945001	-	REG2094	LOG	Cromosoma	64	10.67	SMc00476
<i>cis</i> -sRNA antisentido	1960859	1960925	-	REG2103	LOG	Cromosoma	67	12.28	SMc00548
<i>trans</i> -sRNA	1963559	1963604	+	REG1226	LOG	Cromosoma	46	12.42	
<i>cis</i> -sRNA antisentido	1963617	1963703	+	REG1227	LOG	Cromosoma	87	14.94	SMc00193
<i>cis</i> -sRNA antisentido	1963705	1963890	+	REG1228	LOG	Cromosoma	186	15.20	SmelC457
<i>cis</i> -sRNA antisentido	1963812	1963978	+	REG404	EST	Cromosoma	167	12.72	SMc00193
<i>trans</i> -sRNA	1965407	1965458	+	REG1229	LOG	Cromosoma	52	8.37	
<i>cis</i> -sRNA antisentido	1985004	1985023	+	REG405	EST	Cromosoma	20	6.81	SMc00178
<i>cis</i> -sRNA antisentido	2007635	2007741	+	REG1232	LOG	Cromosoma	107	7.66	SMc04438
<i>trans</i> -sRNA	2019911	2019962	+	REG1234	LOG	Cromosoma	52	14.19	
<i>trans</i> -sRNA	2019980	2020089	+	REG1235	LOG	Cromosoma	110	9.98	
<i>cis</i> -sRNA antisentido	2021746	2021769	+	REG1245	LOG	Cromosoma	24	7.84	SMc00148
<i>cis</i> -sRNA antisentido	2031506	2031854	+	REG1250	LOG	Cromosoma	349	28.89	AL591688_2264
<i>cis</i> -sRNA antisentido	2033749	2033846	+	REG408	EST	Cromosoma	98	7.78	SMc00136



**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	2033896	2033937	+	REG409	EST	Cromosoma	42	5.95	SMc00136
<i>cis</i> -sRNA antisentido	2045474	2045563	+	REG1259	LOG	Cromosoma	90	13.12	SMc04218
<i>trans</i> -sRNA	2061199	2061234	+	REG1260	LOG	Cromosoma	36	10.31	
<i>trans</i> -sRNA	2065075	2065170	+	REG410	EST	Cromosoma	96	10.19	
<i>cis</i> -sRNA antisentido	2067135	2067158	-	REG720	EST	Cromosoma	24	7.58	SMc04241
<i>cis</i> -sRNA antisentido	2070428	2070447	-	REG2130	LOG	Cromosoma	20	6.08	SMc04245
<i>cis</i> -sRNA antisentido	2077139	2077161	+	REG412	EST	Cromosoma	23	6.89	SMc05007
<i>cis</i> -sRNA antisentido	2130129	2130177	-	REG744	EST	Cromosoma	49	11.73	SMc04311
<i>cis</i> -sRNA antisentido	2151258	2151280	-	REG745	EST	Cromosoma	23	5.55	SMc04552
<i>trans</i> -sRNA	2153343	2153392	+	REG414	EST	Cromosoma	50	13.30	
<i>trans</i> -sRNA	2153529	2153576	+	REG415	EST	Cromosoma	48	12.95	
<i>cis</i> -sRNA antisentido	2174978	2175017	-	REG746	EST	Cromosoma	40	10.43	SMc04185
<i>cis</i> -sRNA antisentido	2177536	2177567	+	REG1287	LOG	Cromosoma	32	6.37	SmelC498
<i>cis</i> -sRNA antisentido	2177656	2177687	+	REG419	EST	Cromosoma	32	5.85	SMc04189
<i>cis</i> -sRNA antisentido	2178646	2178697	+	REG1290	LOG	Cromosoma	52	9.35	SMc04191
<i>trans</i> -sRNA	2206415	2206603	+	REG421	EST	Cromosoma	189	54.23	SmelC507
<i>cis</i> -sRNA antisentido	2215393	2215669	+	REG422	EST	Cromosoma	277	48.24	SMc04299
<i>cis</i> -sRNA antisentido	2215733	2215768	+	REG423	EST	Cromosoma	36	6.01	SMc04299
<i>trans</i> -sRNA	2222902	2222930	+	REG1294	LOG	Cromosoma	29	8.57	
<i>trans</i> -sRNA	2229162	2229191	-	REG2167	LOG	Cromosoma	30	6.93	
<i>cis</i> -sRNA antisentido	2247126	2247148	-	REG2181	LOG	Cromosoma	23	6.08	SMc01405
<i>cis</i> -sRNA antisentido	2247170	2247208	-	REG2182	LOG	Cromosoma	39	7.82	SMc01405
<i>cis</i> -sRNA antisentido	2247217	2247288	-	REG2183	LOG	Cromosoma	72	11.47	SMc01405
<i>cis</i> -sRNA antisentido	2247352	2247393	-	REG2184	LOG	Cromosoma	42	7.70	SMc01405
<i>cis</i> -sRNA antisentido	2247364	2247392	-	REG762	EST	Cromosoma	29	7.41	SMc01405
<i>cis</i> -sRNA antisentido	2247427	2247573	-	REG2185	LOG	Cromosoma	147	7.50	SMc01405
<i>cis</i> -sRNA antisentido	2248117	2248220	+	REG1305	LOG	Cromosoma	104	6.53	SMc01488
<i>cis</i> -sRNA antisentido	2265687	2265712	-	REG2191	LOG	Cromosoma	26	6.22	SMc01415
<i>cis</i> -sRNA antisentido	2312423	2312529	-	REG768	EST	Cromosoma	107	9.64	SMc01465
<i>cis</i> -sRNA antisentido	2312429	2312522	-	REG2203	LOG	Cromosoma	94	8.29	SMc01465
<i>cis</i> -sRNA antisentido	2312567	2312602	-	REG769	EST	Cromosoma	36	6.43	SMc01465
<i>cis</i> -sRNA antisentido	2312667	2312707	-	REG770	EST	Cromosoma	41	11.40	SMc01465

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	2312684	2312706	-	REG2204	LOG	Cromosoma	23	9.87	SMc01465
<i>trans</i> -sRNA	2313248	2313337	-	REG2208	LOG	Cromosoma	90	20.18	
<i>trans</i> -sRNA	2320712	2320745	-	REG2209	LOG	Cromosoma	34	6.08	
<i>cis</i> -sRNA antisentido	2323358	2323377	-	REG2210	LOG	Cromosoma	20	8.32	SMc01885
<i>cis</i> -sRNA antisentido	2359499	2359552	+	REG441	EST	Cromosoma	54	6.73	SMc01854
<i>trans</i> -sRNA	2371412	2371435	+	REG1316	LOG	Cromosoma	24	8.18	
<i>trans</i> -sRNA	2393984	2394011	+	REG1319	LOG	Cromosoma	28	6.79	
<i>cis</i> -sRNA antisentido	2427106	2427258	-	REG2250	LOG	Cromosoma	153	9.47	SMc01634
<i>cis</i> -sRNA antisentido	2427337	2427380	-	REG2251	LOG	Cromosoma	44	7.80	SMc01634
<i>trans</i> -sRNA	2482026	2482075	-	REG2263	LOG	Cromosoma	50	12.84	
<i>cis</i> -sRNA antisentido	2485531	2485565	-	REG786	EST	Cromosoma	35	7.17	SMc01742
<i>trans</i> -sRNA	2523779	2523802	+	REG457	EST	Cromosoma	24	5.52	
<i>cis</i> -sRNA antisentido	2544514	2544584	-	REG791	EST	Cromosoma	71	7.27	SMc02695
<i>cis</i> -sRNA antisentido	2547976	2548085	+	REG459	EST	Cromosoma	110	15.11	SMc02700
<i>trans</i> -sRNA	2623412	2623433	-	REG795	EST	Cromosoma	22	6.94	
<i>cis</i> -sRNA antisentido	2642930	2643536	-	REG2308	LOG	Cromosoma	607	28.21	SmelC609
<i>trans</i> -sRNA	2651355	2651389	+	REG1384	LOG	Cromosoma	35	8.00	
<i>trans</i> -sRNA	2695487	2695536	-	REG2338	LOG	Cromosoma	50	23.22	SmelC616
<i>trans</i> -sRNA	2711021	2711066	+	REG492	EST	Cromosoma	46	21.72	
<i>cis</i> -sRNA antisentido	2740175	2740590	-	REG807	EST	Cromosoma	416	41.42	SMc04642
<i>cis</i> -sRNA antisentido	2740266	2740409	-	REG2339	LOG	Cromosoma	144	10.46	SMc04642
<i>cis</i> -sRNA antisentido	2804145	2804185	+	REG1395	LOG	Cromosoma	41	6.53	SMc04422
<i>trans</i> -sRNA	2816973	2816994	-	REG2348	LOG	Cromosoma	22	7.60	
<i>trans</i> -sRNA	2817308	2817923	-	REG2350	LOG	Cromosoma	616	56.63	SmelC625
<i>cis</i> -sRNA antisentido	2825449	2825513	-	REG2355	LOG	Cromosoma	65	6.73	SMc02440
<i>trans</i> -sRNA	2888352	2888417	+	REG1409	LOG	Cromosoma	66	8.33	
<i>trans</i> -sRNA	2942014	2942045	-	REG2384	LOG	Cromosoma	32	12.38	
<i>trans</i> -sRNA	2969082	2969124	-	REG2395	LOG	Cromosoma	43	7.24	
<i>trans</i> -sRNA	2986019	2986470	+	REG1437	LOG	Cromosoma	452	16.49	SmelC671
<i>trans</i> -sRNA	2999051	2999084	+	REG1449	LOG	Cromosoma	34	7.58	
<i>trans</i> -sRNA	3046625	3046658	-	REG2404	LOG	Cromosoma	34	7.52	
<i>trans</i> -sRNA	3046709	3046791	+	REG525	EST	Cromosoma	83	45.44	SmelC689

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	3051893	3051977	-	REG840	EST	Cromosoma	85	8.20	SMc04873
<i>cis</i> -sRNA antisentido	3052134	3052156	-	REG841	EST	Cromosoma	23	6.88	SMc04873
<i>trans</i> -sRNA	3109192	3109216	-	REG2422	LOG	Cromosoma	25	8.77	
<i>trans</i> -sRNA	3109250	3109279	-	REG2423	LOG	Cromosoma	30	5.59	
<i>cis</i> -sRNA antisentido	3169751	3169778	-	REG2440	LOG	Cromosoma	28	10.00	SMc03204
<i>cis</i> -sRNA antisentido	3196448	3196732	+	REG1487	LOG	Cromosoma	285	17.20	SMc03133
<i>cis</i> -sRNA antisentido	3196508	3196534	+	REG540	EST	Cromosoma	27	5.41	SMc03133
<i>cis</i> -sRNA antisentido	3196655	3196711	+	REG541	EST	Cromosoma	57	8.18	SMc03133
<i>trans</i> -sRNA	3233843	3233900	-	REG2448	LOG	Cromosoma	58	6.00	
<i>cis</i> -sRNA antisentido	3256424	3256477	+	REG1498	LOG	Cromosoma	54	6.71	SMc03092
<i>cis</i> -sRNA antisentido	3312983	3313019	-	REG2473	LOG	Cromosoma	37	7.23	AL591688_3694
<i>trans</i> -sRNA	3347615	3347643	+	REG1518	LOG	Cromosoma	29	10.86	
<i>trans</i> -sRNA	3364558	3364655	-	REG2491	LOG	Cromosoma	98	9.88	
<i>trans</i> -sRNA	3367117	3367144	-	REG2498	LOG	Cromosoma	28	5.71	
<i>trans</i> -sRNA	3367146	3367289	-	REG2499	LOG	Cromosoma	144	5.95	
<i>trans</i> -sRNA	3368593	3368630	-	REG2501	LOG	Cromosoma	38	6.14	
<i>trans</i> -sRNA	3368649	3368720	-	REG2502	LOG	Cromosoma	72	5.67	
<i>trans</i> -sRNA	3406459	3406584	+	REG1544	LOG	Cromosoma	126	13.09	
<i>trans</i> -sRNA	3406599	3406628	+	REG1545	LOG	Cromosoma	30	7.44	
<i>trans</i> -sRNA	3407384	3407414	+	REG1546	LOG	Cromosoma	31	6.52	
<i>cis</i> -sRNA antisentido	3412214	3412242	+	REG560	EST	Cromosoma	29	8.01	SMc03297
<i>trans</i> -sRNA	3413556	3413583	-	REG2508	LOG	Cromosoma	28	9.55	
<i>trans</i> -sRNA	3430024	3430099	-	REG2517	LOG	Cromosoma	76	10.08	
<i>trans</i> -sRNA	3430148	3430167	-	REG2518	LOG	Cromosoma	20	5.83	
<i>trans</i> -sRNA	3430212	3430235	-	REG2519	LOG	Cromosoma	24	6.31	
<i>cis</i> -sRNA antisentido	3431263	3431297	+	REG1547	LOG	Cromosoma	35	10.79	SMc03765
<i>trans</i> -sRNA	3432934	3433001	-	REG891	EST	Cromosoma	68	15.27	
<i>trans</i> -sRNA	3496396	3496464	+	REG1562	LOG	Cromosoma	69	6.08	
<i>trans</i> -sRNA	3499737	3499756	-	REG2549	LOG	Cromosoma	20	6.81	
<i>cis</i> -sRNA antisentido	3526728	3526758	+	REG1571	LOG	Cromosoma	31	6.56	SMc03862
<i>cis</i> -sRNA antisentido	3526881	3527002	+	REG1572	LOG	Cromosoma	122	11.13	SMc03862
<i>trans</i> -sRNA	3527027	3527066	+	REG1573	LOG	Cromosoma	40	9.35	

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>trans</i> -sRNA	3527070	3527111	+	REG1574	LOG	Cromosoma	42	8.03	
<i>trans</i> -sRNA	3539589	3539629	+	REG564	EST	Cromosoma	41	11.61	SmelC781
<i>cis</i> -sRNA antisentido	3548337	3548361	-	REG2561	LOG	Cromosoma	25	6.72	SMc03880
<i>cis</i> -sRNA antisentido	3548363	3548404	-	REG2562	LOG	Cromosoma	42	7.85	SMc03880
<i>cis</i> -sRNA antisentido	3548414	3548441	-	REG2563	LOG	Cromosoma	28	5.06	SMc03880
<i>cis</i> -sRNA antisentido	3548463	3548506	-	REG2564	LOG	Cromosoma	44	6.40	SMc03880
<i>trans</i> -sRNA	3557642	3557665	-	REG2570	LOG	Cromosoma	24	6.08	
<i>cis</i> -sRNA antisentido	3652640	3652662	+	REG1598	LOG	Cromosoma	23	15.51	SMc02795
<i>trans</i> -sRNA	1211	1248	+	REG1	EST	pSymA	38	6.99	
<i>trans</i> -sRNA	14378	14399	+	REG14	EST	pSymA	22	5.76	
<i>cis</i> -sRNA antisentido	36359	36465	-	REG137	EST	pSymA	107	25.23	SMa0067
<i>cis</i> -sRNA antisentido	70848	70871	-	REG139	EST	pSymA	24	5.19	SMa0132
<i>cis</i> -sRNA antisentido	70901	70969	-	REG140	EST	pSymA	69	8.09	SMa0132
<i>trans</i> -sRNA	103639	103662	+	REG31	EST	pSymA	24	5.73	
<i>cis</i> -sRNA antisentido	108675	108739	+	REG25	LOG	pSymA	65	6.82	SMa0197
<i>cis</i> -sRNA antisentido	129343	129380	+	REG42	EST	pSymA	38	10.83	SMa0233
<i>trans</i> -sRNA	162181	162207	-	REG146	EST	pSymA	27	5.96	
<i>cis</i> -sRNA antisentido	238999	239037	-	REG302	LOG	pSymA	39	7.75	SMa0447
<i>cis</i> -sRNA antisentido	239442	239476	-	REG149	EST	pSymA	35	9.46	SMa0447
<i>cis</i> -sRNA antisentido	239740	240290	+	REG50	LOG	pSymA	551	21.69	SMa0448
<i>cis</i> -sRNA antisentido	240034	240063	+	REG46	EST	pSymA	30	5.93	SMa0448
<i>cis</i> -sRNA antisentido	240405	240452	+	REG51	LOG	pSymA	48	8.37	SMa0448
<i>trans</i> -sRNA	244565	244596	+	REG53	LOG	pSymA	32	6.08	
<i>trans</i> -sRNA	244705	244773	+	REG54	LOG	pSymA	69	7.13	
<i>trans</i> -sRNA	253665	253734	+	REG53	EST	pSymA	70	11.21	
<i>trans</i> -sRNA	314081	314100	-	REG150	EST	pSymA	20	5.26	
<i>trans</i> -sRNA	314106	314128	-	REG151	EST	pSymA	23	7.80	
<i>trans</i> -sRNA	314116	314222	-	REG307	LOG	pSymA	107	7.31	
<i>cis</i> -sRNA antisentido	314341	314369	-	REG152	EST	pSymA	29	5.19	SMa0590
<i>cis</i> -sRNA antisentido	314385	314428	-	REG153	EST	pSymA	44	9.01	SMa0590
<i>trans</i> -sRNA	340293	340317	-	REG311	LOG	pSymA	25	10.09	
<i>trans</i> -sRNA	342367	342406	+	REG69	LOG	pSymA	40	9.13	

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>trans</i> -sRNA	344816	344933	+	REG70	LOG	pSymA	118	24.11	
<i>trans</i> -sRNA	344935	344988	+	REG71	LOG	pSymA	54	9.54	
<i>trans</i> -sRNA	358240	358291	+	REG76	LOG	pSymA	52	9.01	
<i>trans</i> -sRNA	358482	358569	+	REG77	LOG	pSymA	88	10.77	
<i>trans</i> -sRNA	374988	375016	-	REG347	LOG	pSymA	29	7.14	
<i>trans</i> -sRNA	379666	379694	+	REG90	LOG	pSymA	29	6.08	
<i>cis</i> -sRNA antisentido	390928	390981	-	REG156	EST	pSymA	54	6.41	SMa0717
<i>trans</i> -sRNA	394921	394953	-	REG348	LOG	pSymA	33	9.40	
<i>trans</i> -sRNA	396520	396545	-	REG349	LOG	pSymA	26	6.08	
<i>trans</i> -sRNA	442191	442265	+	REG99	LOG	pSymA	75	15.48	
<i>trans</i> -sRNA	442991	443060	+	REG107	LOG	pSymA	70	8.23	
<i>trans</i> -sRNA	472522	472546	+	REG111	LOG	pSymA	25	9.49	
<i>trans</i> -sRNA	472562	472621	+	REG112	LOG	pSymA	60	8.33	
<i>cis</i> -sRNA antisentido	487749	487800	+	REG66	EST	pSymA	52	8.83	SMa0875
<i>cis</i> -sRNA antisentido	495486	495524	-	REG363	LOG	pSymA	39	7.60	SMa0888
<i>trans</i> -sRNA	501991	502015	+	REG115	LOG	pSymA	25	7.63	
<i>cis</i> -sRNA antisentido	504781	504804	-	REG364	LOG	pSymA	24	8.35	SMa0905
<i>trans</i> -sRNA	526325	526373	+	REG67	EST	pSymA	49	16.33	
<i>trans</i> -sRNA	528791	528820	+	REG124	LOG	pSymA	30	6.97	
<i>trans</i> -sRNA	546339	546369	-	REG368	LOG	pSymA	31	9.02	
<i>trans</i> -sRNA	546390	546424	-	REG369	LOG	pSymA	35	7.67	
<i>trans</i> -sRNA	552802	552997	+	REG125	LOG	pSymA	196	16.41	SmelA035
<i>cis</i> -sRNA antisentido	597694	597730	+	REG138	LOG	pSymA	37	9.45	SMa1092
<i>cis</i> -sRNA antisentido	597951	598222	+	REG139	LOG	pSymA	272	18.63	SMa1092
<i>trans</i> -sRNA	598639	598675	+	REG142	LOG	pSymA	37	9.01	
<i>trans</i> -sRNA	598677	598708	+	REG143	LOG	pSymA	32	6.61	
<i>trans</i> -sRNA	608442	608477	-	REG384	LOG	pSymA	36	6.85	
<i>cis</i> -sRNA antisentido	613671	613702	-	REG385	LOG	pSymA	32	8.76	SMa1124
<i>cis</i> -sRNA antisentido	628503	628530	+	REG164	LOG	pSymA	28	9.16	SMa1151
<i>cis</i> -sRNA antisentido	632399	632626	-	REG391	LOG	pSymA	228	17.19	SMa1155
<i>cis</i> -sRNA antisentido	632718	632794	-	REG392	LOG	pSymA	77	6.55	SMa1155
<i>trans</i> -sRNA	643646	643692	-	REG393	LOG	pSymA	47	9.34	

Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)

<i>trans</i> -sRNA	643738	643760	-	REG394	LOG	pSymA	23	7.37	
<i>cis</i> -sRNA antisentido	675920	675956	-	REG159	EST	pSymA	37	8.56	SMa1231
<i>trans</i> -sRNA	699688	699794	+	REG212	LOG	pSymA	107	6.96	
<i>trans</i> -sRNA	699837	699865	+	REG213	LOG	pSymA	29	5.78	
<i>trans</i> -sRNA	699901	700001	+	REG214	LOG	pSymA	101	9.10	
<i>cis</i> -sRNA antisentido	706299	706376	-	REG434	LOG	pSymA	78	10.26	SMa1294
<i>trans</i> -sRNA	706414	706476	-	REG435	LOG	pSymA	63	21.20	
<i>trans</i> -sRNA	706521	706556	-	REG436	LOG	pSymA	36	6.97	
<i>trans</i> -sRNA	706697	706774	-	REG437	LOG	pSymA	78	13.83	
<i>cis</i> -sRNA antisentido	734029	734139	-	REG445	LOG	pSymA	111	14.46	SMa1341
<i>trans</i> -sRNA	879072	879110	-	REG466	LOG	pSymA	39	6.28	
<i>trans</i> -sRNA	879121	879147	-	REG467	LOG	pSymA	27	6.08	
<i>trans</i> -sRNA	879245	879373	-	REG468	LOG	pSymA	129	9.68	
<i>trans</i> -sRNA	882499	882525	+	REG87	EST	pSymA	27	12.15	SmelA052
<i>cis</i> -sRNA antisentido	911624	911660	-	REG470	LOG	pSymA	37	7.00	AE006469__921
<i>trans</i> -sRNA	917185	917297	-	REG180	EST	pSymA	113	24.51	
<i>cis</i> -sRNA antisentido	920720	920746	+	REG232	LOG	pSymA	27	10.53	SMa1650
<i>cis</i> -sRNA antisentido	940372	940394	+	REG233	LOG	pSymA	23	6.19	SMa1676
<i>trans</i> -sRNA	956888	956953	-	REG475	LOG	pSymA	66	8.99	
<i>trans</i> -sRNA	957730	957793	-	REG479	LOG	pSymA	64	15.02	
<i>trans</i> -sRNA	957765	957793	-	REG182	EST	pSymA	29	7.65	
<i>trans</i> -sRNA	1008584	1008604	+	REG237	LOG	pSymA	21	6.08	
<i>trans</i> -sRNA	1021497	1021533	+	REG239	LOG	pSymA	37	14.82	
<i>trans</i> -sRNA	1021513	1021653	+	REG103	EST	pSymA	141	127.24	SmelA060
<i>trans</i> -sRNA	1082539	1082600	+	REG240	LOG	pSymA	62	12.93	
<i>trans</i> -sRNA	1082584	1082651	+	REG118	EST	pSymA	68	8.21	
<i>trans</i> -sRNA	1085493	1085572	-	REG487	LOG	pSymA	80	10.08	
<i>trans</i> -sRNA	1085634	1085780	-	REG488	LOG	pSymA	147	13.68	
<i>cis</i> -sRNA antisentido	1161469	1161504	+	REG249	LOG	pSymA	36	6.08	SMa2057
<i>cis</i> -sRNA antisentido	1161665	1161695	+	REG250	LOG	pSymA	31	7.09	SMa2057
<i>trans</i> -sRNA	1161748	1161884	+	REG125	EST	pSymA	137	76.67	SmelA072
<i>cis</i> -sRNA antisentido	1194013	1194042	+	REG253	LOG	pSymA	30	6.08	SMa2113

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>trans</i> -sRNA	1220727	1220765	+	REG255	LOG	pSymA	39	6.46	SmelA075
<i>trans</i> -sRNA	1249211	1249240	+	REG257	LOG	pSymA	30	8.29	
<i>trans</i> -sRNA	1253024	1253061	+	REG127	EST	pSymA	38	9.87	
<i>cis</i> -sRNA antisentido	1253129	1253176	+	REG128	EST	pSymA	48	6.51	SMa2237
<i>cis</i> -sRNA antisentido	1328088	1328119	-	REG512	LOG	pSymA	32	8.03	SMa2355
<i>cis</i> -sRNA antisentido	52217	52293	+	REG940	EST	pSymB	77	9.52	SMB20042
<i>cis</i> -sRNA antisentido	74662	74788	+	REG948	EST	pSymB	127	10.64	SMB20067
<i>cis</i> -sRNA antisentido	74796	74860	+	REG949	EST	pSymB	65	7.51	SMB20067
<i>trans</i> -sRNA	75719	75793	+	REG951	EST	pSymB	75	5.88	
<i>cis</i> -sRNA antisentido	87365	87596	-	REG1218	EST	pSymB	232	40.80	SMB20078
<i>trans</i> -sRNA	87430	87532	-	REG2817	LOG	pSymB	103	13.25	
<i>trans</i> -sRNA	87624	87658	-	REG2818	LOG	pSymB	35	6.60	
<i>cis</i> -sRNA antisentido	94934	94982	+	REG2599	LOG	pSymB	49	7.51	SMB20084
<i>cis</i> -sRNA antisentido	94988	95011	+	REG2600	LOG	pSymB	24	7.80	SMB20084
<i>cis</i> -sRNA antisentido	95045	95066	+	REG2601	LOG	pSymB	22	6.19	SMB20084
<i>cis</i> -sRNA antisentido	95068	95170	+	REG962	EST	pSymB	103	15.33	SMB20084
<i>cis</i> -sRNA antisentido	95108	95167	+	REG2602	LOG	pSymB	60	9.72	SMB20085
<i>cis</i> -sRNA antisentido	95176	95213	+	REG2603	LOG	pSymB	38	7.28	SMB20085
<i>cis</i> -sRNA antisentido	101415	101442	-	REG1227	EST	pSymB	28	9.46	SMB20093
<i>trans</i> -sRNA	242111	242134	+	REG991	EST	pSymB	24	6.84	
<i>trans</i> -sRNA	251063	251095	-	REG1270	EST	pSymB	33	11.69	
<i>trans</i> -sRNA	268354	268424	-	REG2820	LOG	pSymB	71	7.36	
<i>trans</i> -sRNA	278901	278939	-	REG1273	EST	pSymB	39	6.85	
<i>cis</i> -sRNA antisentido	414415	414521	+	REG1020	EST	pSymB	107	10.42	SMB20400
<i>cis</i> -sRNA antisentido	465912	465938	+	REG2651	LOG	pSymB	27	8.06	SMB20453
<i>cis</i> -sRNA antisentido	467263	467359	+	REG1029	EST	pSymB	97	26.12	SmelB038
<i>cis</i> -sRNA antisentido	476456	476503	+	REG1030	EST	pSymB	48	11.07	SmelB041
<i>cis</i> -sRNA antisentido	495916	495976	-	REG1283	EST	pSymB	61	5.74	SMB20481
<i>trans</i> -sRNA	545331	545618	-	REG1284	EST	pSymB	288	47.73	SmelB047
<i>trans</i> -sRNA	575765	575900	-	REG1285	EST	pSymB	136	13.19	
<i>cis</i> -sRNA antisentido	580608	580644	+	REG2659	LOG	pSymB	37	5.83	SmelB054
<i>trans</i> -sRNA	583141	583164	+	REG1034	EST	pSymB	24	14.19	

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	594107	594147	+	REG1035	EST	pSymB	41	7.87	SMb20808
<i>cis</i> -sRNA antisentido	602486	602513	-	REG2851	LOG	pSymB	28	6.68	SMb20814
<i>cis</i> -sRNA antisentido	602528	602556	-	REG2852	LOG	pSymB	29	9.32	SMb20814
<i>cis</i> -sRNA antisentido	602704	602798	-	REG2853	LOG	pSymB	95	11.95	SMb20814
<i>trans</i> -sRNA	613226	613256	-	REG2861	LOG	pSymB	31	5.32	
<i>trans</i> -sRNA	613343	613400	-	REG2862	LOG	pSymB	58	7.31	
<i>trans</i> -sRNA	667083	667106	-	REG1291	EST	pSymB	24	6.17	
<i>trans</i> -sRNA	668480	668524	+	REG2664	LOG	pSymB	45	8.46	
<i>trans</i> -sRNA	668539	668566	+	REG2665	LOG	pSymB	28	8.11	
<i>cis</i> -sRNA antisentido	708915	708944	+	REG2666	LOG	pSymB	30	7.61	SMb21072
<i>cis</i> -sRNA antisentido	708954	709068	+	REG2667	LOG	pSymB	115	14.48	SMb21072
<i>trans</i> -sRNA	709161	709268	+	REG2669	LOG	pSymB	108	8.71	
<i>cis</i> -sRNA antisentido	712383	712423	-	REG1292	EST	pSymB	41	12.34	SMb21075
<i>cis</i> -sRNA antisentido	723634	723672	-	REG1293	EST	pSymB	39	5.41	SMb21085
<i>cis</i> -sRNA antisentido	866740	866778	-	REG2900	LOG	pSymB	39	6.27	SMb21290
<i>cis</i> -sRNA antisentido	866810	866986	-	REG2901	LOG	pSymB	177	12.81	SMb21290
<i>cis</i> -sRNA antisentido	866988	867399	-	REG2902	LOG	pSymB	412	28.00	SMb21290
<i>cis</i> -sRNA antisentido	965678	965725	-	REG1305	EST	pSymB	48	6.82	SMb21211
<i>trans</i> -sRNA	998802	998885	+	REG1052	EST	pSymB	84	21.23	
<i>cis</i> -sRNA antisentido	1104353	1104494	-	REG2926	LOG	pSymB	142	18.39	SMb21565
<i>trans</i> -sRNA	1116020	1116145	+	REG1092	EST	pSymB	126	17.07	
<i>cis</i> -sRNA antisentido	1116148	1116283	+	REG1093	EST	pSymB	136	24.16	SMb21577
<i>cis</i> -sRNA antisentido	1116309	1116332	+	REG1094	EST	pSymB	24	5.88	SMb21577
<i>cis</i> -sRNA antisentido	1116343	1116456	+	REG1095	EST	pSymB	114	7.84	SMb21577
<i>cis</i> -sRNA antisentido	1116460	1116656	+	REG1096	EST	pSymB	197	9.65	SMb21577
<i>cis</i> -sRNA antisentido	1116661	1116795	+	REG1097	EST	pSymB	135	10.25	SMb21577
<i>cis</i> -sRNA antisentido	1120069	1120093	+	REG1100	EST	pSymB	25	7.55	SMb21580
<i>cis</i> -sRNA antisentido	1120114	1120150	+	REG1101	EST	pSymB	37	8.04	SMb21580
<i>cis</i> -sRNA antisentido	1124208	1124229	-	REG2940	LOG	pSymB	22	7.25	SMb21585
<i>cis</i> -sRNA antisentido	1181156	1181182	-	REG1324	EST	pSymB	27	7.10	SmelB108
<i>trans</i> -sRNA	1190853	1190892	+	REG2752	LOG	pSymB	40	6.32	
<i>cis</i> -sRNA antisentido	1204675	1204719	+	REG1117	EST	pSymB	45	18.33	SMb20973



**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	1228294	1228316	-	REG2958	LOG	pSymB	23	8.48	SMb21674
<i>cis</i> -sRNA antisentido	1258835	1258900	+	REG2758	LOG	pSymB	66	6.70	SMb20854
<i>cis</i> -sRNA antisentido	1258932	1258976	+	REG2759	LOG	pSymB	45	6.59	SMb20854
<i>cis</i> -sRNA antisentido	1258978	1259001	+	REG2760	LOG	pSymB	24	6.08	SMb20854
<i>cis</i> -sRNA antisentido	1285121	1285184	-	REG1325	EST	pSymB	64	10.09	SMb20878
<i>cis</i> -sRNA antisentido	1306017	1306067	+	REG2770	LOG	pSymB	51	8.52	SMb20898
<i>trans</i> -sRNA	1311375	1311395	-	REG1332	EST	pSymB	21	5.31	
<i>trans</i> -sRNA	1311419	1311451	-	REG1333	EST	pSymB	33	7.78	
<i>cis</i> -sRNA antisentido	1318719	1318756	+	REG2773	LOG	pSymB	38	5.53	SMb20909
<i>cis</i> -sRNA antisentido	1318726	1318756	+	REG1122	EST	pSymB	31	7.10	SMb20909
<i>cis</i> -sRNA antisentido	1318772	1318858	+	REG1123	EST	pSymB	87	6.88	SMb20909
<i>cis</i> -sRNA antisentido	1318860	1319040	+	REG1124	EST	pSymB	181	13.59	SMb20909
<i>cis</i> -sRNA antisentido	1318889	1318933	+	REG2774	LOG	pSymB	45	7.69	SMb20909
<i>cis</i> -sRNA antisentido	1318944	1319040	+	REG2775	LOG	pSymB	97	9.80	SMb20909
<i>trans</i> -sRNA	1325308	1325328	+	REG2776	LOG	pSymB	21	5.83	
<i>trans</i> -sRNA	1325446	1325620	+	REG1125	EST	pSymB	175	61.54	SmelB126
<i>trans</i> -sRNA	1325449	1325475	+	REG2777	LOG	pSymB	27	9.43	
<i>cis</i> -sRNA antisentido	1331039	1331158	+	REG2778	LOG	pSymB	120	7.04	SMb21399
<i>cis</i> -sRNA antisentido	1331135	1331174	+	REG1126	EST	pSymB	40	6.33	SMb21399
<i>cis</i> -sRNA antisentido	1331248	1331305	+	REG1127	EST	pSymB	58	9.20	SmelB128
<i>cis</i> -sRNA antisentido	1331255	1331326	-	REG1334	EST	pSymB	72	9.34	SmelB128
<i>trans</i> -sRNA	1335217	1335292	+	REG2779	LOG	pSymB	76	6.97	
<i>trans</i> -sRNA	1341186	1341205	-	REG2976	LOG	pSymB	20	9.61	SmelB130
<i>trans</i> -sRNA	1370561	1370720	-	REG1337	EST	pSymB	160	14.05	
<i>trans</i> -sRNA	1370722	1370804	-	REG1338	EST	pSymB	83	11.38	
<i>trans</i> -sRNA	1370874	1370900	-	REG1339	EST	pSymB	27	6.34	
<i>trans</i> -sRNA	1370902	1370993	-	REG1340	EST	pSymB	92	15.28	
<i>cis</i> -sRNA antisentido	1372300	1372426	+	REG1136	EST	pSymB	127	21.58	SMb21443
<i>cis</i> -sRNA antisentido	1380655	1380690	-	REG1356	EST	pSymB	36	6.31	SMb21450
<i>cis</i> -sRNA antisentido	1380692	1380724	-	REG1357	EST	pSymB	33	6.79	SMb21450
<i>cis</i> -sRNA antisentido	1380759	1381097	-	REG1358	EST	pSymB	339	40.57	SMb21450
<i>trans</i> -sRNA	1383648	1383676	-	REG1363	EST	pSymB	29	7.79	

**Tabla I-7. Candidatos correspondientes a las categorías de *trans*-sRNAs y *cis*-sRNAs antisentido identificados en nuestro estudio (continuación)**

<i>cis</i> -sRNA antisentido	1402637	1402746	-	REG1366	EST	pSymB	110	43.69	SMb21478
<i>trans</i> -sRNA	1404287	1404318	-	REG1370	EST	pSymB	32	6.93	
<i>cis</i> -sRNA antisentido	1406076	1406101	-	REG2983	LOG	pSymB	26	13.65	SMb21687
<i>trans</i> -sRNA	1406127	1406148	-	REG2985	LOG	pSymB	22	5.16	
<i>cis</i> -sRNA antisentido	1406190	1406386	-	REG2986	LOG	pSymB	197	16.37	SMb21485
<i>cis</i> -sRNA antisentido	1406200	1406268	-	REG1376	EST	pSymB	69	18.22	SMb21485
<i>cis</i> -sRNA antisentido	1406274	1406301	-	REG1377	EST	pSymB	28	9.65	SMb21485
<i>cis</i> -sRNA antisentido	1441337	1441375	+	REG2783	LOG	pSymB	39	6.08	SMb21517
<i>trans</i> -sRNA	1446036	1446079	-	REG1378	EST	pSymB	44	19.72	
<i>trans</i> -sRNA	1448701	1448750	-	REG1382	EST	pSymB	50	10.03	SmelB147b
<i>trans</i> -sRNA	1450376	1450406	+	REG1164	EST	pSymB	31	6.91	
<i>trans</i> -sRNA	1450453	1450526	+	REG1165	EST	pSymB	74	11.44	
<i>cis</i> -sRNA antisentido	1509223	1509246	+	REG2787	LOG	pSymB	24	9.24	SMb20707
<i>cis</i> -sRNA antisentido	1598691	1598798	+	REG1183	EST	pSymB	108	9.30	AL591985_1542
<i>cis</i> -sRNA antisentido	1605984	1606007	+	REG2814	LOG	pSymB	24	10.14	SMb20591
<i>trans</i> -sRNA	1678520	1678539	-	REG3020	LOG	pSymB	20	9.67	

<sup>1</sup> Todos los sRNAs de nueva anotación se indican sombreados

<sup>2</sup> La categoría GenDB se corresponde en el caso de sRNAs antisentido de nueva anotación al ORF con el que solapan. En el caso de los anotados en la base de datos se indica el nombre su nombre en la misma.

## OLIGONUCLEÓTIDOS UTILIZADOS

Tabla I-8. Oligonucleótidos utilizados en este trabajo.

Nombre	Aplicación	Secuencia
Smr7C	Detección mediante Northern de Smr7C	5'-ACCAGATGAGGACAAAGGCCTCATC-3' 5'-GATGAGGCCTTTGTCCTCATCTGGT-3'
Smr9C	Detección mediante Northern de Smr9C	5'-CGCGTGATCTTTAATCCGTTTCCGG-3' 5'-CCGAAACGGATTAAGATCACGCG-3'
Smr14C2	Detección mediante Northern de Smr14C	5'-TGCTTGATCTGATTGGCAACCGGGA-3' 5'-TCCCGTTGCCAATCAGATCAAGCA-3'
Smr15C1	Detección mediante Northern de Smr15C1	5'-GAGGAGAAAGCCGCTAGATGCACCA-3' 5'-TGGTGCATCTAGCGCTTTCTCCTC-3'
Smr15C2	Detección mediante Northern de Smr15C2	5'-ACTGGGAGGAGAAGCCACCAAAGAT-3' 5'-ATCTTTGGTGGCTTCCTCCCACT-3'
Smr22C	Detección mediante Northern de Smr22C	5'-TACTAGGTAGGTGGGCACCGTATGC-3' 5'-GCATACGGTGCCACCTACCTAGTA-3'
Smr45C	Detección mediante Northern de Smr45C	5'-CCGCACCGTCGTTGCTTCAAGATGT-3' 5'-ACATCTTGAAGCAACGACGGTGGCG-3'
Smr35B	Detección mediante Northern de Smr35B	5'-TGGTAAGCGATGATGAGGAAGGTTCG-3' 5'-CGACCTTCCTCATCATCGCTTACCA-3'
5'O-7C	Amplificación mediante RACE 5'	5'-TCTCCCCCTACTCCGGAAAC-3'
5'I-7C	Amplificación mediante RACE 5'	5'-CTACTCCGGAAACAGGAGTAGACCA-3'
5'O-9C	Amplificación mediante RACE 5'	5'-ATCTTTAATCCGTTTCCGGC-3'
5'I-9C	Amplificación mediante RACE 5'	5'-ATCCGTTTCCGGCGGCAGCGAACAA -3'
5'O-14C2	Amplificación mediante RACE 5'	5'-TGATTGGCAACCGGGAGGAGGTGT-3'
5'I-14C2	Amplificación mediante RACE 5'	5'-CAACCGGGAGGAGGAGTGTGCGCAAT -3'
5'O-15C1	Amplificación mediante RACE 5'	5'-GAAGGTTTCAAACCTCCAGA -3'
5'I-15C1	Amplificación mediante RACE 5'	5'-TGCAGCGGCTGGGGAGGAGAAA -3'
5'O-15C2	Amplificación mediante RACE 5'	5'-AAACCTCCAGAGGGGAACAGCTGCT -3'
5'I-15C2	Amplificación mediante RACE 5'	5'-AACAGCTGCTGCGGTGGCACT -3'
5'O-22C	Amplificación mediante RACE 5'	5'-TACTAGGTAGGTGGGCACCGTATGC -3'
5'I-22C	Amplificación mediante RACE 5'	5'-AGGGACAGTCCCCTTGAGAT -3'
3'O-22C	Amplificación mediante RACE 3'	5'-ATGTGCTAGAGTTCCCCTG -3'
3'I-22C	Amplificación mediante RACE 3'	5'-TAGAGTCCCCTGACAGGGA-3'
5'O-35B	Amplificación mediante RACE 5'	5'-AAACCGCACCGACCTTCCTCAT -3'
5'I-35B	Amplificación mediante RACE 5'	5'-CGACCTTCCTCATCATCGCTTACCA -3'
5'O-45C	Amplificación mediante RACE 5'	5'-TTGGAGAGCCGGCAGACACCAGAA-3'
5'I-45C	Amplificación mediante RACE 5'	5'-CCGCACCGTCGTTGCTTCAAGATGT-3'
RNasa P	Detección mediante Northern de RNasa P	5'-TCCCTATTCGGTCTTGCTCCCGGTG-3'

SRP	Detección mediante Northern de SRP	5'-AGAAGGTGGGAGGCTGGCACGATGA-3'
Hfq_Fw	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-TCCGCAACGGGCTCTTCTAT-3'
Hfq_Rv	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-TGCCCGCTTCGTATAGCCT-3'
3xFlag	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-CTAGAGATTACAAAGACCATGACGGTGATTA TAAAGATCATGACATCGACTACAAGGATGACGATG ACA-3'
3xFlag-i	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-AGCTTGTCATCGTCATCCTTGTAGTCGATGTCATG ATCTTTATAATCACCGTCATGGTCTTTGTAATCT-3'
5HfqTag	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-TCTAGATTTCCCTGATGCGGACAC-3'
3HfqTag	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-TCTAGAGGCAGCTTCTCATTCTC-3'
5FlxTag	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-AAGCTTTGACGGGATGTGAGGAAC-3'
3FlxTag	Generación del mutante <i>hfq</i> <sup>FLAG</sup>	5'-AAGCTTAAATGAAGCCGACGGTGT-3'
5SRNAFw	RT-PCR	5'-TGGTTCIGCGGGGTGGCTGC-3'
5SRNARv	RT-PCR y detección Northern 5S RNA	5'-TACTCTCCCGCTCTTAAGACGAA-3'
16S-RTFw	RT-PCR	5'-GGCTAGCGTIGTTCGGAATT-3
16S-RTRv	RT-PCR	5'-TCAGCGTCAGTTCAGACCA-3
7C-RTFw	RT-PCR	5'-GAGCGGCACTCATACAATGCC-3
7C-RTRv	RT-PCR	5'-CCGCCACGGCGCGGAGGCTGC-3
15C1-RTFw	RT-PCR	5'-AGGGCCACAAGAAACTTGGG-3
15C1-RTRv	RT-PCR	5'-AGCTGGTGCATCTAGCGGCT-3
15C2-RTFw	RT-PCR	5'-AGAGGGCCACAAGAAAAACT-3
15C2-RTRv	RT-PCR	5'-AGCTGATGCATCTTTGGTGGC-3

## PLÁSMIDOS UTILIZADOS.

Tabla I-9. Construcciones utilizadas en este trabajo.

Nombre	Descripción
pGEMHfq	pGEM-T con la región genómica y ORF de <i>hfq</i>
pKS3xFlag	Fragmento codificante en pBluescriptII KS+
pKS3xFlag5	pKS3xFlag con región promotora <i>hfq</i>
pKSHfq3xFlag	pBluescriptII KS+ con el ORF <i>hfq</i> <sup>FLAG</sup>
pGEMHfq3xFlag	pGEM-T con el ORF <i>hfq</i> <sup>FLAG</sup>
pK18Hfq3xFlag	pK18mobsacB con el ORF <i>hfq</i> <sup>FLAG</sup>



# **CAPÍTULO 2**

**BIOLOGÍA DE LA  
CHAPERONA  
DE RNA Hfq**



## II-1. INTRODUCCIÓN

La proteína Hfq fue descubierta a finales de los años 60 como representante de una de las primeras familias de proteínas de unión a RNA identificadas, conocidas genéricamente como Sm o LSm. Este grupo de proteínas poseen una distribución global en los tres dominios de la vida, *Eubacteria*, *Archaea* y *Eucariota* y la característica que las identifica es la presencia de una estructura cuaternaria multimérica en forma de anillo. Por su capacidad de unión a moléculas diversas de RNA Hfq y las proteínas tipo Sm son considerados reguladores globales de la expresión génica a nivel post-transcripcional. Las proteínas Sm incluyen componentes de la maduración de RNAs mensajeros, adaptadores de maquinarias de procesamiento de RNAs y chaperonas de RNAs ribosómicos (rRNAs), transferentes (tRNAs) y RNAs nucleolares pequeños (snoRNAs) (Wilusz and Wilusz, 2005).

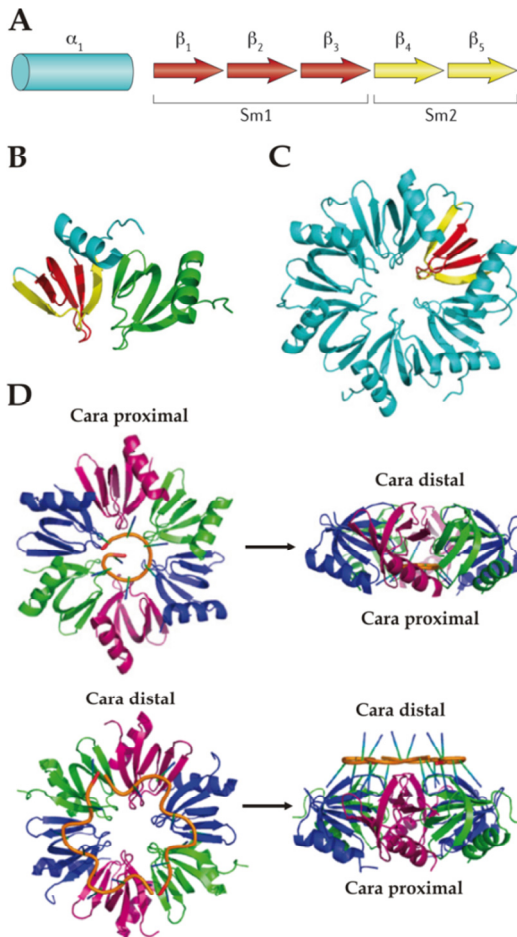
Hfq fue originalmente identificada en la bacteria modelo *E. coli*, como el factor del huésped, esencial para la replicación del bacteriófago Q $\beta$  (Host Factor Q $\beta$  bacteriophage; Franze de Fernandez *et al.*, 1968). Posteriormente se observó, que incluso en ausencia del fago, se detectaban niveles de este factor comparables con la cantidad de proteínas ribosómicas presentes en la célula (Kajitani *et al.*, 1994). Estudios bioquímicos posteriores confirmaron que en *E. coli* Hfq es una proteína abundante (unas 10.000 moléculas por célula en fase logarítmica), termo-resistente y con capacidad para unirse a ácidos nucleicos, preferentemente a zonas de cadena sencilla ricas en A-U de RNAs (Hori and Yanazaki, 1974; de Haseth and Uhlenbeck, 1980; de Haseth and Uhlenbeck, 1980).

### II-1.1. ESTRUCTURA DE LA PROTEÍNA Hfq

La estructura en forma de anillo, característica de las proteínas LSm, se genera, en el caso de Hfq, por ensamblaje de subunidades idénticas, aunque puede existir un plegamiento a partir de hetero-oligómeros en algunos miembros de la familia, como por ejemplo en proteínas del espliceosoma eucariótico (Vogel and Luisi, 2011).

En Hfq y otras proteínas Sm presentes en bacterias y arqueas la estructura de estas subunidades y la organización cuaternaria están muy conservadas en los diversos grupos taxonómicos, no siendo así el caso de la secuencia aminoacídica primaria (Hermann *et al.*, 1995). El núcleo conservado de las subunidades de Hfq, que consiste en un pliegue  $\alpha$ - $\beta_{1-5}$ , comprende aproximadamente 65 residuos y, en la mayoría de los casos, es la única parte que ha sido estructuralmente resuelta. Sin embargo, la estructura y función del extremo C-terminal variable de las proteínas Hfq permanecen aún sin dilucidar (Vecerek *et al.*, 2008).





**Fig. II-1. Estructura de Hfq.** A) Elementos que forman parte de la estructura secundaria de la proteína, incluyendo los motivos ultraconservados Sm1 y Sm2. B) y C) Cada unidad estructural la componen una  $\alpha$ -hélice y cinco láminas  $\beta$  ( $\alpha$ - $\beta_{1-5}$ ), donde las láminas  $\beta$  se disponen de forma antiparalela. Una de las láminas ( $\beta_2$ ) gira y se curva, permitiendo que se forme una estructura cerrada, del tipo barril (parte c). Las láminas  $\beta_4$  y  $\beta_5$  de la periferia de cada subunidad de Hfq exponen sus extremos amino, que pueden interaccionar con otras subunidades. D) Las caras para la interacción con RNAs (en naranja) están presentes a ambos lados del anillo. La cara proximal (la que contiene el N-terminal expuesto) incluye residuos del motivo Sm2. (Tomado de Vogel and Luisi, 2011)

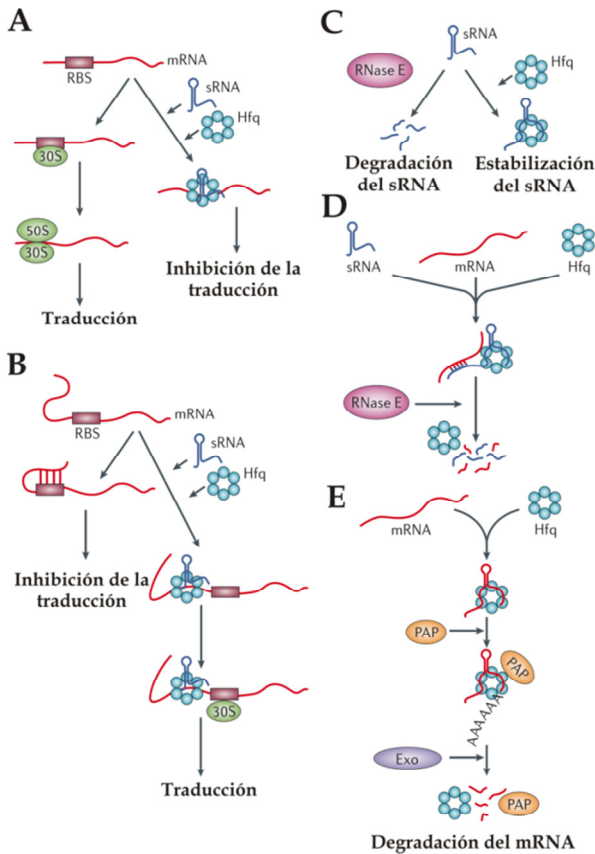
El núcleo consta de dos motivos conservados (32 y 14 aminoácidos, respectivamente), denominados Sm1 y Sm2 (Hermann *et al.*, 1995). Esta característica es la que identifica a Hfq y a las proteínas Sm como originadas a partir una proteína ancestral. El motivo Sm1 abarca las tres primeras cadenas ( $\beta_{1-3}$ ) de la estructura común  $\alpha$ - $\beta_{1-5}$  del núcleo estructural, y está altamente conservado en todos los miembros de esta familia (Kambach *et al.*, 1999) (Fig. II-1A, B y C). El motivo Sm2 está formado por las dos últimas cadenas ( $\beta_{4-5}$ ) y difiere entre las distintas proteínas Hfq bacterianas y más aún con las equivalentes en arqueas y eucariotas. Dado que los elementos del motivo Sm2 ocupan la interzona entre las distintas subunidades, ésta puede ser la causa de que este motivo influya en el número de subunidades que forman el anillo y la especificidad de sus interacciones, sobre todo en el caso de hetero-oligómeros observados en algunas proteínas Sm. Estos hetero-oligómeros pueden aparecer en bacterias que expresan múltiples homólogos de Hfq (Ramos *et*

*al.*, 2011). Además, las variantes de esta proteína con dos sets de dominios Sm parece que existen y pueden expresar dímeros unidos de forma natural (Yang *et al.*, 2010).

La arquitectura en forma de anillo expone dos caras para potenciales interacciones con ácidos nucleicos (Fig. II-1D). Para distinguirlas, normalmente se denomina “cara proximal” a la superficie que expone el extremo amino y la hélice  $\alpha$ , y “cara distal” a la contraria. Estudios de mutagénesis en ambas superficies de la proteína Hfq de *E. coli*, han revelado que existe una selección en la unión de RNAs influenciada por la cara de la proteína. Concretamente, se ha establecido un modelo de cómo Hfq puede interaccionar con los sRNAs y sus mRNA diana simultáneamente (Mikulecky *et al.*, 2004). Sin embargo, las dos superficies no actúan exclusivamente como “cara sRNA” y “cara mRNA”, dado que existen RNAs que se pueden unir a ambas partes (Fender *et al.*, 2010). La naturaleza repetitiva de cada superficie permite el contacto en diversos sitios con la misma molécula de RNA, y facilita el reclutamiento de RNAs competidores para favorecer el intercambio con RNAs unidos establemente (Fender *et al.*, 2010).

### II-1.2. MECANISMOS MOLECULARES Y RIBO-REGULACIÓN

Diversas especies de RNAs co-inmunoprecipitan con Hfq, pero el mayor enriquecimiento corresponde a sRNAs y mRNAs (Zhang *et al.*, 2003, Sittka *et al.*, 2008, Sittka *et al.*, 2009). Existen diversos mecanismos de regulación post-transcripcional mediados por Hfq que se esquematizan en la Fig. II-2 (Valentin-Hansen *et al.*, 2004, Aiba, 2007; Frohlich and Vogel, 2009; Chao and Vogel, 2010; Gottesman and Storz, 2010; Papenfort and Vogel, 2010).. En el primero de ellos, Hfq puede suprimir o activar la síntesis proteica por ayuda de un sRNA específico que se une a la región 5' de su mRNA diana (Fig. II-2A y B). Por esta misma vía, Hfq puede activar la traducción guiando la unión del sRNA a la región 5' del mensajero y produciendo un cambio estructural que expone el sitio de unión a ribosoma. En otro mecanismo, Hfq puede proteger al sRNA de la degradación por las ribonucleasas o promover esta misma degradación junto con el mRNA (Fig. II-1C y D). Por último, Hfq puede cooperar en el reciclaje de RNAs por asociación con los extremos 3' accesibles a la poliadenilación y la consecuente degradación exonucleotídica 3'-5' (Fig. II-E). En cada caso, el mecanismo preciso de acción de Hfq parece depender de información estructural contenida en las moléculas de RNA con las cuales interacciona.



**Fig. II-2. Modelos de actividad aceptados para Hfq.** A) Hfq en asociación con un sRNA puede secuestrar el sitio de unión a ribosoma (RBS) del mRNA diana, bloqueando la unión de las subunidades 30S y 50S y reprimiendo la traducción. B) En algunos mRNAs, existe una estructura secundaria en la región 5' no traducida (UTR) que puede enmascarar el RBS e inhibir la traducción. Un complejo formado por Hfq y su sRNA específico puede activar la traducción de este mRNA por exposición de la región del inicio de traducción para la unión de 30S. C) Hfq puede proteger al sRNA del ataque por ribonucleasas, generalmente RNasaE. D) Hfq puede inducir la degradación (generalmente vía RNasaE) de algunos sRNAs y sus mRNAs dianas. E) Hfq puede inducir la poliadenilación de un mRNA vía poli-A polimerasa (PAP), que puede generar una señal para la degradación 3'-5' por exorribonucleasas (Exo). (Tomado de Vogel and Luisi, 2011).

### II-1.3. LOCALIZACIÓN SUBCELULAR

Hfq puede localizarse en los distintos compartimentos celulares en bacterias; citoplasma (Azam *et al.*, 2000), membrana citoplasmática (Diestra *et al.*, 2009) o nucleoide (Kajitani *et al.*, 1994). La localización cercana a la membrana se ha relacionado con el supuesto de que sRNAs dependientes de Hfq regulen mRNAs que codifican proteínas de la cara interna de la membrana (Kawamoto *et al.*, 2005) o con su asociación a proteínas de membrana como RNasaE (Belasco, 2010). Algunos experimentos de inmunoprecipitación han demostrado que existen complejos Hfq-RNasa E y que esta asociación podría relacionarse con procesos de degradación/maduración de RNAs en el complejo nucleoproteico conocido como degradosoma (Morita *et al.*, 2005), pero actualmente no existen pruebas definitivas para apoyar esta hipótesis. La localización potencial en el DNA del nucleoide se ha apoyado en la visualización del dúplex DNA-proteína con una aparente constante de disociación de aproximadamente 400 nm (Updegrove *et al.*, 2010). Sin embargo, las propiedades de unión de Hfq a DNA han sido escasamente estudiadas, y la

profundización en este tema podría arrojar luz a diversos aspectos de la regulación génica que no son explicados por los efectos a nivel de RNA.

Otros estudios iniciales sugirieron que Hfq co-localiza con el ribosoma y, específicamente, con la subunidad 30S (Sukhodolets and Garges, 2003; Worhunsky *et al.*, 2003; Koleva *et al.*, 2006). Sin embargo, un estudio reciente desecha la idea de cualquier asociación ribosómica (Vecerek *et al.*, 2010) aunque ésta podría ser transitoria y dependiente de una traducción activa. Esto podría explicar la observación de que sRNA y mRNA diana pueden ser cargados sobre las subunidades 30S en una traducción temprana de los complejos de pre-iniciación (Worhunsky *et al.*, 2003).

#### II-1.4. DISTRIBUCIÓN Y BIOLOGÍA EN BACTERIAS

Al menos el 50% de las especies bacterianas cuya secuencia genómica es conocida poseen un ortólogo del gen *hfq* identificado en *E. coli*.

Hfq está presente en la mayoría de las grandes divisiones del dominio Bacteria. Sin embargo, en muchas bacterias en las que ha ocurrido una reducción importante del genoma debido a su estilo de vida (parásitos intracelulares obligados), Hfq está ausente, i.e. los géneros *Rickettsia*, *Chlamydia*, o *Mycoplasma* (Sun *et al.*, 2002). La presencia de este *locus* no se circunscribe a este dominio bacteriano, sino que también se extiende al dominio Archaea, incluso en organismos que carecen de genes tipo Sm (Sauter *et al.*, 2003). Esta distribución generalizada hace pensar que el origen de la proteína Hfq es monofilético, es decir, apareció en un ancestro común a todos los organismos (Sun *et al.*, 2002).

Dado que Hfq es un regulador global de la expresión génica y además presente en la mayoría de los grupos bacterianos, se ha observado que mutaciones en el gen *hfq* influyen severamente en la fisiología bacteriana, alterando la tasa de crecimiento, la morfología celular o la tolerancia a estreses ambientales (Tsui *et al.*, 1994). En diversas enterobacterias y otros patógenos intracelulares facultativos de mamíferos, estas deficiencias comprometen seriamente la virulencia, la motilidad, la invasión celular y/o la supervivencia en su nicho intracelular (Sonnleitner *et al.*, 2003; Christiansen *et al.*, 2004; Ding *et al.*, 2004; McNealy *et al.*, 2005; Sharma and Payne, 2006; Sittka *et al.*, 2007; Kulesus *et al.*, 2008). El fenotipo pleiotrópico conferido por mutaciones en el *locus hfq* es común también a todos los miembros de las  $\alpha$ -proteobacterias en donde la función de Hfq ha sido estudiada. Es el caso de *B. abortus*, donde los mutantes defectivos en Hfq muestran una reducción en la tasa de crecimiento, osmosensibilidad e incapacidad para sobrevivir dentro de los

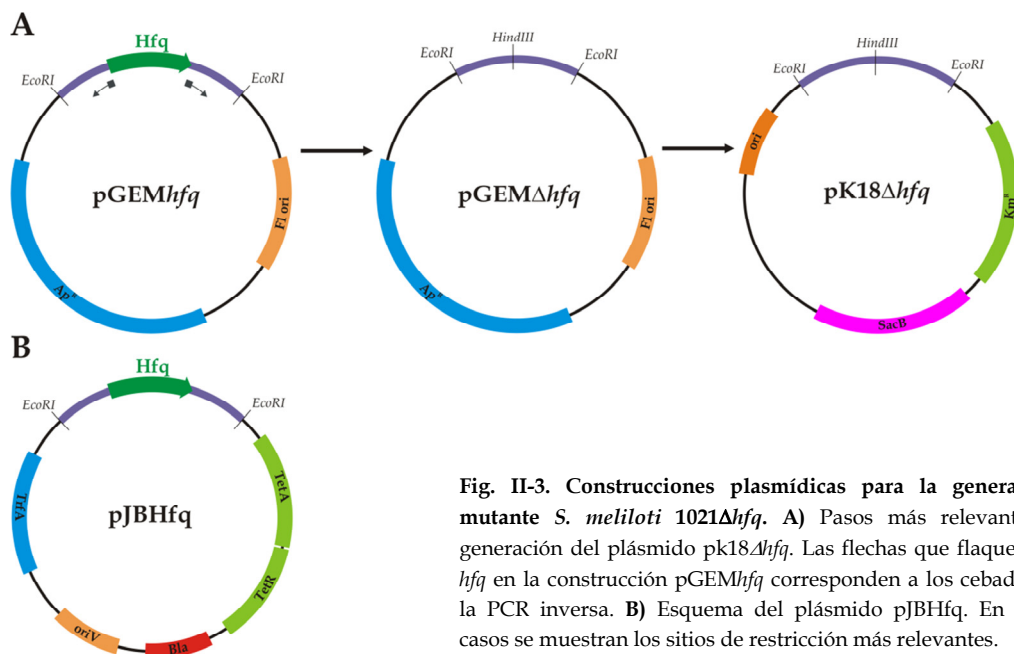
macrófagos a los que infecta, lo que sugiere un papel fundamental de esta proteína en el establecimiento y mantenimiento de infecciones intracelulares crónicas (Robertson and Roop, 1999; Roop *et al.*, 2002). Además de su contribución general a la tolerancia al estrés, se ha descrito también la influencia de Hfq en el proceso de fijación biológica de nitrógeno en  $\alpha$ -proteobacterias de vida libre (*Rhodobacter capsulatus*) y simbióticas (*Azorhizobium caulinodans* y *R. leguminosarum* bv. *viciae*). En estos microorganismos Hfq actúa como regulador post-transcripcional positivo de *nifA*, el gen que codifica el activador transcripcional principal de los genes que especifican el complejo de la nitrogenasa (Kaminski *et al.*, 1994; Drepper *et al.*, 2002; Zhang and Hong, 2009).

En este capítulo se describe la combinación de abordajes genéticos y genómicos para la identificación de procesos y genes (mRNAs) dependientes de Hfq y, en consecuencia, potencialmente ribo-regulados en *S. meliloti*.

## II-2. DISEÑO EXPERIMENTAL

### II-2.1. CONSTRUCCIÓN DEL MUTANTE *S. MELILOTI* $\Delta$ *hfq*

El procedimiento para la generación de un mutante de *S. meliloti* 1021 por delección del gen cromosómico *hfq* se esquematiza en la Fig X. Utilizando DNA genómico de esta bacteria como molde se amplificó, una región de 1.684 pb integrada por el ORF *hfq* de 243 nt y las regiones flanqueantes (714 pb hacia arriba y 727 pb hacia abajo del gen), con una polimerasa corregidora de errores *Pfu* y la pareja de cebadores Hfq\_Fw/Hfq\_Rv. Este fragmento de DNA se insertó en el plásmido pGEM-T Easy, generando pGEM*hfq*. A continuación, este plásmido se utilizó como molde para una PCR inversa, con la *Pfu* como enzima y con la pareja de cebadores divergentes *hfqi\_1/hfqi\_2*, que flaquean el ORF *hfq* y que contienen el sitio interno de restricción *HindIII*. El producto de amplificación se digirió con *HindIII* y se autoligó, generando el plásmido pGEM $\Delta$ *hfq*, que contiene la región genómica de 1.447 pb de *S. meliloti* 1021, donde el gen *hfq* ha sido reemplazado por el sitio *HindIII*. Esta región se recuperó de pGEM $\Delta$ *hfq* como fragmento *EcoRI*, y se clonó en el plásmido *pk18mobsacB* (no replicativo en *S. meliloti*), dando lugar a *pk18* $\Delta$ *hfq*. Éste fue movilizado a la cepa *S. meliloti* 1021 mediante cruce triparental y los fenómenos de doble recombinación se seleccionaron para el aislamiento de la cepa *S. meliloti* 1021 $\Delta$ *hfq* entre los transconjugantes.



**Fig. II-3. Construcciones plasmídicas para la generación del mutante *S. meliloti* 1021 $\Delta$ *hfq*.** A) Pasos más relevantes en la generación del plásmido pK18 $\Delta$ *hfq*. Las flechas que flanquean al gen *hfq* en la construcción pGEM*hfq* corresponden a los cebadores para la PCR inversa. B) Esquema del plásmido pJBHfq. En todos los casos se muestran los sitios de restricción más relevantes.

Para expresar Hfq bajo el control de su propio promotor y complementar el mutante obtenido se generó un fragmento de DNA de 842 pb que contiene la secuencia codificante de Hfq, además de la región de 571 nt situada por encima del gen, por amplificación con la enzima *Pfu* utilizando la pareja de oligonucleótidos 5Hfq\_C/3Hfq\_C y el plásmido pGEM*hfq* como molde. El producto de PCR se insertó en pGEM-T Easy dando lugar a pGEMHfq del que fue recuperado como fragmento *Eco*RI y finalmente insertado en este sitio de restricción del *polilinker* del plásmido de bajo número de copias pJB3Tc19, generando así pJBHfq. Este plásmido fue movilizado mediante cruce triparental típico al mutante *S. meliloti* 1021 $\Delta$ *hfq*.

## II-2.2. ANÁLISIS DEL PERFIL TRANSCRIPTÓMICO DE MUTANTE *S. MELILOTI* 1021 $\Delta$ *hfq*

Para la consecución del análisis de la hibridación sobre *microarrays* Sm14kOLI se tomaron los cDNAs correspondientes a las muestras de RNA total de la cepa control 1021 y el mutante 1021 $\Delta$ *hfq* y se marcaron con Cy5 y Cy3, respectivamente. De cada hibridación se realizaron tres réplicas biológicas procedentes de cultivos independientes de ambas cepas y el experimento se repitió dos veces, para obtener un total de 6 señales de hibridación que se tradujeron en 12 datos en cada sonda del chip.

## II-2.3. CONSTRUCCIÓN DE LA FUSIÓN HFQ:GFP Y ESTUDIO DE LA LOCALIZACIÓN DE LA PROTEÍNA EN *S. MELILOTI*

Para estudiar la localización de Hfq en células en vida libre de *S. meliloti* se diseñó una fusión traduccional de ésta al extremo amino de la proteína fluorescente GFP según el diseño experimental propuesto en Valencia-Burton and Broude, 2007, que se esquematiza en la Fig. II-4.

El primer paso es la generación de dos construcciones que expresen las dos proteínas que se van a fusionar, ambas en el mismo tipo de plásmido. En nuestro caso, se diseñaron las construcciones pGEM $hfq$ -tga (que contiene la secuencia codificante de Hfq sin el codón de terminación TGA) y pGEMeGFP (con la secuencia completa de la GFP). Para la construcción del plásmido pGEM $hfq$ -tga, se amplificó un fragmento que contiene la región promotora y la secuencia codificante del gen  $hfq$  (sin el codón TGA) con Phusion® DNA polimerasa (NEB) y la pareja de cebadores 5Hfq\_C/3Hfq\_fus utilizando pGEM $hfq$  como molde. Este fragmento se clonó en pGEM-T Easy en la orientación adecuada dando lugar a pGEM $hfq$ -tga. La segunda construcción, pGEMeGFP, se generó por restricción *Bam*HI de pBBs-eGFP, un plásmido que expresa GFP desde el promotor sintético syn creado en nuestro laboratorio. El fragmento generado se insertó en la orientación adecuada en pGEM-T, dando lugar a pGEMeGFP.

Una vez que se dispuso de las dos construcciones pGEM $hfq$ -tga y pGEMeGFP, se realizó la restricción con *Sac*II de la primera y *Nde*I de la segunda, generándose las versiones linearizadas de ambos plásmidos. Estos dos fragmentos se utilizaron como molde en una PCR anidada con la pareja de cebadores *ChimericHfqPhus/ChimericeGFPPhus*, que contienen una secuencia que codifica un polipéptido con 10 aminoácidos, flexible y rico en serinas que sirve de puente entre las dos proteínas. El amplicón se digirió con 10 U de la enzima *Dpn*I (NEB), se purificó de gel y se circularizó con 1 U de DNA ligasa (Roche), dando lugar a pGEMHfq:GFP. La fusión Hfq:GFP se recuperó de este plásmido y se insertó en pJB3Tc19 como fragmento *Eco*RI generando pJBHFQ:GFP, que finalmente se movilizó mediante cruce triparental a la cepa *S. meliloti* 1021 $\Delta hfq$ . Para visualizar la fluorescencia de las bacterias transformadas con pJBHFQ:GFP se tomaron 10  $\mu$ l de un cultivo líquido de las mismas en medio TY hasta fase logarítmica (D.O. 0.6) y se colocaron en un portaobjetos, para inmediatamente inmovilizar las células con 20  $\mu$ L de una solución de agarosa 1% en PBS. Las muestras se visualizaron en un microscopio invertido Leica DMI4000B con un filtro de fluorescencia específico para GFP (520 nm).

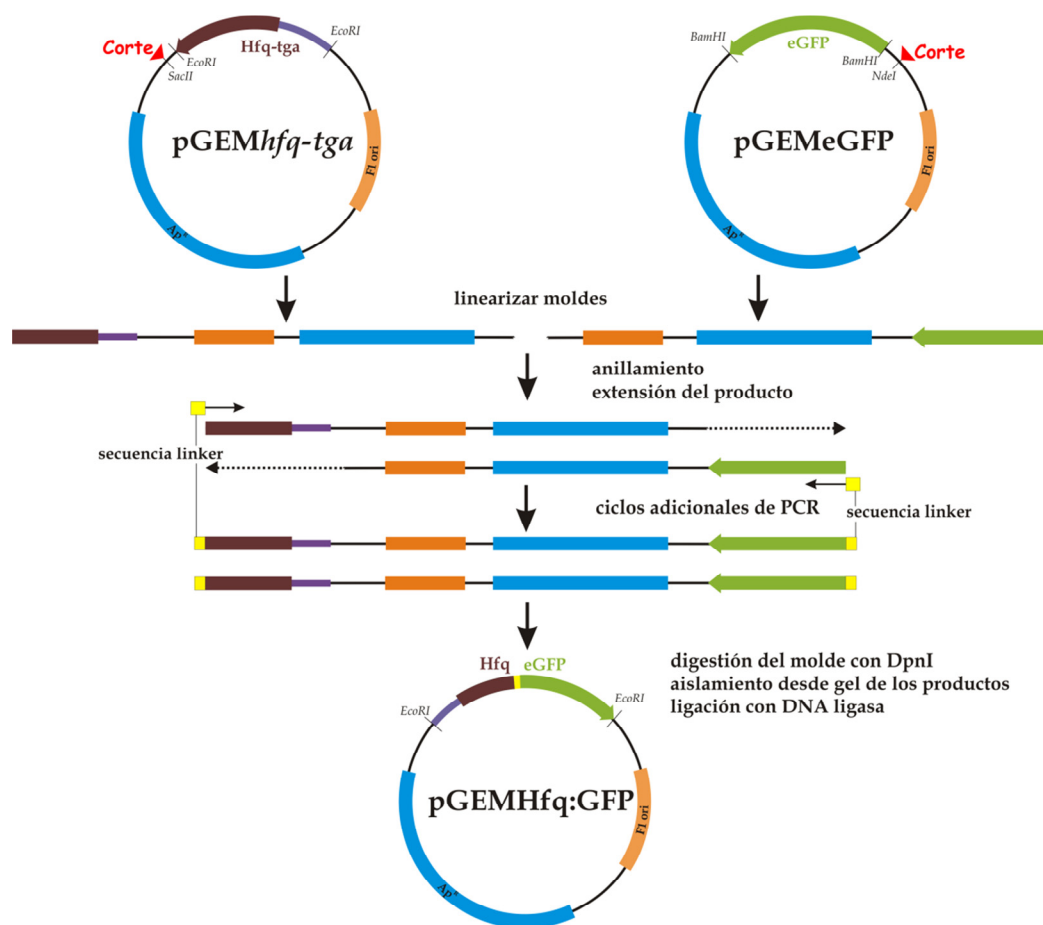


Fig. II-4. Etapas en la creación de la fusión traduccional Hfq:eGFP. Explicación en el texto

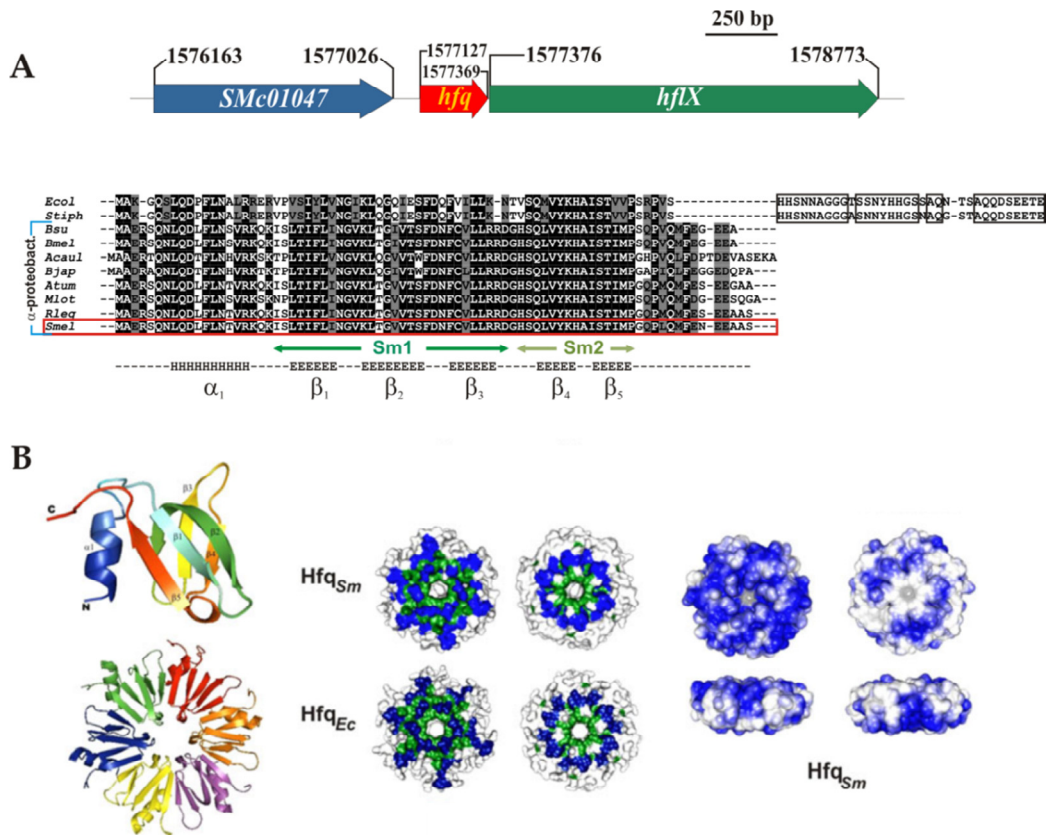
## II-3. RESULTADOS

### II-3.1. LA PROTEÍNA Hfq DE *S. MELILOTI*

El gen *hfq* corresponde al ORF *SMc01048* (anotado inicialmente como *nrfA*) del genoma de *S. meliloti* (<http://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi>), que se encuentra anotado entre las posiciones 1.277.127 y 1.577.369 del cromosoma de la cepa de referencia 1021 (Galibert *et al.*, 2001) (Fig. II-5A). El locus *hfq* de *S. meliloti* parece formar un operón dicistrónico con el gen *hflX*, situado hacia abajo, que codifica una proteína probable de unión a GTP. Hacia arriba de *hfq* se sitúan *SMc01047* y *trkA* que codifican una D-alanina aminotransferasa y un transportador de potasio, respectivamente. Inmediatamente hacia arriba del gen *trkA* se encuentra la unidad génica específica del sistema de asimilación de nitrógeno *ntr*



(*ntrB-ntrC-ntrY-ntrX*). Este ordenamiento genómico está conservado esencialmente en todos los endosimbiontes fijadores de nitrógeno del orden Rhizobiales. La excepción es la ausencia de homólogos de *trkA* o *SMc01047* entre el operón *ntr* y *hfq* en algunas especies (e.g. *M. loti*, *R. leguminosarum* bv *viciae*). Por el contrario, la región situada por encima del gen *hfq* en *S. meliloti* diverge totalmente de la de patógenos intracelulares animales (i.e. *Brucella* sp.) Los genomas de enterobacterias y  $\alpha$ -proteobacterias sólo comparten el gen *hflX* en esta región cromosómica.



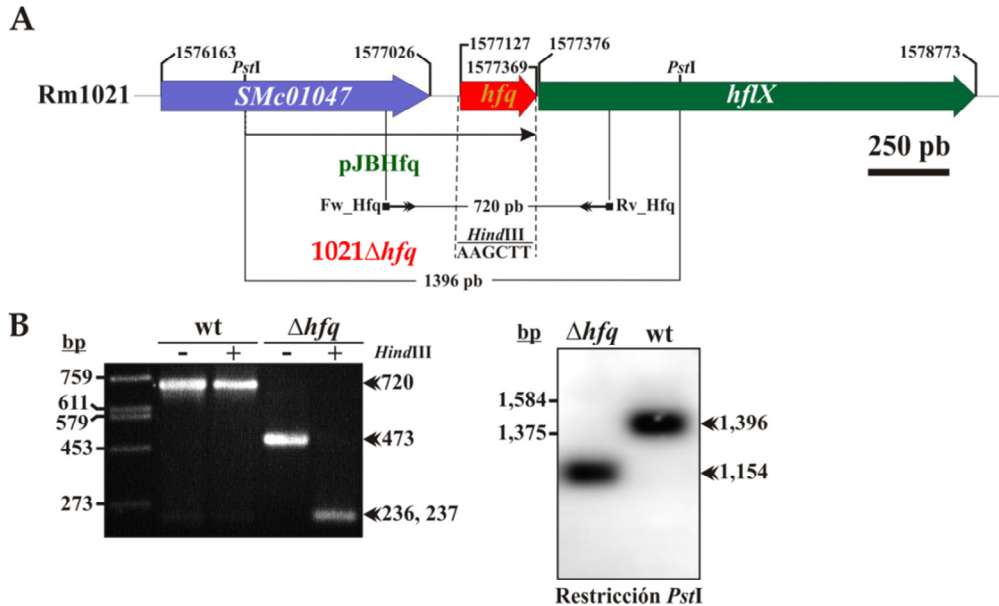
**Fig. II-5. Caracterización *in silico* del gen y la proteína Hfq de *S. meliloti*.** A) Organización de la región genómica de *hfq* (arriba) y alineamiento múltiple correspondiente a la secuencia de aminoácidos de proteínas Hfq codificadas en los genomas de enterobacterias y  $\alpha$ -proteobacterias. El diagrama genómico está dibujado a escala. Las abreviaturas correspondientes al alineamiento son: *Ecol*, *E. coli*; *Styph*, *S. thyphimurium*; *Bsu*, *B. suis*; *Bmel*, *B. melitensis*; *Acaul*, *A. caulinodans*; *Atum*, *A. tumefaciens*; *Mlot*, *M. loti*; *Rleg*, *R. leguminosarum*; *Smel*, *S. meliloti*. Las especies pertenecientes a subdivisión  $\alpha$  de las proteobacterias se indican a la izquierda. Los aminoácidos sombreados corresponden al menos a un 80% de identidad y el extremo C-terminal presente únicamente en enterobacterias se muestra encuadrado, además de los motivos ultraconservados Sm1 y Sm2. En la parte inferior se muestra la secuencia  $\alpha$ - $\beta_{1-5}$  correspondiente al núcleo principal (H, hélice; E, lámina  $\beta$ ) B) Modelo estructural predicho para el hexámero de la proteína Hfq de *S. meliloti* (Hfq<sub>Sm</sub>) utilizando la estructura de la Hfq de *E. coli* (Hfq<sub>Ec</sub>) como molde (parte izquierda) y representación tridimensional de la estructura cuaternaria de ambas proteínas (parte central), en la que se muestran los residuos conservados y los variables, así como la distribución de cargas electrostáticas (parte derecha) de la Hfq de *S. meliloti* (en azul). Parte B tomada de Sobrero and Valverde, 2011.

En *S. meliloti* 1021 el gen *hfq* codifica una proteína de 80 aminoácidos con el 72% de similaridad y el 45% de identidad con la proteína Hfq bien caracterizada de *E. coli* y el 77%-100% de identidad con sus homólogos en  $\alpha$ -proteobacterias. El análisis del alineamiento múltiple de algunas de estas secuencias aminoacídicas reveló que la conservación entre las proteínas Hfq de enterobacterias y rizobios está restringida a los dominios Sm1 y Sm2, que conforman el anillo de unión a ácidos nucleicos, mientras que los genomas de enterobacterias codifican una proteína con una extensión en el extremo C-terminal que no está presente en sus homólogas de  $\alpha$ -proteobacterias. Estudios *in silico* posteriores de Hfq de *S. meliloti* predicen la existencia de la secuencia  $\alpha$ - $\beta_{1-5}$  en el núcleo central y una estructura tridimensional similar a la obtenida por cristalización de la proteína de *E. coli* (Sobrero and Valverde, 2011) (Fig. II-5B).

### II-3.2. CARACTERIZACIÓN FUNCIONAL DEL GEN *hfq* EN *S. MELILOTI*

#### II-3.2.1. GENERACIÓN DE UN MUTANTE *hfq* EN *S. MELILOTI*

Como abordaje genético para la caracterización funcional de Hfq en *S. meliloti* se generó un mutante en la cepa de referencia 1021 mediante una doble recombinación mediada por *pk18mobsacB* que resultó en la delección completa de la secuencia codificante de esta proteína. Este *locus* fue reemplazado por un sitio de restricción *HindIII*, respetando la pauta de lectura del gen *hflX*, inmediatamente posterior a *hfq*. La amplificación de DNA genómico en lisados celulares de las cepas silvestre y mutante utilizando cebadores que flanquean el gen *hfq* generó en cada caso un único fragmento del tamaño esperado; 720 nt el correspondiente al alelo silvestre y 492 nt el del que no contiene *hfq*. Confirmando la introducción del sitio de restricción *HindIII* en lugar del gen *hfq*, la digestión de este segundo amplicón con este enzima generó dos fragmentos de tamaños casi idénticos (236 y 237 nt), que se observan como banda única en los geles de agarosa (Fig. II-6B, panel izquierdo). Finalmente, la hibridación del DNA genómico de ambas cepas previamente digerido con *PstI*, con el fragmento de 492 nt como sonda también confirmó la delección de *hfq* en la cepa *S. meliloti* 1021 $\Delta$ *hfq*., (Fig. II-6, panel derecho).



**Fig. II-6. Caracterización genética del mutante *S. meliloti* 1021 $\Delta hfq$ .** **A)** Región genómica del gen *hfq* (centro), en la que se indica la sustitución del lugar correspondiente al locus *hfq* por el sitio de restricción *HindIII* en el mutante 1021 $\Delta hfq$ , y el fragmento clonado en *pJBHfq* para la complementación de la mutación. Los sitios más relevantes en las verificaciones experimentales, así como los cebadores utilizados en las mismas, también se indican. **B)** Aproximaciones experimentales para la verificación de la mutación en 1021: análisis mediante Southern-blot sobre DNA genómico digerido con *PstI* (izquierda) del mutante 1021 $\Delta hfq$  y la cepa salvaje 1021; gel de agarosa correspondiente a los fragmentos de amplificación (derecha) de las mismas cepas, indicando la restricción (+) o no (-) de los mismos con la enzima *HindIII*.

### II-3.2.2. ANÁLISIS FENOTÍPICO DE LOS MUTANTES *hfq*.

#### II-3.2.2.1. Fenotipo en vida libre

##### II-3.2.2.1.1. Curvas de crecimiento

El mutante *S. meliloti* 1021 $\Delta hfq$  forma colonias de idéntica morfología a las de la cepa silvestre 1021 cuando crece en placa en cualquier medio de cultivo (TY o MM), aunque ya en el agar se adivina una pauta de crecimiento más lenta que la de la cepa de referencia. La monitorización del crecimiento de ambas bacterias en cultivo líquido confirmó que la mutante en el gen *hfq* presentaba un retraso a lo largo de toda la fase logarítmica alcanzando la fase estacionaria con menor densidad óptica en medio completo. El plásmido de bajo número de copias *pJBHfq*, que expresa el gen *hfq* desde su propio promotor complementó como era esperado este fenotipo de crecimiento (Fig. II-7).

Como se había observado previamente en otras bacterias estos resultados anticipan también un efecto pleiotrópico de la mutación de *hfq* en *S. meliloti*.

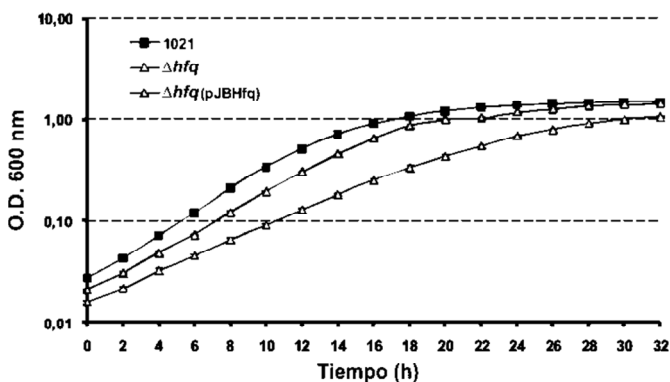


Fig. II-7. Fenotipo de crecimiento del mutante *S. meliloti* 1021 $\Delta hfq$ . Curvas de crecimiento en medio TY de la cepa salvaje de *S. meliloti* 1021, y el mutante 1021 $\Delta hfq$ . El crecimiento se midió a D.O. de 600 nm de cultivos independientes y por triplicado a intervalos de 2 horas. Leyendas: 1021, cepa salvaje de referencia;  $\Delta hfq$ , mutante por delección en 1021;  $\Delta hfq$  (pJBHfq), mutante *hfq* complementado con pJBHfq.

### II-3.2.2.1.2. Array fenotípico Biolog

Para profundizar en la caracterización en vida libre del mutante *S. meliloti* 1021 $\Delta hfq$ , esta cepa y la de referencia 1021 se sometieron a un análisis mediante los denominados *arrays* fenotípicos por la empresa Biolog (Phenotype microarrays™, Biolog® [[http://www.biolog.com/\\_pmMicrobialCells.shtml](http://www.biolog.com/_pmMicrobialCells.shtml)]). Estos *arrays* se comercializan en placas de 96 pocillos en las que se testa la capacidad de las bacterias para utilizar multitud de fuentes de energía (compuestos carbonados y nitrogenados, derivados de fósforo o azufre, etc) o su resistencia a diversos agentes (fungicidas, antibióticos, etc), mediante estimación indirecta de su crecimiento por medida de la tasa respiratoria como producción de NADH en cada caso. En este experimento se utilizaron en primer lugar 10 *arrays* correspondientes a múltiples fuentes nutritivas [192 fuentes de carbono, 96 fuentes de nitrógeno simples, 96 fuentes de fósforo y azufre, 96 microelementos (aminoácidos, vitaminas), 292 péptidos simples]. En placas independientes fueron testados además, 96 osmolitos y 96 elementos relacionados con el pH (ver anexo situado al final de este capítulo). Los resultados enviados por Biolog®, correspondientes a dos réplicas biológicas, revelaron que el mutante *S. meliloti* 1021 $\Delta hfq$  tiene patrones similares de utilización de las fuentes de carbono que la cepa silvestre. Sin embargo, la utilización de compuestos nitrogenados fue diferente entre ambas cepas en algunos casos. El mutante *hfq* respiró mejor que la cepa parental en presencia de los dipéptidos glicina-L-prolina (PM1 H01) y metionina-prolina (PM7 C06), así como del ácido p-hidroxifenil acético (PM1 H02).

A este set fenotípico se añadieron otros 10 *arrays* correspondientes a 960 elementos relacionados con sensibilidad química en bacterias (antibióticos, detergentes, bactericidas, etc) (ver anexo). Los datos de respiración (también derivados de dos experimentos independientes) mostraron que el mutante *S. meliloti* 1021 $\Delta$ hfq crece mejor que *S. meliloti* 1021 en presencia de bromuro de domifeno (PM15B D08) y toliilfluánida (PM20 BH05, H06, H07), ambos detergentes con actividad fungicida que actúan a nivel de membrana, y del ácido pipemídico (PM18CB08), un antibiótico del tipo quinolona que inhibe la actividad de la DNA topoisomerasa. Sin embargo, se observó que las tasas de respiración de *S. meliloti* 1021 $\Delta$ hfq fueron inferiores a las de la cepa salvaje en presencia de los antibióticos geneticina (PM13B E06,E07,E08) y tobramicina (PM12B F02,F03), ambos del tipo de los aminoglucósidos con actividad en la síntesis proteica así como y del telurito potásico (PM16A F04), un anión tóxico.

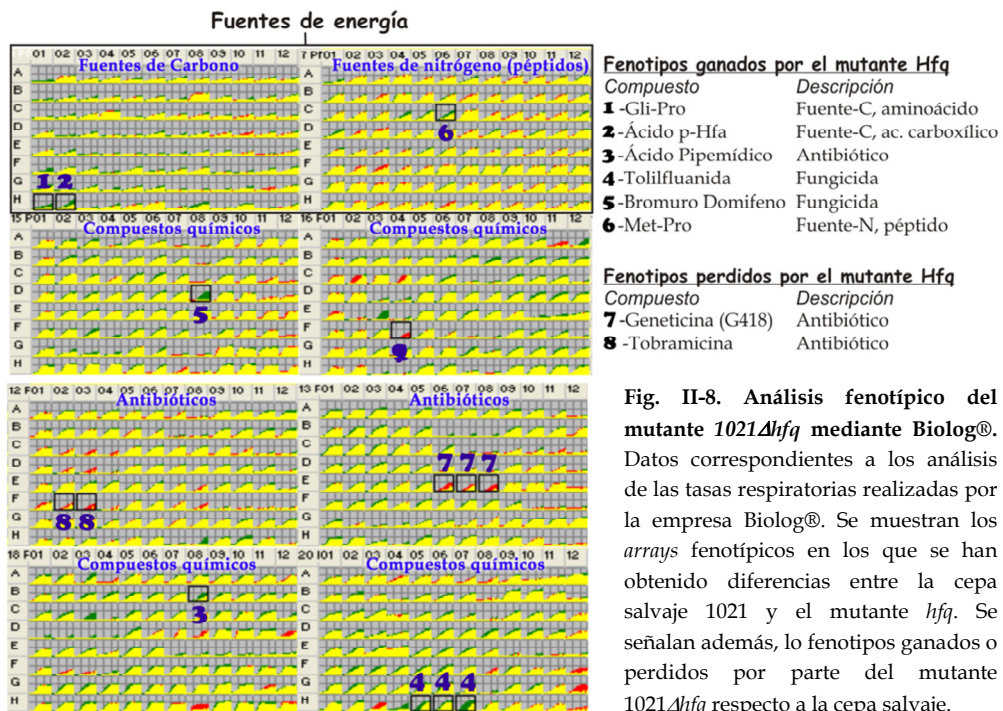


Fig. II-8. Análisis fenotípico del mutante 1021 $\Delta$ hfq mediante Biolog®. Datos correspondientes a los análisis de las tasas respiratorias realizadas por la empresa Biolog®. Se muestran los *arrays* fenotípicos en los que se han obtenido diferencias entre la cepa salvaje 1021 y el mutante *hfq*. Se señalan además, lo fenotipos ganados o perdidos por parte del mutante 1021 $\Delta$ hfq respecto a la cepa salvaje.

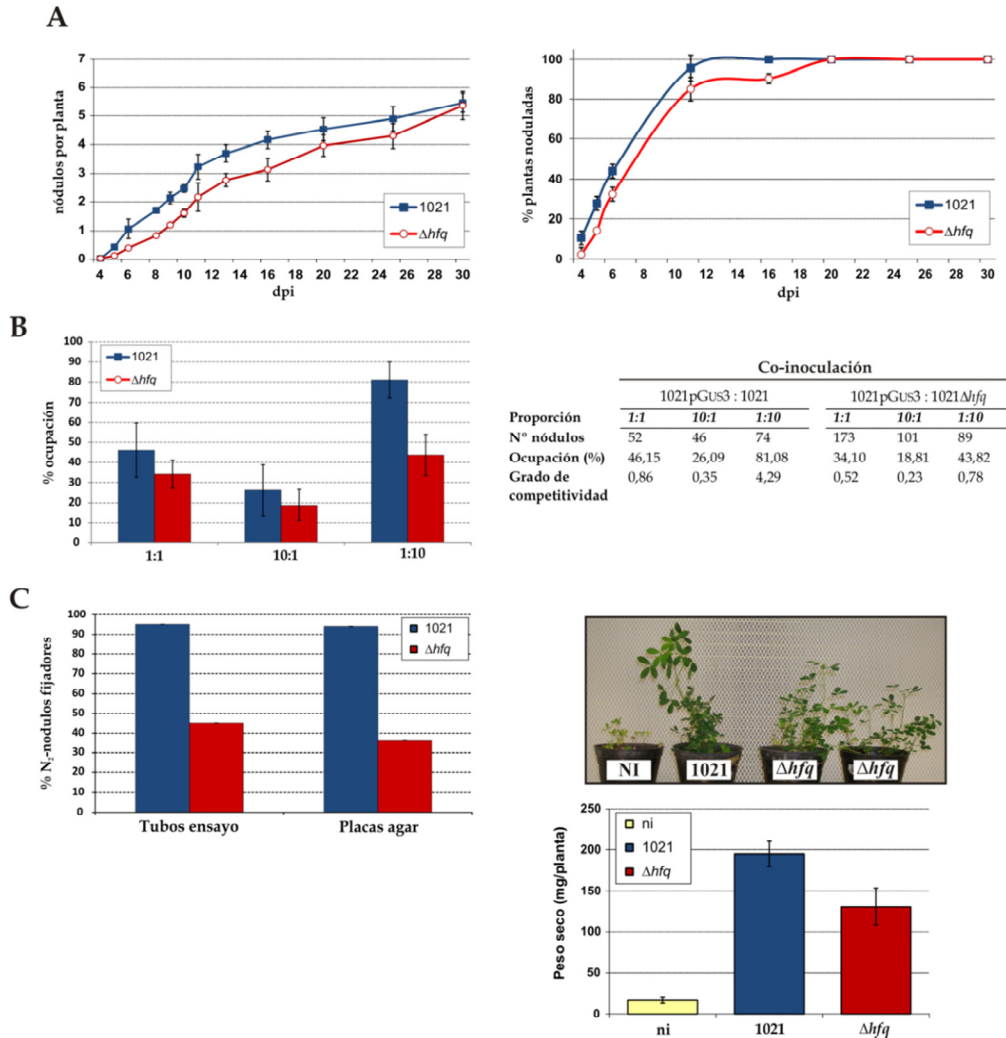
### II-3.2.2.2. Fenotipo durante la interacción simbiótica con *M. sativa*

#### II-3.2.2.2.1. Nodulación, competitividad y eficiencia de la simbiosis

Para evaluar el comportamiento simbiótico del mutante *hfq* en *S. meliloti* se llevaron a cabo en primer lugar experimentos de nodulación y de competitividad. Para los primeros, se inocularon independientemente con suspensiones bacterianas

de la cepa salvaje (1021) y el mutante *hfq* (1021 $\Delta$ *hfq*) sets de 24 plantas crecidas en cultivo hidropónico en tubos de ensayo. El número de nódulos por planta inducidos por cada cepa (infectividad) y el porcentaje de plantas noduladas (cinética de nodulación) se contabilizaron a intervalos diarios post-inoculación (dpi) en tres experimentos independientes. No se observaron diferencias significativas en el inicio de la nodulación (i.e. tiempo de aparición del primer nódulo) o en el número medio de nódulos por planta al final del experimento cuando se compararon ambas cepas (Fig. II-9A). Además, el mutante *hfq* fue capaz de nodular el 100% de las plantas inoculadas, al igual que la cepa 1021. Sin embargo la cinética de nodulación del mutante fue discretamente más lenta que la de la cepa silvestre como se infiere de ambas gráficas. Estos resultados revelan que la pérdida de Hfq no afecta a la capacidad de *S. meliloti* de formar nódulos en raíces de alfalfa, aunque el retraso en la nodulación del mutante probablemente denote cierta influencia de esta proteína en la adaptación de la bacteria al ambiente rizosférico.

Para los ensayos de competitividad se inocularon sets de 24 plantas de alfalfa crecidas en tubo de ensayo con mezclas en diferentes proporciones (1:1, 1:10 y 10:1) de la cepa mutante (*S. meliloti* 1021 $\Delta$ *hfq*) y una derivada de la silvestre 1021, que expresa constitutivamente el gen reportero *gus* desde el plásmido pGUS3 (*S. meliloti* 1021 pGUS3), [ver materiales y métodos]. En una serie simultánea de experimentos se sustituyó la cepa mutante por la silvestre *S. meliloti* 1021 en cada mezcla de co-inoculación, como control de la posible influencia del plásmido pGUS3 en la competitividad por la nodulación. Después de 30 días las raíces de todas las plantas inoculadas se sometieron a tinción GUS contándose posteriormente los nódulos totales de forma que los nódulos blancos se consideraron ocupados exclusivamente por la cepa *S. meliloti* 1021 $\Delta$ *hfq* o la control *S. meliloti* 1021 en su caso. En cada ensayo, a partir del número total de nódulos blancos y azules se calcularon dos parámetros: % de nódulos ocupados por la cepa testada (control 1021 o 1021 $\Delta$ *hfq*) respecto al total y grado de competitividad, calculado como la razón entre los nódulos ocupados por la cepa testada y la reportera (*S. meliloti* pGUS3) en cada caso. Los resultados de estos experimentos se muestran en la Fig. II-9B. En los experimentos control se observó que cuando las cepas *S. meliloti* 1021 y *S.*



**Fig. II-9. Fenotipo simbiótico del mutante 1021Δhfq en *S. meliloti*.** A) Infectividad y cinéticas de formación de nódulos de la cepa salvaje 1021 y el mutante 1021Δhfq determinados como el número de nódulos por planta (gráfico de la izquierda) y % de plantas noduladas (gráfico de la derecha), respectivamente. Cada punto representa la media ± el error típico de las determinaciones en dos sets de 24 plantas crecidas en cultivo hidropónico en tubos de ensayo. Dpi, días post-inoculación B) Ensayos de competitividad entre la cepa *S. meliloti* 1021 y el mutante 1021Δhfq. La ocupación nodular se expresa como porcentaje de nódulos invadidos por cada cepa problema. El grado de competitividad se expresa como la relación entre el número de nódulos ocupados por la cepa problema y la cepa control (GUS). C) Eficiencia simbiótica de las cepas 1021 y 1021Δhfq. El histograma de la izquierda representa el % de nódulos fijadores de nitrógeno (rosáceos) inducidos por cada depa en plantas crecidas en tubos de ensayo (dos sets de 24 plantas) y en placas de agar (5 placas con 10 plantas cada una) a los 30 dpi. Gráficos de la derecha: Desarrollo de plantas crecidas en jarras Leonard de 30 dpi con las cepas 1021 y 1021Δhfq y medida del peso seco expresada como la media ± error estándar de determinaciones en 24 plantas individuales. Ni, no inoculado.

*meliloti* 1021(pGUS3) se co-inoculan en proporción 1:1 el porcentaje de nódulos ocupado por la primera fue del 46,15%, que corresponde a un grado de competitividad próximo al 1. Cuando las proporciones de las cepas co-inoculadas no

fueron equivalentes (1:10 o 10:1) ambos parámetros variaron en correlación con la abundancia de cada cepa en las mezclas de co-inoculación. De estos resultados concluimos que el plásmido pGUS3 no afecta de forma significativa la competitividad por la nodulación de *S. meliloti* 1021, por lo que la cepa reportera *S. meliloti* 1021(pGUS3) puede utilizarse como referencia (silvestre) en estos experimentos. Por su parte, el mutante *S. meliloti* 1021 $\Delta$ hfq apenas fue capaz de ocupar el 34% de los nódulos inducidos en alfalfa cuando se inoculó en presencia de la cepa *S. meliloti* 1021(pGUS3) en proporción 1:1, lo que supone un grado de competitividad de 0,52. Cuando la proporción del mutante en la mezcla de co-inoculación fue 10 veces inferior a la cepa reportera este porcentaje de ocupación disminuyó hasta el 18,81% (grado de competitividad 0,23), mientras que en el caso contrario, cuando *S. meliloti* 1021 $\Delta$ hfq se inoculó en proporción 10 veces superior a la cepa de referencia el porcentaje de nódulos ocupados por el mutante ni siquiera alcanzó el 50% (43,82%), manteniendo su grado de competitividad por debajo de 1 (0,78). En resumen, estos datos revelan que la mutación en el locus *hfq* disminuye la competitividad de *S. meliloti* 1021 por la ocupación de nódulos en plantas de alfalfa.

Las mayores diferencias en el comportamiento simbiótico entre la cepa silvestre y el mutante *hfq* se observaron cuando se analizó el número final de nódulos fijadores (i.e. nódulos rosados que expresan activamente la leg-hemoglobina) inducidos por cada cepa cuando se inocularon independientemente en plantas de alfalfa. Este parámetro se determinó en plantas crecidas en tubos de ensayo o en placas de agar. Al final del experimento (30 dpi) el 95% de los nódulos generados por la cepa salvaje eran rosas, indicativo de una fijación de nitrógeno activa, independientemente de las condiciones de crecimiento de las plantas, mientras que el 55% (tubos de ensayo)-64% (placas de agar) de nódulos inducidos por el mutante fueron blancos (Fig. II-9C, panel izquierdo). Además, el primer nódulo rosa inducido por la cepa salvaje apareció en torno a los 13 dpi. Por el contrario, este tiempo se estimó en 18 dpi en plantas inoculadas con la cepa 1021 $\Delta$ hfq. Finalmente, el crecimiento de las plantas de alfalfa noduladas las cepas *S. meliloti* 1021 y 1021 $\Delta$ hfq también se comparó en experimentos realizados en jarras Leornard durante 30 días (Fig. II-9C, paneles sección derecha). Al final del experimento, las plantas inoculadas con el mutante *hfq* mostraron hojas con un color verde pálido y alcanzaron aproximadamente la mitad de la altura de las plantas inoculadas con *S. meliloti* 1021. Las determinaciones del peso seco de plantas individuales confirmaron esta percepción; el peso medio de las plantas inoculadas con la cepa 1021 $\Delta$ hfq fue aproximadamente del 64% (130 mg por planta) del de las inoculadas con la cepa salvaje (200 mg por planta). Estos resultados

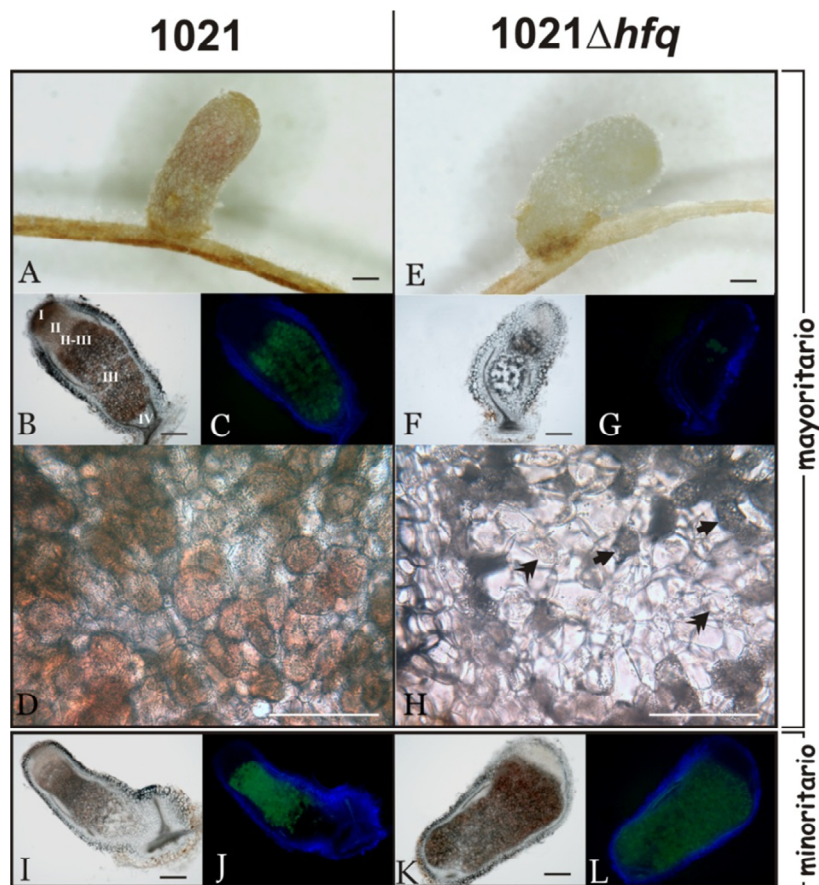


indican que Hfq también influye en estadios tardíos del proceso simbiótico y es requerida para el establecimiento de una simbiosis eficiente.

#### II-3.2.2.2.2. Histología y organogénesis nodular

Para analizar con más detalle el fenotipo endosimbiótico asociado a la mutación *hfq*, diseñamos experimentos de microscopía óptica de nódulos inducidos por la cepa salvaje *S. meliloti* 1021 y su derivada 1021 $\Delta$ *hfq* 30 dpi de las plantas de alfalfa (Fig. II-10). Los nódulos salvajes eran elongados y de color rosa indicativo de una fijación de nitrógeno activa (Fig. II-10A). La microscopía de campo claro de secciones longitudinales de estos nódulos reveló las sucesivas zonas características de la histología de nódulos indeterminados: el meristemo apical o zona I, la zona de infección o zona II, la interzona II-III donde comienza la diferenciación de los bacteroides, la zona de fijación activa de nitrógeno o zona III ocupada por bacteroides maduros y, finalmente, la zona de senescencia proximal o zona IV, que contiene células vegetales aparentemente vacías de bacteroides a microscopía de campo claro (Fig. II-10B) (Vasse *et al.*, 1990). Las imágenes combinadas de la misma sección nodular observadas bajo filtros para el verde y el azul (520 nm y 470 nm, respectivamente), confirmaron la colonización uniforme de los tejidos nodulares centrales por bacteroides diferenciados con fluorescencia verde (Fig. II-10C). Una magnificación de la sección de la zona III de fijación de nitrógeno mostró signos evidentes de expresión de la leg-hemoglobina en la mayoría de las células vegetales, que aparecen homogéneamente invadidas por bacteroides, que se observan como pequeñas vesículas (Fig. II-10D). La proporción minoritaria (5-10%) de nódulos blancos inducidos por la cepa silvestre (probablemente de nueva aparición y en desarrollo) mostraron una histología similar incipiente (Fig. II-10I-J).

Una gran proporción (~65%) de nódulos inducidos por la cepa 1021 $\Delta$ *hfq* a los 30 días de inoculadas las plantas fueron blancos y menos elongados que los silvestres, revelando deficiencias simbióticas (típico fenotipo Fix<sup>-</sup>, Fig. II-10E). El resto de nódulos se observaron rosados y exhibieron una histología comparable a la de los silvestres (Fig. II-10I-J, K-L). La observación bajo microscopía de campo claro de secciones longitudinales de los nódulos Fix<sup>-</sup> reveló que los tejidos ocupados por bacteroides se restringían a la interzona II-III, donde incluso la autofluorescencia propia de éstos era mucho menor que la de esta zona en los nódulos salvajes, cuando se observaron a 520 nm (Fig. II-10F-G). La zona situada por debajo, hasta la base del nódulo, no mostró signos típicos de la zona III de fijación pareciéndose más a los

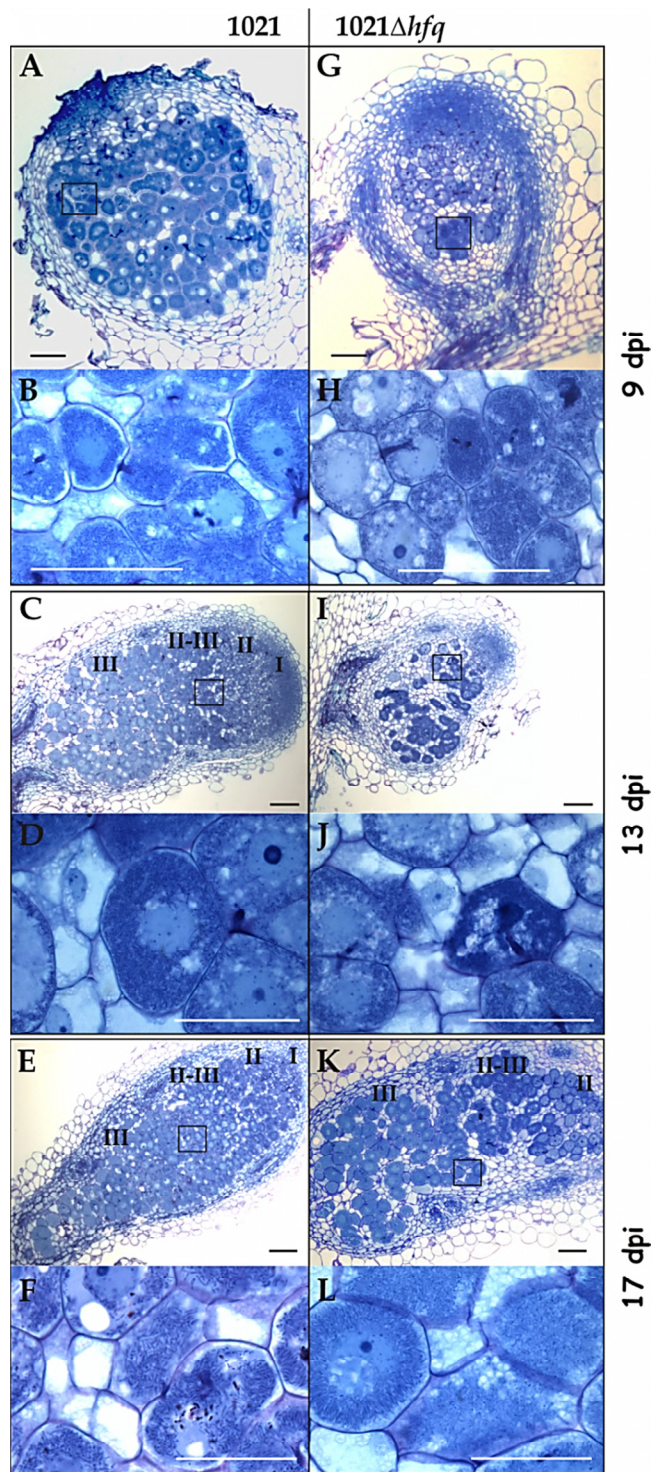


**Fig. II-10. Fenotipo endosimbiótico del mutante *S. meliloti* 1021 $\Delta$ *hfq*.** Imágenes ampliadas representativas de nódulos inducidos en plantas de alfalfa por la cepa 1021 (A) y por el mutante 1021 $\Delta$ *hfq* (E). Microscopía de campo claro de secciones longitudinales de los mismos nódulos (B y F); las zonas características de la histología de nódulos fijadores de nitrógeno indeterminados se indica en (B). Imágenes combinadas de las mismas secciones nodulares observadas con filtros para el verde y el azul (520 nm y 470 nm, respectivamente) (C y G). Magnificación de las imágenes de los tejidos centrales del nódulo (D y H); nódulos inducidos por 1021 $\Delta$ *hfq* son escasamente invadidos por bacterias y muestran signos de senescencia prematura: degradación de leg-hemoglobina (flechas simples) y restos celulares (dobles flechas). Los paneles inferiores (I-J, K-L) muestran la fracción minoritaria de nódulos ocupados por cada cepa, blancos en el caso de 1021 y rosáceos en el caso del mutante 1021 $\Delta$ *hfq*. Barra de escala: 250  $\mu$ m.

tejidos de senescencia (zona IV) de nódulos silvestres. El detalle ampliado de esta zona (Fig. II-10H), evidenció efectivamente reminiscencias histológicas de la zona IV, con mayoría de células vegetales libres de bacterias diferenciadas y evidencias de lisis bacteriana, como se infiere de la visualización de restos celulares (Timmers *et al.*, 2000). Las escasas células que albergan bacteroides no aparecen rosas como en los nódulos salvajes, sino oscuras, probablemente como consecuencia de la degradación de leg-hemoglobina concomitante a la muerte bacteriana. Esta histología confirma que las bacterias de la cepa mutante *S. meliloti* 1021 $\Delta$ *hfq* mantienen cierta capacidad

para infectar el huésped y diferenciarse a bacteroides, aunque la supervivencia de éstos en las células vegetales parece comprometida, resultando en una senescencia prematura de los nódulos y, en consecuencia, en el fenotipo mayoritario Fix observado en los mismos. Como complemento a estos experimentos se realizó un seguimiento de la organogénesis nodular en plantas de alfalfa inoculadas independientemente con la cepa silvestre y el mutante *hfq*. Para ello se recolectaron un total de 6 nódulos inducidos por cada cepa 3, 13 y 17 dpi que se fijaron, se incluyeron en una resina de alta resolución (Technovit), y se visualizaron al microscopio bajo campo claro en secciones de 4  $\mu\text{m}$  teñidas con azul de toluidina. En la Fig. X se muestran secciones representativas de ambos tipos de nódulos a cada tiempo de análisis. En nódulos inducidos por la cepa salvaje 1021, a los 9 dpi se observa que el tejido nodular se encuentra mayoritariamente ocupado por células vegetales que contienen bacteroides diferenciados en su interior (Fig. II-11A y B). A los 13 dpi estos nódulos se ven claramente elongados siendo ya evidente en los mismos la zonación típica de nódulos indeterminados, con casi la totalidad de células vegetales de la zona III de fijación de nitrógeno ocupadas por bacteroides (Fig. II-11C y D), situación que se consolida 17 dpi (Fig. II-11E y F). Sin embargo, en nódulos inducidos por el mutante 1021 $\Delta hfq$ , al noveno día de desarrollo se observa una mayor proporción de células vegetales vacías que en los silvestres, aunque en las ocupadas es también evidente la presencia de bacteroides (Fig. II-11G y H). A los 13 dpi (Fig. II-11I y J) estos nódulos no son tan elongados como los silvestres ni muestran una zonación definida, manteniendo también una baja proporción de células vegetales invadidas por bacteroides, que no obstante aumenta en nódulos observados 17 dpi, que recuerdan a los nódulos silvestres de 13 días de desarrollo (Fig. II-11K y L).

Estos resultados confirman la capacidad infectiva y de diferenciación morfológica del mutante *S. meliloti*  $\Delta hfq$  aunque ponen de manifiesto un retraso evidente en la organogénesis de los nódulos inducidos por éste respecto a la de los silvestres, que probablemente en la mayoría de los casos no debe llegar a término como consecuencia de su senescencia temprana.



**Fig. II-11. Organogénesis inducida por *S. meliloti* 1021 y 1021 $\Delta$ hfq.** Imágenes representativas de secciones longitudinales de 4  $\mu$ m de grosor, teñidas con azul de toluidina, de nódulos fijadores inducidos en plantas de alfalfa por la cepa 1021 (A-F) y no fijadores por el mutante 1021 $\Delta$ hfq (E-L), a 9 dpi, 13 dpi y 17 dpi. Detalle del nódulo completo (A, C y E para 1021 y G, I y K para el mutante) y magnificación de la zona encuadrada (B, D y F para 1021 y H, J y L para el mutante). En todos los casos, los bacteroides se visualizan como células alargadas fuertemente teñidas situadas en el interior de las células vegetales. La zonación típica de histologías de nódulos indeterminados se indica en los casos en los que es perceptible: I, zona invasión; II, zona de diferenciación; II-III, interzona de diferenciación-fijación; III, zona de fijación. Barra de escala: 50  $\mu$ m.

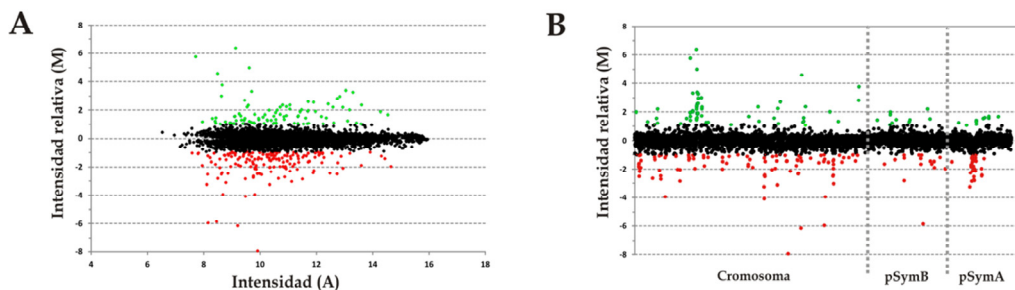
### II-3.2.3. ANÁLISIS TRANSCRIPTÓMICO/PROTEÓMICO DE LOS MUTANTES

#### hfq

Los cambios globales en la abundancia de transcritos dependientes de Hfq se investigaron por comparación de los perfiles transcriptómicos sobre *microarrays* Sm14kOLI (<http://www.cebitec.uni-bielefeld.de/transcriptomics/transcriptomics-facility/sm14koli.html>) de la cepa silvestre *S. meliloti* 1021 y su derivada 1021 $\Delta$ hfq crecidas en medio completo TY hasta fase logarítmica (D.O 0.5-0.6). Se realizaron tres réplicas biológicas (cultivos y obtención de RNA) y dos técnicas (marcaje del cDNA e hibridación) del experimento para obtener un total de 12 datos posibles de hibridación sobre cada sonda del *microarray* (cada oligonucleótido está duplicado en los *microarrays* Sm14kOLI). La normalización y el tratamiento estadístico de los datos se llevaron a cabo con el programa EMMA versión 2.8.2. Una primera cuantificación de la fluorescencia sobre cada sonda reveló señales de hibridación detectables en 10.494 (35%) de los 29.956 oligonucleótidos impresos sobre estos *microarrays*. El primer test de significación ( $P \leq 0,05$ ) reveló que 4.858 de las señales detectadas correspondían a genes (o *loci* no anotados en las RIGs) diferencialmente expresados (valores de intensidad relativa  $M \neq 0$ ), de los cuales 2.347 estaban reprimidos ( $M < 0$ ) y 2.545 aparecían inducidos ( $M > 0$ ) en el mutante *S. meliloti* 1021 $\Delta$ hfq. Esta lista inicial fue finalmente filtrada eliminando las RIGs y considerando para su análisis posterior los genes (ORFs) cuyas señales de hibridación ( $P \leq 0,05$ ) fueron de intensidad  $A \geq 8$  e intensidad relativa (ratio de expresión)  $M \geq 1$  o  $\leq -1$  (diferencia de al menos 2 veces en los niveles de mRNA), valores ambos calculados a partir de al menos 8 datos de hibridación por sonda ( $n \geq 8$ ) del total de 12 posibles.

La aplicación de estos criterios redujo la lista de genes diferencialmente expresados a 168 (2,7% de los 6206 ORFs anotados en el genoma de *S. meliloti* 1021), de los cuales 91 fueron catalogados como reprimidos y 77 como inducidos en *S. meliloti* 1021 $\Delta$ hfq (Fig. I-12A). Los resultados de este análisis se depositaron en la base de datos ArrayExpress, bajo el número de acceso A-MEXP-1760.

La distribución por replicones de los 168 genes dependientes de Hfq reveló que 103 (61%) son de localización cromosómica y 65 se encuentran en los megaplásmidos; 45 (27%) en pSymA y 20 (12%) en pSymB (Fig. I-12B). Considerando como referencia el contenido en ORFs de cada replicón en *S. meliloti* 1021, con el 54% de ellos anotados en el cromosoma, el 21% en pSymA y el 25% restante en pSymB, los datos transcriptómicos revelan un ligero sesgo del impacto de la actividad de Hfq sobre funciones codificadas en el cromosoma y en el plásmido simbiótico pSymA. Este sesgo es particularmente evidente hacia pSymA cuando se consideran sólo los genes reprimidos (91) en el mutante *S. meliloti* 1021 $\Delta$ hfq34 de los cuales (37%) se localizan en



**Fig. II-12. Análisis transcriptómico de *S. meliloti* 1021 $\Delta$ *hfq* sobre microarrays Sm14kOLI. A)** Gráfico M/A correspondiente a las señales estadísticamente significativas incluyendo regiones intergénicas (RIGs) y secuencias codificantes (ORFs). **B)** Representación gráfica del valor M de las mismas señales mostrando la distribución en los tres replicones de *S. meliloti*. En ambos casos, los valores destacados en rojo se corresponden a genes reprimidos en el mutante, mientras que los valores en verde corresponden a los inducidos.

este plásmido simbiótico, casi doblando el porcentaje esperado por su contenido en ORFs.

En base a las anotaciones de la secuencia genómica de *S. meliloti* 1021 (<http://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgiand>) y a la base de datos KEGG (<http://www.genome.jp/kegg/>), 137 (82%) de los 168 genes con expresión alterada en el mutante *hfq* pudieron asignarse a alguna categoría funcional, mientras que los 31 restantes (18%) mostraron homología global o parcial con secuencias de las bases de datos correspondientes a posibles genes de función desconocida (proteínas hipotéticas) (Fig. II-13) (el listado completo de los genes diferencialmente expresados en el mutante *hfq* y sus categorías funcionales se pueden consultar en la Tabla II-1 en el apéndice de este capítulo). Más de la mitad de los genes con homología funcional codifican proteínas y enzimas posiblemente relacionadas con procesos de transporte (45 genes) y metabolismo (39) de pequeñas moléculas, principalmente de azúcares y aminoácidos, mientras que los 53 restantes están relacionados con la regulación de la transcripción (5 genes), el transporte de electrones o la fijación de nitrógeno (17; incluyendo los genes reguladores *fixK1/fixK2*) y otros procesos celulares (31) como el metabolismo central intermediario o el de macromoléculas. Entre estos últimos se identificaron varios genes con posible función en la traducción y reciclaje de RNA o proteínas como *SMc01929* que codifica la RNasa J, *SMc03796* que codifica una posible endorribonucleasa L-PSP implicada en la maduración de mRNAs *SMA1126*, *degP4* y *degP1* que aparecen anotados como determinantes genéticos de diferentes tipos de proteasas, y *rplS/spmA*, ambos codificantes de proteínas ribosómicas. Todos estos genes salvo *SMA1126* y *degP4*, se encontraron inducidos en el mutante.

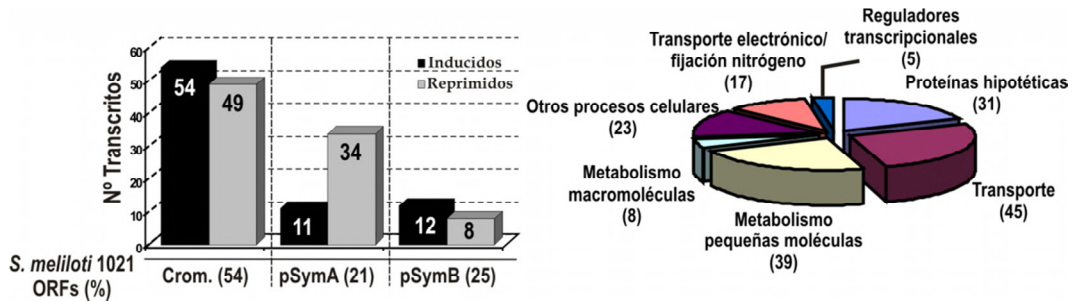


Fig. II-13. Categorías funcionales de los genes diferencialmente expresados en el mutante *S. meliloti*  $\Delta hfq$ . El histograma muestra el número de genes diferencialmente expresados y su distribución en los tres replicones de *S. meliloti*, cromosoma (Crom.), pSymA y pSymB. La distribución de los ORFs anotados en el genoma se indica como referencia. La adscripción de estos genes en categoría funcionales según la base de datos KEGG y *S. meliloti* se muestra a la derecha en diagramas de sectores. El número de genes de cada categoría se muestra entre paréntesis.

### II-3.2.3.1. Efecto de Hfq en el transporte y metabolismo en *S. meliloti*

Entre los 91 genes reprimidos en el mutante *hfq*, 40 (44%) codifican funciones relacionadas con el transporte (14 genes) y el metabolismo de pequeñas moléculas (26 genes), la mayoría de ellas sustratos carbonados (Fig. II-13). Once de estos genes forman parte de operones que codifican los componentes estructurales (i.e. la proteína periplásmica de unión a solutos, la permeasa o la proteína de unión a ATP) de los transportadores ABC de mio-inositol (genes *ibpA*, *iatA* e *iatP*),  $\alpha$ -glucósidos (*aglE* y *aglF*), fructosa (*frcB* y *frkK*), ribosa (*SMc02031*), glicerol (*SMc02514* y *SMc02519*), y otros ácidos/alcoholes orgánicos (*SMB20144*) (Mauchline *et al.*, 2006). Los niveles del transcrito de otro gen, *SMB20072*, que codifica una proteína periplásmica de unión a soluto inducida por mio-inositol (Mauchline *et al.*, 2006) fueron particularmente inferiores en el mutante (más de 32 veces respecto a la cepa silvestre;  $M = -5,87$ ). Sin embargo, este gen parece formar una unidad transcripcional monocistrónica, aparentemente no agrupada con otros genes de captación y utilización de azúcares. Los dos genes restantes que codifican transportadores y se encuentran reprimidos en el mutante estarían implicados en el metabolismo de glicina-betaína (*SMc04439*) y de hierro (*SMc04317*). La eficiencia reducida para la captación de fuentes primarias de carbono en el mutante *S. meliloti* 1021 $\Delta hfq$  está acompañada por la represión de 8 genes relacionados con el catabolismo de azúcares: *iolC*, *iolD*, *iolE* e *iolB* que integran el operón de la utilización del mio-inositol, *SMc01163* que codifica una posible oxidorreductasa de glucosa-fructosa, *SMc00982* correspondiente a una dioxigenasa, y dos genes parálogos que codifican una alcohol deshidrogenasa, *adhA1* y *adhA2*, probablemente implicada en la fermentación de sustratos carbonados. La pérdida de Hfq también se resultó en la disminución de la abundancia del transcrito *SMA1227*, que codifica un regulador transcripcional de la

superfamilia Crp, que contribuye al control de rutas del metabolismo central del carbono en bacterias a través de la unión a cAMP (Gorke and Stulke, 2008). Además de la represión de genes de rutas relacionadas con la producción de energía, algunos transcritos que codifican componentes de la cadena de transporte de electrones como CycA, EtfA1 o Sma1170 (posible citocromo c) fueron también menos abundantes en el mutante. Otro conjunto de genes reprimidos en *S. meliloti* 1021 $\Delta$ hfq incluye algunos implicados en procesos celulares sustentados energéticamente por el catabolismo de azúcares como los de la biosíntesis de aminoácidos (*ilvC*, *SMc03211*, *SMc03252*, *SMc03253*, *pheAa*, *mtbC*, *SMc02045* y *glyA1*), vitaminas (*cobP*, *SMc04342*) y purinas/pirimidinas (*purU1*, *pyrC*).

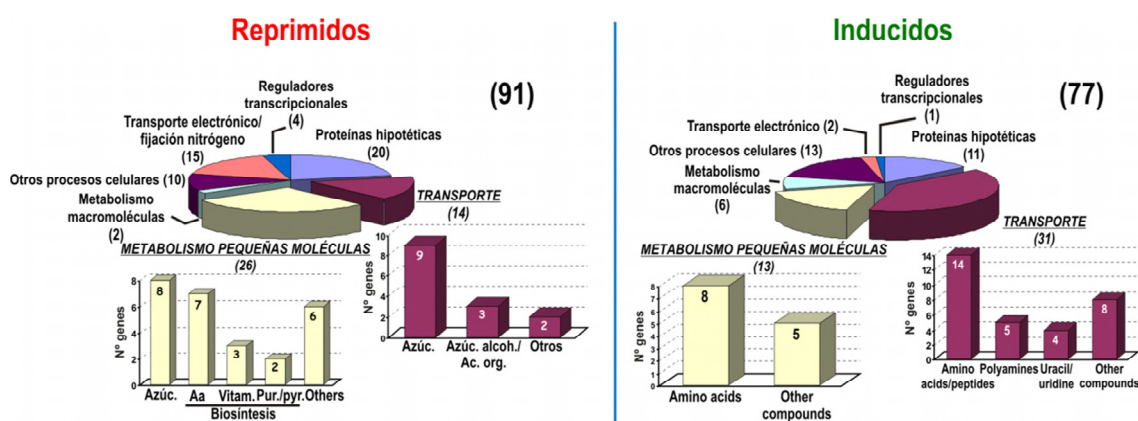


Fig. II-14. Influencia de Hfq en el transporte y metabolismo de pequeñas moléculas en *S. meliloti*. Distribución funcional de los transcritos inducidos y reprimidos en el mutante 1021 $\Delta$ hfq. Entre paréntesis se muestra el número de genes pertenecientes a cada categoría. Los histogramas detallan la subdivisión de los genes metabólicos y de transporte.

Más del 50% de los transcritos más abundantes en el mutante *hfq* que en su cepa parental codifican proteínas para el transporte y catabolismo de diferentes compuestos nitrogenados. Entre ellos están los implicados en la captación de L-aminoácidos y pequeños péptidos (*aapJ*, *aapQ*, *aapP*, *oppB*, *oppC*, *SMc00140*, *SMc01597*, *SMc02259*, *Smb21572*, *Smb20605*), de aminoácidos de cadena ramificada (*livH*, *livM*, *livG*, *livF*, *livK*), de uracil/uridinas (*SMc01823*, *SMc01824*, *SMc01825*, *SMc01827*), de aminas derivadas de azúcares (*Smb21151*) o de otros sustratos nitrogenados complejos como poliaminas, espermidina y putrescina (*SMc01966*, *SMc01965*, *SMc01963*). Consecuentemente, la pérdida de Hfq también resulta en la inducción de algunos genes relacionados con la utilización o modificación de aminoácidos y otros compuestos nitrogenados. Los transcritos correspondientes a tres genes específicos del sistema de ensamblaje de la glicina, *gcvP*, *gcvH* y *gcvT* (valores de M 2,06, 2,02 y 3,32, respectivamente), así como al gen *SMc01930* (M 3,26) que codifica una

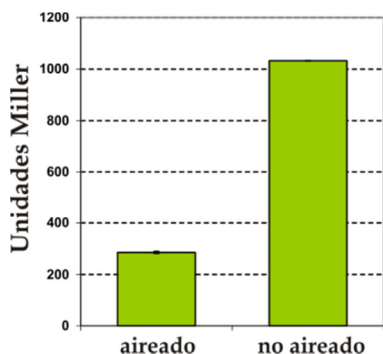


epimerasa de metilmalonil-CoA que opera en el catabolismo de aminoácidos de cadena ramificada, aparecieron particularmente sobrerrepresentados en el mutante.

#### **II-3.2.3.2. Hfq contribuye a la regulación de *fixK1/fixK2***

Un conjunto importante de transcritos codificados en pSymA cuya acumulación disminuye en el mutante *S. meliloti* 1021 $\Delta$ *hfq* incluye los correspondientes a las dos copias del regulador *fixK* (*fixK1* y *fixK2*) y a los genes dependientes de éste, que en *S. meliloti* codifican los componentes de la cadena respiratoria asociada al complejo nitrogenasa (*fixN1*, *fixQ1*, *fixP1*, *fixG*, *fixQ2*, y las dos copias de *fixM*). Este hallazgo fue sorprendente ya que está bien documentado que la transcripción de los genes de fijación de nitrógeno (*fix/nif*) están regulados por oxígeno en los nódulos de leguminosas y sólo se inducen bajo condiciones de microaerofilia en bacterias en vida libre (Dixon and Kahn, 2004). Sin embargo, también se ha descrito que una disminución moderada de la concentración de oxígeno (hasta un 5%) en la fase gaseosa de un cultivo es suficiente para desencadenar la autofosforilación dependiente de ATP de la hemoproteína FixL en la cascada FixLJ-FixK de transducción de señal (Gong *et al.*, 1998). En *S. meliloti* la forma fosforilada de FixJ, no sólo activa la transcripción de los genes reguladores *fixK1/fixK2*, sino también la de *nifA*, el activador transcripcional de los genes estructurales del complejo de la nitrogenasa. La expresión de *nifA* exige, sin embargo, condiciones de microaerobiosis más restrictivas (Dixon and Kahn, 2004). Por tanto, la represión de los genes *fix* en el mutante *hfq* sólo se podría explicar si en nuestras condiciones de cultivo (tubos de ensayo de 15 mL con 10 ml de medio) hay transcripción de *fixK1/fixK2* en la cepa silvestre *S. meliloti* 1021. Para testar esta hipótesis se determinó la actividad  $\beta$ -galactosidasa de la cepa *S. meliloti* pMF457 (Foussard *et al.*, 1997), que expresa una fusión transcripcional del promotor de *fixK* con el gen reportero *lacZ*, en las condiciones de cultivo habituales para la transcriptómica y en otros cultivos más aireados (idénticos tubos de 15 ml con 3 ml de medio). Los resultados obtenidos (Fig. II-15) revelaron valores de actividad  $\beta$ -galactosidasa 4 veces superiores en los cultivos menos aireados, que confirman la expresión de *fixK1/K2* en las condiciones en que fueron investigados los perfiles transcriptómicos de la cepa silvestre *S. meliloti* 1021 y el mutante *hfq*.

Experimentos de RT-PCR independientes llevados a cabo con posterioridad en nuestro laboratorio confirmaron que, efectivamente Hfq contribuye a la regulación post-transcripcional de *fixK1/fixK2*, e incluso de *nifA*, en condiciones microaerófilas.

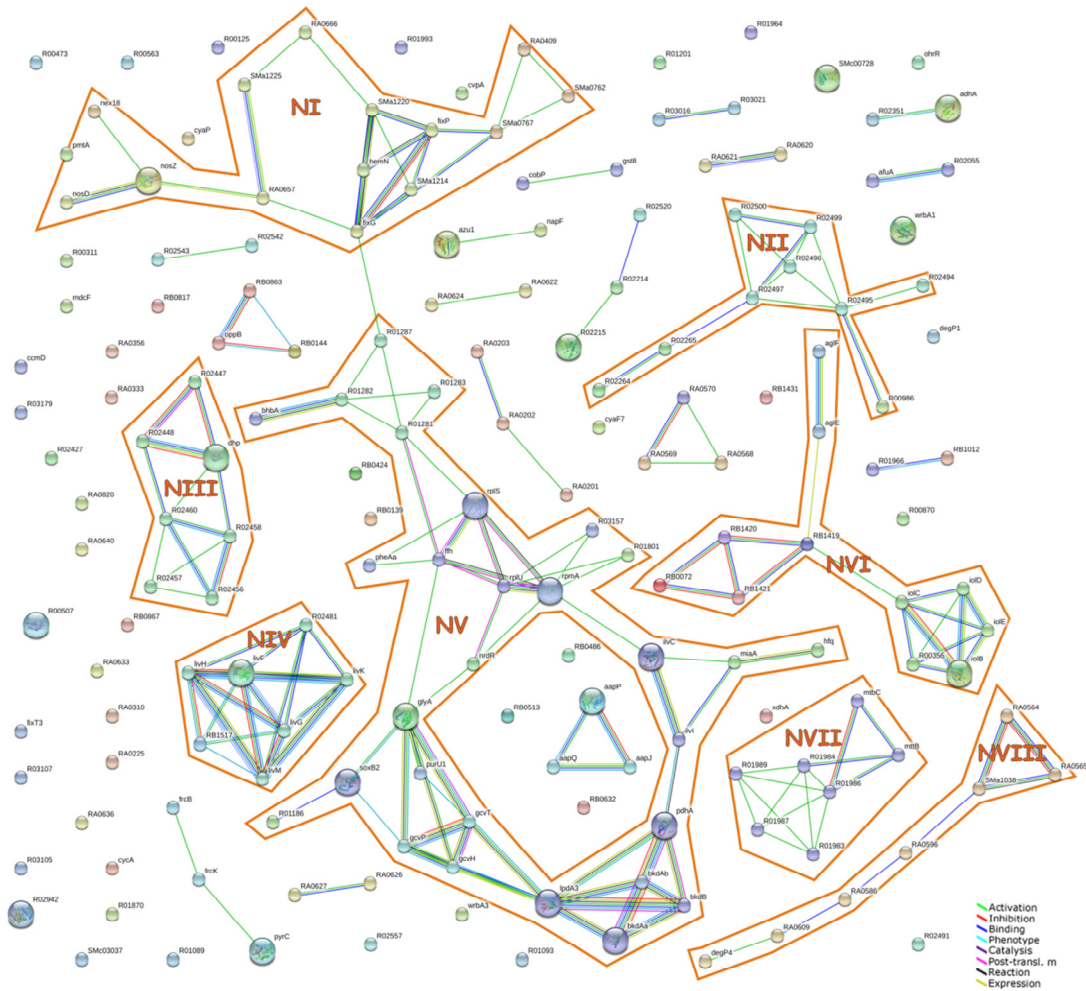


**Fig. II-15.** Actividad  $\beta$ -galactosidasa en fusiones *fixK::lacZ*. Unidades Miller resultantes de la actividad  $\beta$ -galactosidasa de la construcción pMF457 en las condiciones de cultivo aireado (3 ml) y no aireado (10 ml), en tubos de ensayo de 15 ml.

### II-3.2.3.3. Redes génicas afectadas por la mutación en *hfq*

Los 168 genes con expresión alterada en el mutante *Hfq* fueron finalmente agrupados en redes génicas mediante la herramienta STRING (<http://string-db.org/>). Esta aplicación utiliza la información contenida en las bases de datos sobre los genes de interés y/o sus homólogos para inferir posibles relaciones funcionales entre los mismos basadas en su co-regulación en determinados procesos celulares demostrada experimentalmente, en la interacción predicha o verificada experimentalmente de sus productos proteicos, en la sintenia de sus respectivas regiones genómicas o cualquier otra observación *in silico*.

En nuestro caso STRING identificó ocho redes génicas principales que involucran unos 90 de los 168 genes analizados y que se representan en la Fig. II-16. La primera (NI) comprende los genes *fix*, tanto los reguladores *fixK1/fixK2*, como todos los que componen la cadena electrónica asociada a la nitrogenasa. La segunda (NII) corresponde al sistema de utilización de putrescina/espermidina. La red NIII contiene los genes implicados en el metabolismo de las pirimidinas. NIV comprende genes relacionados con el metabolismo de leucinas (genes *liv*). La quinta red (NV) es la más heterogénea, ya que aparece integrada por genes relacionados con la síntesis de glicinas (*gcvT*, *gcvH* y *gcvP*), o con la captación de isovalerato (genes *bkp*), que intervienen en la síntesis de aminoácidos y genes relacionados con síntesis de subunidades ribosomales (*rplS*, *rplN*, *rpmA*), las redes NVI y NVII comprenden genes relacionados con la captación de azúcares como el mio-inositol y de aminas (metilaminas), respectivamente. Finalmente, la última red NVIII contiene un conjunto de proteínas hipotéticas que aparecen relacionadas en las bases de datos y que podrían actuar como enzimas oxidasas del cobre.



**Fig. II-16. Agrupamiento en redes genéticas de los transcritos diferencialmente expresados en el mutante *hfq*.** Los puntos observados representan cada uno de los genes diferencialmente regulados, en nomenclatura STRING (R). Las redes genéticas (señaladas en naranja) corresponden a un índice de confianza de 0,4 (probabilidad de interacción). NI, genes *fix* relacionados con la cadena de transporte electrónico asociada a la nitrogenasa; NII, sistema espermidina/putrescina; NIII, sistema de transporte y utilización de pirimidinas; NIV, sistema de transporte de aminoácidos de cadena ramificada; NV, red de genes relacionados con el metabolismo de aminoácidos; NVI, genes relacionados con el sistema de captación de azúcares; NVII, red relacionada con la captación de metilamina y tetrahidrofolato; NVIII, red de proteínas hipotéticas relacionadas con oxidasas del cobre. El tipo de interacción entre los genes integrantes de cada red se indican en la leyenda situada en la esquina derecha.

#### II-3.2.3.4. Análisis proteómico del mutante *S. meliloti* 2011-3.4

Como aproximación independiente para investigar la función de Hfq en *S. meliloti*, se determinó el perfil proteómico de la cepa salvaje *S. meliloti* 2011 y su mutante 2011-3.4, obtenido por inserción del plásmido pK18 mediante simple recombinación homóloga en el gen *hfq*, afectando la traducción del dominio Sm2 de la proteína. La generación del mutante y el estudio proteómico del mismo se llevaron

a cabo por Roke I. Oruezabal y Edgardo Jofré en el marco de una colaboración con el grupo del Dr. Rafael Rivilla (Universidad Autónoma de Madrid). Los resultados del mismo se describen brevemente a continuación.

El análisis de geles bidimensionales teñidos con Coomassie de bacterias crecidas en medio TY hasta fase logarítmica (D.O<sub>600</sub> 0.5-0.8) reveló una media de 293 spots de los cuales 33 corresponden a polipéptidos individuales con una acumulación diferencial en la cepa salvaje y en el mutante (Tabla II-2, apéndice). La espectrofotometría de masas (MALDI-TOF) reveló que 28 de esas proteínas están codificadas en genes de localización cromosómica, 4 en pSymB y sólo 1 en pSymA, confirmando que el principal papel de Hfq es sobre genes cromosómicos. De estas 33 proteínas, 21 están inducidas y 12 reprimidas en el mutante 2011-3.4. La base de datos KEGG asignó las proteínas diferencialmente expresadas a tres categorías funcionales: transporte (12 proteínas), metabolismo de pequeñas moléculas (8) y chaperonas y/o factores de estrés (4), mientras que las 9 restantes se catalogaron como implicadas en la traducción (i.e. factor Tig y factor P de elongación Efp) o proteínas hipotéticas con función desconocida (7).

La comparación de los perfiles transcriptómicos y proteómicos descritos en este trabajo reveló que existen 9 genes identificados como diferencialmente regulados en los mutantes *hfq* que comparten ambas aproximaciones experimentales. Estos genes codifican proteínas que forman parte de los componentes periplásmicos de transportadores ABC de mio-inositol (IbpA), fructosa (FrcB),  $\alpha$ -glucósidos (AglE), aminoácidos (SMc02259), leucina (LivK) y L-aminoácidos (AapJ y AapP), además de dos enzimas relacionadas con el catabolismo del mio-inositol, IolE e IolD.

### II-3.3. CARACTERIZACIÓN DE LA POBLACIÓN DE mRNAs QUE SE UNEN A LA PROTEÍNA *Hfq* EN *S. MELILOTI*

El análisis correspondiente al CoIP-RNA obtenido de la cepa de *S. meliloti* 2011*hfq*<sup>FLAG</sup> en fase logarítmica de crecimiento, reveló que existen 1.178 mRNAs que se encuentran enriquecidos en un orden de 5 o más respecto de la cepa control 2011. Estos mRNAs abarcan la multitud de procesos celulares, tales como la traducción y degradación de mRNAs, metabolismo general y transporte, respiración, etc.

El análisis conjunto de los perfiles transcriptómicos y proteómicos reveló el solapamiento de 80 genes con la fracción de mRNAs unidos a Hfq y detectados por RNA-Seq. Las categorías de estos 80 genes correspondieron a 18 transportadores ABC (fundamentalmente implicados en el intercambio de aminoácidos y péptidos), 4 genes correspondientes a chaperonas (genes *groES* y la propia Hfq), 20 enzimas

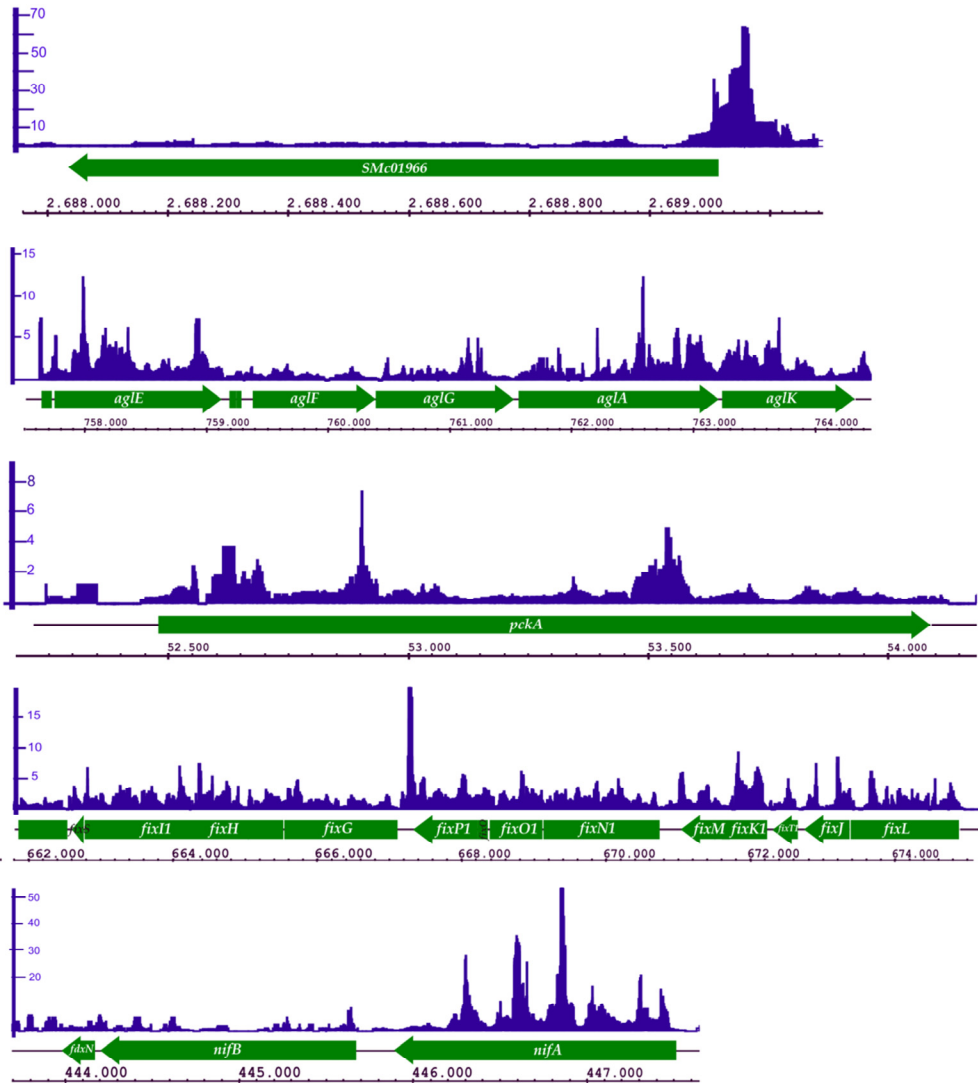


Fig. II-17. Patrones de distribución en la cobertura del enriquecimiento de mRNAs asociados con Hfq. Diferentes mRNAs coincidentes en experimentos de transcriptómica/proteómica y RNA-Seq, en donde la secuencia codificante (OFR) y la dirección de la traducción se representan como una flecha verde. El eje vertical en cada mRNA indica el enriquecimiento de la cepa *hfq*<sup>FLAG</sup> sobre el control (cepa silvestre), en cada nucleótido. La escala correspondiente a la coordenada genómica se indica en la parte inferior de cada panel.

metabólicas, 5 genes relacionados con la fijación de nitrógeno, 17 proteínas hipotéticas y 16 que abarcan otros procesos celulares. Cuando se compararon las listas correspondientes a los genes diferencialmente expresados en *1021Δhfq*, obtenidos en experimentos de *microarrays*, y los mRNAs asociados con Hfq en los experimentos de RNA-seq, se vio que 65 *loci* eran comunes a ambas listas (ver Tabla I-1 apéndice). Entre ellos se pueden destacar los genes implicados en el transporte de aminoácidos (e.g. *livK*, *livH*; *appP*, *appQ* y *appJ*; *dppA1*, *dppA2*), el transportador de  $\alpha$ -

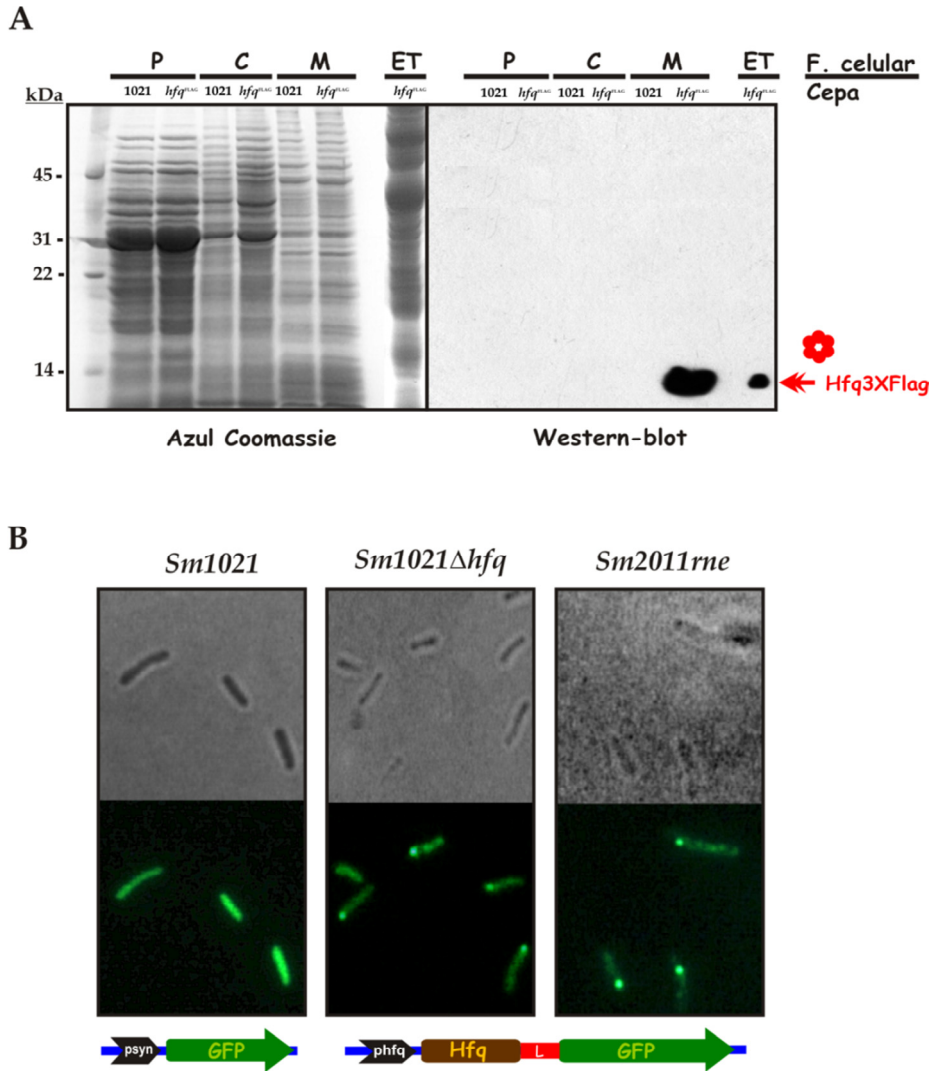
glucósidos (*aglE*), el propio transcrito de Hfq y varias enzimas metabólicas (e.g. la hidrolasa de prolina *SMc03253*, la tetrahidrofolato reductasa *SMc03205* y la alcohol deshidrogenasa *adhA*). El análisis de las regiones en las que estos genes se encuentran implicados reveló que, en muchos casos, la cobertura del enriquecimiento observado en el CoIP-RNA se correspondía con la región 5' del mRNA, sugiriendo una posible regulación por parte de sRNAs (e.g. *SMc01966*) (Fig. II-17, primer panel). En el caso de mRNAs policistrónicos, la cobertura se extendió en muchos casos a la totalidad de los elementos que conforman el sistema multigénico, e.g. el operón *agl* (Fig. II-17, segundo panel), en donde se observó la asociación con Hfq de todos los genes que conforman el transportador. La comparación de los resultados obtenidos en los perfiles proteómicos con el análisis del RNA asociado a Hfq únicamente mostró coincidencia en 5 genes correspondientes a los transportadores de aminoácidos *livK*, *appPJ*, *SMc02259* y el transportador de glucósidos *aglE*. Además, se detectó la presencia del transcrito completo del regulador de la gluconeogénesis *pckA* (Fig. II-17, tercer panel), aunque su enriquecimiento medio no fue superior a 4.

Es destacable la presencia en estas muestras de los reguladores principales de la fijación de nitrógeno *fixK* y *nifA* (Fig. II-17, cuarto y quinto panel). El caso de *fixK* es muy llamativo ya que se asocian con Hfq, además de los reguladores *fixK1/K2*, todos los integrantes de los genes que conforman la cadena de transporte electrónico asociada a la nitrogenasa, relacionándose con los experimentos de transcriptómica, en donde se encontraban reprimidos. El único gen correspondiente al sistema estructural de la nitrogenasa que mostró una asociación con Hfq fue *nifD*, que mostró un enriquecimiento medio de 6 repartido por toda la longitud del transcrito. Todos estos resultados explicarían las bases moleculares del fenotipo simbiótico observado en el mutante *1021Δhfq*.

#### II-3.4. LOCALIZACIÓN DE LA PROTEÍNA Hfq EN *S. MELILOTI*

Como primer abordaje para la determinación de la localización de Hfq en células de *S. meliloti*, se analizó la presencia de la proteína recombinante Hfq<sup>FLAG</sup> en distintas fracciones celulares de la cepa *S. meliloti* *1021hfq<sup>FLAG</sup>*, derivada de la de referencia 1021. Esta cepa expresa también una proteína Hfq marcada en su extremo C-terminal por tres unidades del epítipo FLAG y fue obtenida de forma idéntica a *S. meliloti* *2011hfq<sup>FLAG</sup>* según el protocolo descrito con detalle en el Capítulo 1.

Las cepas *S. meliloti* 1021 y *S. meliloti* 1021 *hfq<sup>FLAG</sup>* se cultivaron independientemente en medio TY hasta una D.O. (600 nm) de 0.6. Los respectivos



**Fig. II-18. Localización sub-celular de Hfq en *S. meliloti* 1021.** A) SDS-PAGE de las distintas fracciones proteicas; membrana, citoplasma y periplasma teñido con Coomassie (izquierda) y membrana PVCF del mismo gel sometida a análisis por Western-blot utilizando anticuerpos anti-Flag (derecha). De cada fracción subcelular (P, periplasma; C, citoplasma; M, membrana; ET, extracto total) se cargó una muestra de la cepa salvaje 1021 (wt) y otra de la cepa problema 1021 *hfq*<sup>FLAG</sup> (*hfq*<sup>FLAG</sup>), salvo en el extracto total, que se cargó únicamente una muestra procedente de la cepa problema, como control. B) Células de *S. meliloti* visualizadas a microscopía de campo claro (arriba) y fluorescencia (abajo). Las cepas visualizadas se indican en cada panel y el plásmido que portan se muestra abajo.

lisados celulares se sometieron a un protocolo de fraccionamiento, obteniéndose finalmente para cada cepa tres fracciones proteicas correspondientes a periplasma, citoplasma y membrana que se resolvieron en un gel SDS-PAGE al 12 % para someterse a Western-blot con anticuerpos anti-FLAG. Los resultados muestran que se detectó únicamente la banda correspondiente a la proteína Hfq<sup>FLAG</sup> (~12 kDa) en la

fracción de membrana de la cepa modificada que, sin embargo, no es visible como era esperado en ninguna fracción celular de la cepa silvestre 1021 (Fig. II-18A).

Otra aproximación utilizada en este trabajo para localizar Hfq *in vivo*, fue la visualización de la proteína mediante fluorescencia en el interior de la célula bacteriana. Para ello (lo de para tal efecto no es una expresión adecuada) se utilizó la cepa *S. meliloti* 1021 $\Delta$ hfq (pJBHfq:GFP) que expresa la fusión traduccional Hfq:eGFP desde el promotor del gen *hfq*. La peculiaridad de esta fusión radica en que las dos proteínas no están estrechamente unidas, sino que entre ellas existe un péptido corto (10 aminoácidos) rico en serinas y glicinas que le confieren una gran flexibilidad, lo que permite a ambas mantener su independencia y, por tanto, su funcionalidad (Valencia-Burton *et al.*, 2007). De hecho, la proteína recombinante que expresa el plásmido pJBHfq:eGFP complementa el fenotipo de crecimiento observado en el mutante *S. meliloti* 1021 $\Delta$ hfq.

Cuando se visualizaron al microscopio de fluorescencia bacterias de la cepa silvestre *S. meliloti* 1021 que expresan constitutivamente eGFP desde el plásmido pBBSyn-eGFP se observó fluorescencia homogénea en toda la célula. Sin embargo, la fusión de Hfq a eGFP analizada en *S. meliloti* 1021 $\Delta$ hfq, relocalizó esta fluorescencia, concentrándola en uno o varios puntos brillantes (*foci*) situados en los polos de las células (Fig. II-18B). Esta localización sub-celular ha sido observada en proteínas relacionadas con procesos diversos como división celular o fenómenos asociados a membrana y/o nucleoide, por lo que la correlación de estos hallazgos con los de los Western-blots requeriría de más verificaciones experimentales.

### II-3.5.1. Hfq Y RNasa E

En *E. coli* la proteína Hfq interacciona, entre otros elementos celulares, con la ribonucleasa E (RNasa E), proteína que forma parte del complejo del degradosoma y que en esta bacteria localiza uniformemente en membrana (Ikeda *et al.*, 2011).

La comparación de las secuencias aminoacídicas de las RNasa E de *E. coli* y *S. meliloti* reveló una identidad discreta entre ambas del 51%. Los motivos conservados se restringen a la región N-terminal donde se localiza el centro catalítico de estos enzimas (Callaghan *et al.*, 2005) mientras que la región C-terminal, que en *E. coli* especifica la localización en membrana (Khemici *et al.*, 2008) y la asociación del enzima con Hfq y otros componentes del degradosoma (Ikeda *et al.*, 2011), no se encuentra apenas conservada en ambas proteínas. (Fig. II-19).

Para estudiar si, como ocurre en *E. coli*, RNasa E y Hfq co-localizan en células de *S. meliloti*, la fusión Hfq:GFP se movilizó al mutante *S. meliloti*



2011mTn5STM.4.07.G10 (colección de mutantes de *S. meliloti* de la Universidad de Bielefeld; Pobigaylo *et al.*, 2006). Esta bacteria, derivada de la cepa silvestre *S. meliloti* 2011, tiene una inserción del transposón mini-Tn5 en el gen que codifica RNasa E (SMc01336), afectando la traducción de la región C-terminal de esta proteína a partir del residuo 678 de su secuencia aminoacídica, que no compromete la viabilidad de las bacterias (<http://www.rhizogate.de>) (Fig. II-19). La visualización de estas células al microscopio reveló una distribución de la fluorescencia en focos de disposición polar similar a la observada en la cepa *S. meliloti* 1021Δ*Hfq* (pJBHfq:eGFP), (Fig. II-18B). Estos resultados sugieren que, en *S. meliloti*, Hfq no se asocia con la ribonucleasa E, o al menos no lo hace con la región no catalítica de ésta, anticipando que los extremos C-terminales no conservados podrían ser funcionalmente distintos en ambas proteínas.

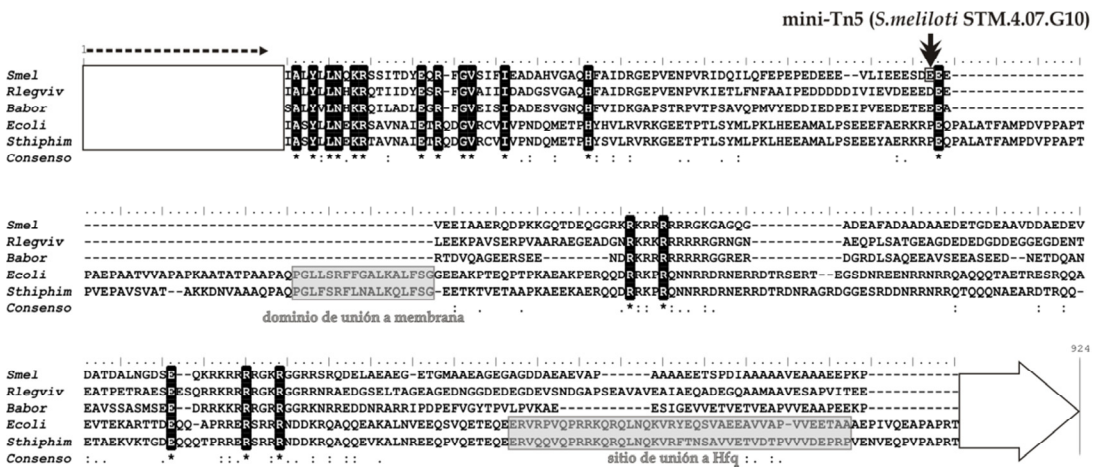


Fig. II-19. El gen *SMc01336* de *S. meliloti*. Alineamiento de la secuencia aminoacídica de las RNasa E de *S. meliloti* (Smel), *R. leguminosarum* bv. *viciae* (Rlegvic), *B. abortus* (Babor), *E. coli* (Ecoli) y *S. tiphymurium* (Stiphim). El alineamiento muestra la región que comprende la región C-terminal de las cinco RNasas E. El sitio de inserción del mini-Tn5 insertado en el mutante 2011mTn5STM.4.07.G10 se indica con doble flecha, además de los motivos de unión a membrana y a la proteína Hfq conservados en las RNasas E entebacterianas.

## II-4. DISCUSIÓN

La chaperona de RNA Hfq es un regulador post-transcripcional global que controla redes génicas que operan en el establecimiento de las interacciones entre bacterias patógenas con sus huéspedes eucariotas (Toledo-Arana *et al.*, 2007). Sin embargo, su papel en las interacciones beneficiosas microbio-huésped no ha sido investigado en detalle. En este capítulo se han presentado aproximaciones genéticas y genómicas a la función de Hfq en *S. meliloti* como bacteria de vida libre y durante la interacción simbiótica con su leguminosa huésped, alfalfa. Como se resume en el

modelo mostrado en la Fig. I-20, nuestros resultados sugieren la implicación de Hfq en rutas bacterianas que afectan al metabolismo central, a la competencia rizosférica, a la supervivencia en el interior de las células vegetales y a la fijación simbiótica de nitrógeno. Con anterioridad a este trabajo se publicaron algunos estudios que describen la contribución de Hfq a la regulación de los genes de la fijación de nitrógeno en varios diazotrofos del grupo de las  $\alpha$ -proteobacterias (Kaminski *et al.*, 1994; Kaminski and Elmerich, 1998; Drepper *et al.*, 2002; Zhang and Hong, 2009). Por tanto, nuestros resultados revelan una función más amplia de Hfq en bacterias fijadoras de nitrógeno que afecta a otros procesos relacionados con la interacción de éstas con sus huéspedes vegetales, añadiendo así nuevas evidencias que anticipan una función universal de esta proteína en el establecimiento de infecciones intracelulares crónicas.

### *Hfq influye en el crecimiento y en el metabolismo central de S. meliloti*

La pérdida de la función de Hfq afecta al crecimiento en vida libre de *S. meliloti*, confirmando el efecto pleiotrópico de esta mutación observado también en otras bacterias (Tsui *et al.*, 1994). Los *microarrays* fenotípicos (Biolog) revelaron que mientras los patrones de utilización de diversas fuentes de carbono por la cepa silvestre y la mutante en el gen *hfq* fueron esencialmente idénticos, la utilización de algunos compuestos nitrogenados (i.e. aminoácidos y dipéptidos) y la resistencia a ciertos agentes químicos fueron fenotipos alterados por la ausencia de Hfq. Las diferencias en la utilización de fuentes de nitrógeno fueron incluso más evidentes en experimentos similares publicados por otros autores en los que los medios de cultivo utilizados para los ensayos se optimizaron para *S. meliloti* (Gao *et al.*, 2010). En conjunto, estos resultados anticipan alteraciones (e.g. en los componentes estructurales de la membrana) en el mutante *hfq* que resultan en cambios específicos en el transporte y metabolismo de pequeñas moléculas.

Para profundizar en las bases moleculares de este fenotipo hemos combinado el análisis del perfil transcriptómico del mutante *S. meliloti* 1021 $\Delta$ *hfq* con el del proteómico de *S. meliloti* 2011-3.4, en la fase exponencial de crecimiento de las bacterias en medio completo TY. Estos experimentos identificaron 168 transcritos y 33 polipéptidos que muestran una acumulación diferencial entre las cepas silvestres, 1021 y 2011, y sus respectivos mutantes en el gen *hfq* con 9 genes en común para ambos sets. Las diferencias entre las cepas salvajes 1021 y 2011, ya descritas en la Introducción de esta memoria, podrían ser una explicación parcial a la escasa coincidencia entre proteínas y transcritos regulados por Hfq en ambos fondos genéticos. Sin embargo, esto ha sido también observado en experimentos similares en

*Salmonella* y probablemente refleje el impacto global de esta proteína a diferentes niveles en la expresión génica; la propia transcripción (i.e. Hfq tiene cierta capacidad para unir DNA), la estabilidad de los diferentes transcritos y la traducción (Ansong *et al.*, 2009). En cualquier caso, ambas aproximaciones convergen en la identificación de genes que codifican el componente periplásmico de transportadores ABC y enzimas metabólicas como las categorías funcionales mayoritariamente afectadas por la mutación Hfq. Otros estudios similares y simultáneos al nuestro aumentaron el número de genes y proteínas alterados en el mutante *hfq* a 483 y 50, respectivamente, probablemente debido a la utilización de otras condiciones de cultivo (i.e. fase logarítmica tardía o estacionaria de crecimiento) y a la aplicación de criterios menos estrictos en la selección final de los genes diferencialmente expresados (i.e.  $M>0$  o  $M<0$ ) (Barra-Bily *et al.*, 2010; Gao *et al.*, 2010). No obstante, la distribución funcional de estos genes/proteínas es esencialmente coincidente con la derivada de nuestros experimentos. El papel de Hfq en la regulación masiva del transporte de nutrientes y el metabolismo central parece ser universal en bacterias pues también ha sido evidenciado por estudios transcriptómicos/proteómicos de mutantes *hfq* en otros modelos experimentales como *E. coli*, *S. typhimurium*, *P. aeruginosa* y *Y. pestis* (Sonnleitner *et al.*, 2006; Guisbert *et al.*, 2007; Sharma *et al.*, 2007; Sittka *et al.*, 2007; Geng *et al.*, 2009). La pérdida de Hfq resultó también en la represión general de genes que codifican proteínas que operan en los mecanismos para la producción central de energía basada en el transporte y catabolismo de azúcares, además de otros genes que pueden participar en la biosíntesis de los componentes estructurales de macromoléculas como proteínas o ácidos nucleicos (i.e. aminoácidos y nucleótidos). Este conjunto de alteraciones probablemente impongan una reprogramación de la actividad metabólica en los mutantes *hfq* sesgada hacia el ciclo de los ácidos tricarboxílicos (TCA) y la gluconeogénesis, que es apoyada por la inducción generalizada de los genes del transporte de péptidos y aminoácidos y algunos del metabolismo del nitrógeno así como del enzima PckA (i.e. fosfoenolpiruvato carboxiquinasa) en ausencia de la proteína (Fig. I-20). Además, algunos genes que codifican componentes de la cadena de transporte electrónico aparecieron reprimidos en el mutante, lo que probablemente contribuya a la reducción de la fuerza protón motriz a través de la membrana citoplasmática. Estos cambios metabólicos podrían explicar el fenotipo de crecimiento observado en el mutante *hfq*.

Un segundo análisis de los datos transcriptómicos/proteómicos en relación al contenido en mRNAs del CoIP-RNA obtenido en fase exponencial de crecimiento identificó 70 mRNAs comunes a ambas listas. Esto supone que escasamente el 7% de los mRNAs representados en CoIP-RNA presentaron expresión dependiente de Hfq, mientras que un 42% de los mRNAs/proteínas diferencialmente expresados en los

mutantes *hfq* aparecieron asociados a la proteína. La escasa coincidencia entre ambos grupos de datos puede atribuirse a cuatro factores principales. En primer lugar, alguno de los genes desregulados en los mutantes *hfq* codifican factores de transcripción, lo que también alteraría la expresión de los genes dependientes de éstos; en otras palabras, no todos los mRNAs diferencialmente expresados en los mutantes tienen que ser necesariamente dianas directas de Hfq. Segundo, puede haber un número considerable de mRNAs representados en CoIP-RNA que no satisfacen el estricto factor de enriquecimiento impuesto en nuestro análisis para su catalogación como asociados a Hfq. De la misma forma sería también mayor el número de genes catalogados como diferencialmente expresados en los mutantes *hfq* utilizando en los análisis valores de M más bajos (Gao *et al.*, 2010). En tercer lugar, algunas de las interacciones mRNA-Hfq pueden no tener efecto cuantificable mediante nuestra metodología sobre los niveles de los transcritos y/o las proteínas codificadas por éstos. Finalmente, los extremos precisos de las regiones 5'/3'-UTR de los mRNAs no han sido todavía establecidos en *S. meliloti* por lo que nuestros cálculos para estimar los mRNAs representados en CoIP-RNA han tenido en cuenta tan solo las regiones mapeadas total o parcialmente en las secuencias codificantes de cada mensajero.

Los mRNAs dependientes de Hfq que aparecieron también enriquecidos en CoIP-RNA (identificados en la Tabla II-3) codifican mayoritariamente funciones de transporte y enzimas metabólicas. Hfq puede contribuir a su regulación de dos formas: *i*) por asociación directa con los mRNAs en regiones diversas de éstos, afectando su traducción y/o estabilidad (Vytvytska *et al.*, 2000; Joanny *et al.*, 2007; Sittka *et al.*, 2008), y *ii*) a través de su interacción con sRNAs con efectos similares a los anteriores.. La distribución de los cDNAs derivados de estos mRNAs fue en general heterogénea en cada *loci*. No obstante, en muchos de ellos, particularmente en los correspondientes a proteínas periplásmicas de transportadores ABC de compuestos nitrogenados, se observó una cobertura preferente de sus regiones 5', que incluye su posible inicio de transcripción y los primeros codones de la secuencia codificante. Estas son mayoritariamente las regiones de interacción de los *trans*-sRNAs con sus mensajeros diana, por lo que es previsible que los mRNAs con este patrón de enriquecimiento en el CoIP-RNA estén regulados a nivel post-transcripcional por sRNAs dependientes de Hfq. En este sentido en el catálogo de mRNAs representados en CoIP-RNA de *Salmonella* hay muchos (e.g. los transportadores *dppA* o *oppA*) que son dianas experimentalmente confirmadas de sRNAs (e.g. GcvB) expresados por esta enterobacteria (Sittka *et al.*, 2008).

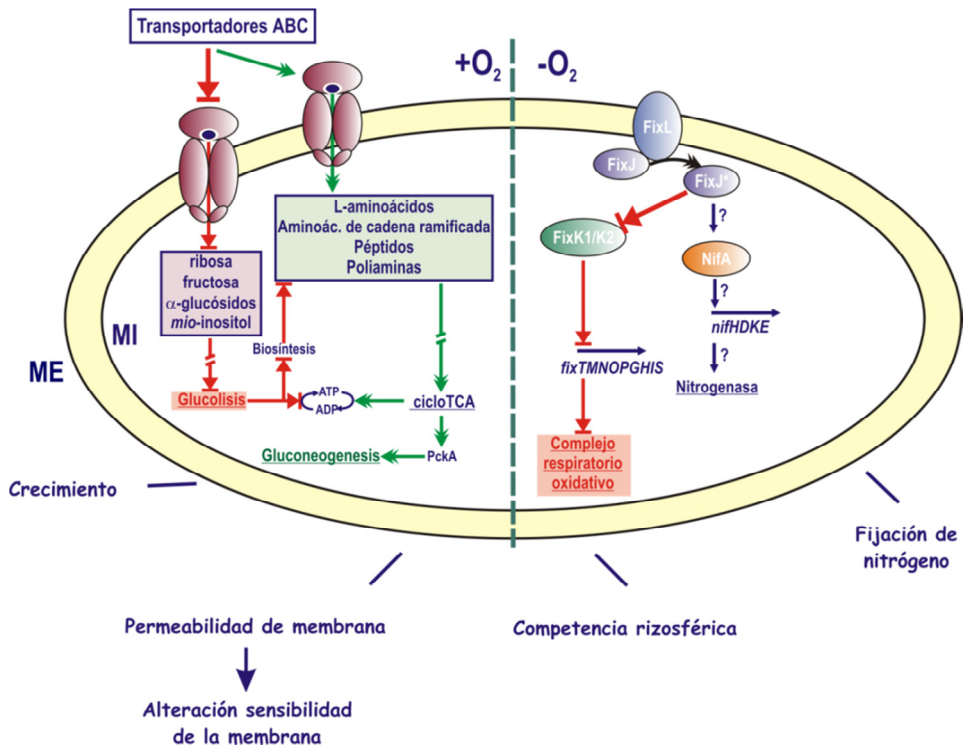


Fig. II-20. Resumen de las rutas y fenotipos asociados a la mutación *hfq* en *S. meliloti*. Las flechas dobles indican rutas favorecidas y las flechas interrumpidas, las desfavorecidas en ausencia de Hfq. +O<sub>2</sub>, condiciones aerobias; -O<sub>2</sub>, condiciones microaerófilas.

### La pérdida de Hfq afecta a diversos estadios de la simbiosis *S. meliloti*-alfalfa

Los eventos tempranos de la interacción simbiótica de los rizobios con sus leguminosas huésped implican una activa colonización de las rizosfera y la subsiguiente respuesta a compuestos específicos presentes en los exudados de las raíces (i.e. flavonoides) para desencadenar la cascada de señalización de los factores Nod que deriva en la organogénesis nodular (Prell and Poole, 2006; Jones *et al.*, 2007; Gibson *et al.*, 2008). La rizosfera es un nicho ecológico complejo que provee a la bacteria de un amplio abanico de compuestos carbonados y nitrogenados. El éxito ecológico del microsimbionte demanda una gran plasticidad metabólica, que en *S. meliloti* está garantizada por un amplio conjunto de genes que codifican factores de transcripción, transportadores ABC y enzimas metabólicas (Galibert *et al.*, 2001). Nuestros resultados muestran que el mutante *hfq* de *S. meliloti*, cuando es inoculado independientemente, es capaz de nodular plantas de alfalfa de forma similar a la cepa salvaje, aunque se observa un ligero retraso en sus cinéticas de nodulación. Estos resultados evidencian que la mutación en el locus *hfq* no compromete la percepción y la producción de las señales simbióticas específicas (i.e. flavonoides y

factores Nod) que desencadenan la organogénesis nodular pero sugieren que la adaptación de la bacteria en la rizosfera está comprometida. De hecho, en presencia de la cepa parental la capacidad del mutante *hfq* para inducir nódulos en alfalfa está significativamente disminuida. La eficiencia de las bacterias para incorporar y metabolizar compuestos carbonados como trehalosa, sacarosa, maltosa o *mio*-inositol, abundantes en la rizosfera, es muy importante para competir con éxito por la ocupación de los nódulos (Fry *et al.*, 2001; Jensen *et al.*, 2005). Los operones *iolRCDEB* y *aglEFGAK* que especifican las funciones para el transporte y catabolismo de *mio*-inositol y  $\alpha$ -glucósidos, respectivamente, están fuertemente reprimidos en el mutante *hfq* por lo que estas alteraciones podrían suponer una desventaja decisiva para la colonización competitiva de la rizosfera de plantas de alfalfa. Una caracterización fenotípica de mutantes *hfq* de *S. meliloti* algo más extensa que la presentada en este trabajo reveló también deficiencias de los mismos en la motilidad así como patrones de acumulación de moléculas autoinducidas del *quorum sensing* sustancialmente diferentes a los observados en la cepa silvestre (Gao *et al.*, 2010). Estos fenotipos pueden comprometer también la captación de nutrientes y la coordinación adecuada de las respuestas colectivas de las bacterias a las señales de la planta, con el consiguiente impacto negativo en su competencia rizosférica por la nodulación. Aunque el mutante *S. meliloti* 1021 $\Delta$ *hfq* es capaz de inducir nódulos en raíces de plantas de alfalfa (fenotipo Nod<sup>+</sup>), una proporción importante de ellos parecen no fijadores (65%-74%). Además, hemos observado un retraso significativo en el comienzo del proceso de fijación simbiótica de nitrógeno (i.e. expresión de la leg-hemoglobina) en el resto de los nódulos inducidos por el mutante (36%- 45%) en comparación a la cepa salvaje 1021. Como era de esperar, estas deficiencias afectan de forma negativa al rendimiento de la simbiosis (i.e. crecimiento de las plantas). El fenotipo Fix<sup>-</sup> de mutantes *hfq* de *S. meliloti* ha sido también observado en otros dos estudios independientes (Barra-Bily *et al.*, 2010b; Gao *et al.*, 2010) de forma que todas estas evidencias indican un impacto de la actividad de Hfq también en estadios intermedios o tardíos de la simbiosis. Un examen detallado de los nódulos blancos inducidos por la cepa 1021 $\Delta$ *hfq* reveló una leve expresión de la leg-hemoglobina y escaso número de células vegetales ocupadas por bacteroides fijadores de nitrógeno. Además, en estos nódulos se observaron signos de senescencia prematura. No obstante, la monitorización de la organogénesis nodular en el mutante *hfq* y su cepa parental revelaron que aunque la cinética de formación de las estructuras nodulares es más lenta en el mutante, éste mantiene cierta capacidad para diferenciarse a bacteroide. Estos hallazgos sugieren que la pérdida de Hfq afecta a la capacidad de *S. meliloti* de sobrevivir en el ambiente intracelular dentro del huésped. Este fenotipo parece ser común a todas las bacterias patógenas intracelulares en las que se ha

investigado la función de Hfq, con independencia de su grado de relación filogenética (Robertson and Roop, 1999; Roop *et al.*, 2002; Sonnleitner *et al.*, 2003; Christiansen *et al.*, 2004; Ding *et al.*, 2004; Sittka *et al.*, 2007).

En el interior de los nódulos las fuentes de energía disponibles para el microsimbionte parecen estar más definidas y son menos limitantes que en el ambiente rizosférico. Las plantas proveen a los rizobios de ácidos dicarboxílicos como fuente de carbono prioritaria para la obtención de la energía que demanda la fijación simbiótica de nitrógeno (Prell and Poole, 2006). Sin embargo, estudios más recientes sugieren que compuestos nitrogenados como los aminoácidos de cadena ramificada (e.g. leucina o valina) son también esenciales para el desarrollo y persistencia intracelular de los bacteroides (Prell *et al.*, 2009). Los rizobios adquieren los aminoácidos de cadena ramificada a través de los transportadores Liv y Aap esencialmente que, no obstante, parecen tener afinidad por un amplio espectro de sustratos (Prell and Poole, 2006). Los mRNAs (proteínas) codificados por ambos sistemas de transporte están inducidos en el mutante *hfq* y representados en CoIP-RNA. Sin embargo, los genes para su catabolismo no fueron catalogados como diferencialmente expresados por nuestro análisis aunque en otros estudios similares aparecieron discretamente reprimidos en el mutante *hfq* (Gao *et al.*, 2010), posiblemente debido a los diferentes criterios impuestos en cada caso para el tratamiento de los datos transcriptómicos. La descoordinación entre el transporte y catabolismo de estos compuestos podría justificar, aunque sólo parcialmente, el fenotipo endosimbiótico observado en los mutantes *hfq*.

Además de nutrientes, los rizobios perciben y responden a otras señales procedentes de la planta para la colonización exitosa de los nódulos (Jones *et al.*, 2007; Gibson *et al.*, 2008). Éstas incluyen especies reactivas de oxígeno liberadas en respuesta a la infección (Soto *et al.*, 2009), péptidos transportados al interior de las bacterias por el producto del gen *bacA* para iniciar la diferenciación del bacteroide (Marlow *et al.*, 2009; Van de Velde *et al.*, 2010), la acidez de los compartimentos intracelulares (Glazebrook *et al.*, 1993) o el ambiente microaerófilo que demanda la actividad del complejo nitrogenasa para la reducción del nitrógeno atmosférico (Dixon and Kahn, 2004).

Entre las proteínas reprimidas en el mutante *hfq* (i.e. 2011-3.4) se identificaron las chaperonas GroEL2, GroEL3, GrpE e IbpA. El genoma de *S. meliloti* contiene cuatro operones *groESL* y el gen adicional *groEL*, siendo *groEL1* indispensable para la nodulación y para la fijación de nitrógeno (Ogawa and Long, 1995; Bittner *et al.*, 2007). Se puede especular, por tanto, que las chaperonas dependientes de Hfq

podrían contribuir a paliar el estrés prologando al que son sometidas las bacterias endosimbióticas dentro de las células vegetales, fundamentalmente durante la infección de los pelos radiculares. Sin embargo, las proteínas esenciales para la resistencia al estrés oxidativo durante esta etapa de la simbiosis son, entre otras, las superóxido dismutasas (SodB y SodC), la cisteína sintasa (CysK), la bacterioferritina (Brf), la glutatión sintetasa (GshB1) o la peroxiredoxina codificada por el gen *SMc00072* en *S. meliloti*. La comparación del perfil proteómico de la cepa de referencia *S. meliloti* 1021 y el de su mutante *hfq* después de someter a las bacterias a una situación generalizada de estrés (i.e. fase estacionaria tardía de crecimiento) reveló también la represión de todas estas proteínas en la cepa mutante (Barra-Bily *et al.*, 2010), lo que justificaría la cinética de infección más lenta observada en ausencia de Hfq.

Por otra parte, los perfiles transcriptómicos revelaron que la acumulación de los transcritos de los reguladores FixK1/FixK2 y los correspondientes a los genes regulados por éstos, que codifican los componentes de la cadena de transporte de electrones asociada a la nitrogenasa, es dependiente de Hfq. La cobertura de esta región por CoIP-RNA abarcó también todo este mRNA policistrónico. Sorprendentemente, aunque la transcriptómica no reveló expresión diferencial de *nifA* en el mutante *hfq*, probablemente porque la expresión de este gen requiere condiciones de microaerobiosis más estrictas que la de los genes *fix*, su mRNA también apareció representado en CoIP-RNA. Experimentos de RT-PCR con RNA total obtenido en condiciones microaeróbicas han confirmado también la represión del gen *nifA* de *S. meliloti* en ausencia de Hfq, tal como había sido demostrado previamente en otras  $\alpha$ -proteobacterias diazotróficas como *A. caulinodans* o *R. capsulatus* (Kaminski and Elmerich, 1998; Drepper *et al.*, 2002; Barra-Bily *et al.*, 2010b). La ausencia de ambos reguladores, FixK1/K2 y NifA, en el mutante *hfq* afectaría de forma severa a la síntesis del complejo nitrogenasa, comprometiendo la fijación biológica de nitrógeno en los bacteroides diferenciados a término. . Curiosamente, estudios recientes han demostrado que la función de NifA se extiende más allá de la regulación de la fijación de nitrógeno, de forma que la mutación de su gen afecta otros procesos celulares como la producción de determinadas proteínas extracelulares o el *swarming*, que son también relevantes para el establecimiento de una simbiosis eficiente (Gong *et al.*, 2007; Barra-Billy *et al.*, 2010b). ( Estudios recientes en *R. leguminosarum* bv. *viciae* sugieren que Hfq y RNaseE forman parte del complejo del degradosoma, de forma que la regulación post-transcripcional de *nifA* implicaría la interacción de Hfq con su mRNA posibilitando el procesamiento de la región 5'-UTR por RNaseE para hacer así accesible el sitio SD a los ribosomas, en un mecanismo independiente de sRNAs (Zhang and Hong, 2009). Dado el hecho



evidente de la asociación del mRNA de *nifA* a Hfq en *S. meliloti* se puede especular con la existencia de un mecanismo similar en esta bacteria para la regulación de la traducción de NifA.

En resumen, la combinación de datos transcriptómicos, proteómicos y de secuenciación masiva de CoIP-RNA explica el fenotipo pleiotrópico de la mutación en el gen *hfq* a la vez que pone en evidencia posibles mecanismos de regulación génica post-transcripcional mediados por esta proteína.

### *Localización subcelular de Hfq en S. meliloti*

La maquinaria bacteriana de degradación de RNAs implica frecuentemente a proteínas que se asocian a la cara interna de la membrana citoplasmática, como por ejemplo RNasa E (Taghbalout and Rothfield, 2007,2008; Khemici *et al.*, 2008). De forma similar, RNasa III y RNasa P también han sido detectadas en asociación con la membrana bacteriana (Miczak *et al.*, 1991), incluso la poli (A) polimerasa (Jasiecki and Wegrzyn, 2005). En *E. coli*, se ha descrito que Hfq se localiza preferentemente en membrana, formando un complejo multiproteico con RNasa E (Diestra *et al.*, 2009; Ikeda *et al.*, 2011). Nuestros resultados de Western-blot revelaron que Hfq se recupera exclusivamente entre las proteínas de membrana después del fraccionamiento celular. La observación de bacterias que expresan una fusión traduccional Hfq::GFP reveló la concentración de la fluorescencia en *foci* situados en uno o en los dos polos de la célula. Este patrón de fluorescencia no excluye la asociación de Hfq a membrana aunque también puede tener su origen en la concentración de componentes celulares de diversos compartimentos, o de los elementos excluidos del nucleoide (Khemici and Carpousis, 2004; Khemici *et al.*, 2008). Un problema que ha sido documentado en la bibliografía es que en experimentos de microscopía fluorescente convencional no se obtiene la suficiente resolución para dilucidar la distribución de moléculas en la fina capa que compone la cara interna de la membrana citoplasmática (Huang *et al.*, 2008), con lo que la correcta visualización de elementos en este compartimento plantearía una dificultad añadida.

Nuestros resultados no muestran tampoco ninguna deslocalización de la fluorescencia en un mutante viable de RNasa E de *S. meliloti*. Aunque la asociación de Hfq con esta ribonucleasa está muy estudiada en *E. coli*, todavía es desconocida en el orden Rhizobiales, exceptuando el caso de *R. leguminosarum* bv. *viciae*, donde se ha visto que Hfq co-precipita con los complejos RNasa E, sin que haya evidencia directa de su interacción (Zhang and Hong, 2009). En *S. meliloti* no existe información alguna sobre esta asociación y, además, no están caracterizados ni los motivos estructurales de RNasa E y ni su papel en el degradosoma. La capacidad de unión a la membrana

interna y a otras proteínas que integran el degradosoma (i.e. Hfq) está especificada por la región C-terminal de la RNasa E de *E. coli*, que es muy divergente a la de su homóloga en *S. meliloti* (Khemici *et al.*, 2008; Ikeda *et al.*, 2011). La caracterización del complejo multiproteico del degradosoma en *S. meliloti* proporcionará, por tanto, nueva información relevante para entender los mecanismos de regulación génica post-transcripcional que involucran a Hfq.



## **ANEXO CAPÍTULO 2**



## TRANSCRITOS DIFERENCIALMENTE EXPRESADOS EN *S. MELILOTI* OBTENIDOS EN HIBRIDACIÓN DE MICROARRAYS Sm14kOLI.

Tabla II-1. Transcritos diferencialmente expresados en el mutante 1021 $\Delta$ hfq

TRANSCRITOS REPRIMIDOS EN EL MUTANTE 1021 $\Delta$ hfq		
TRANSPORTE	Función predicha	Valor M
<i>Transportadores ABC mio-inositol</i>		
<i>ibpA</i> (SMb20712)*	myo-inositol transporter/periplasmic binding protein	-1,24
<i>iatA</i> (SMb20713)	ATP-binding protein	-2,82
<i>iatP</i> (SMb20714)	Inner membrane component/Permease	-1,05
<i>Transportadores ABC <math>\alpha</math>-glucósidos/ carbohidratos/ ács. orgánicos/ alcoholes</i>		
<i>aglE</i> (SMc03061)*	Periplasmic solute-binding protein	-3,94
<i>aglF</i> (SMc03062)	Inner membrane component/Permease	-1,21
<i>Otros genes de transporte</i>		
SMc02514	ABC transporter glycerol-3-phosphate and glycerol/periplasmic solute-binding component	-2,05
SMc02519	Putative glycerol-3-phosphate and glycerol ABC transporter/ATP-binding component	-2,85
<i>frcB</i> (SMc02171)*	Putative fructose ABC-type transport system/periplasmic component	-1,04
SMc02031	Ribose transport system permease	-1,29
SMc04317	ABC transporter/periplasmic solute-binding component/Fe uptake	-1,04
SMc04439	Putative ABC-transporter ATP-binding protein/histidine, glycine betaine	-1
SMb20072	ABC transporter/sugars/periplasmic solute-binding protein, induced by myo-inositol	-5,87
SMb20144	ABC transporter/periplasmic solute-binding protein/carbohydrates, org. acids, alcohols	-1,27
<b>METABOLISMO DE PEQUEÑAS MOLÉCULAS</b>		
<i>Operón del catabolismo de mio-inositol</i>		
<i>iolC</i> (SMc01165)	Putative sugar kinase	-1,2
<i>iolD</i> (SMc01166)*	Putative malonic semialdehyde oxidative decarboxylase	-2,01
<i>iolE</i> (SMc00433)*	Putative myo-inositol catabolism protein/Xylose isomerase-like, TIM barrel	-1,79
<i>iolB</i> (SMc00432)	Putative myo-inositol catabolism protein/Cupin, RmlC-type	-1,67
<i>Otros genes relacionados con el metabolismo de pequeñas moléculas</i>		
SMc01163	Glucose-fructose oxidoreductase	-1,48
<i>ilvC</i> (SMc04346)	Ketol-acid reductoisomerase/branched-chain amino acid biosynthesis	-1,83
SMc03211	Putative 4-hydroxyphenylpyruvate dioxygenase/aromatic amino acids metabolic process	-2,48
SMc03253	L-proline cis-4-hydroxylase/peptidyl-amino acid modification	-2,09
<i>pheAa</i> (SMc03858)	Putative chorismate mutase/aromatic amino acids biosynthetic process	-1,18
<i>mtbC</i> (SMc04331)	Putative dimethylamine corrinoid protein/methionine biosynthetic process	-2,16
<i>cobP</i> (SMc04305)	Bifunctional adenosylcobalamin biosynthesis protein CobP	-1,41
SMc04342	Dihydropteroate synthase (DHPS)/folic acid and derivative biosynthetic process	-1,61
<i>mttB1</i> (SMc04330)	Putative trimethylamine methyltransferase/methanogenesis	-1,97
SMc02045	D-3-phosphoglycerate dehydrogenase/Acetylation Amino-acid biosynthesis	-1,05
<i>bkdAb</i> (SMc03202)	Probable 2-oxoisovalerate dehydrogenase beta subunit/glycolysis oxidoreductase	-1,04
SMc02689	Probable aldehyde dehydrogenase	-1,12

<i>pyrC</i> (SMc02166)	Probable dihydroorotase/pyrimidine base biosynthetic process	-1,18
<i>glyA1</i> (SMc01770)	Probable serine hydroxymethyltransferase/ glycine metabolic process	-1,22
SMc00982	Putative dioxygenase Rieske 2Fe-2S family protein/carbon compounds	-1,13
<i>purU1</i> (SMc03205)	Putative formyltetrahydrofolate deformylase/purine ribonucleotide biosynthesis	-2,71
<i>soxB2</i> (SMc03933)	Putative sarcosine oxidase subunit B/tetrahydrofolate metabolic process	-2,65
<i>frcK</i> (SMc02167)	Putative fructose transport system kinase	-1,36
<i>adhA1</i> (SMa1296)	Alcohol dehydrogenase, Zn-dependent class III/fermentation	-3,24
SMa1156	Alcohol dehydrogenase, Zn-dependent class III/Energy metabolism, carbon	-1,93
<i>hemN</i> (SMa1266)	HemN coproporphyrinogen III oxidase/porphyrin biosynthetic process	-2,27
SMa1166	Protein containing an alpha/beta hydrolase fold/alpha-beta hydrolase	-1,15
<i>azu1</i> (SMa1243)	Azu1 pseudoazurin	-1,71

---

## OTROS PROCESOS CELULARES

---

### *Transporte electrónico/fijación de nitrógeno*

<b>Operon <i>fixN1-O1-Q1-P1-G-H-I1-S1</i></b>		
<i>fixN1</i> (SMa1220)	FixN1 cytochrome c oxidase subunit I/Cytochrome c oxidase, subunit I	-1,27
<i>fixQ1</i> (SMa1214)	FixQ1 nitrogen fixation protein/Cbb3-type cytochrome oxidase component	-1,78
<i>fixP1</i> (SMa1213)	FixP1 di-heme c-type cytochrome/Cytochrome c oxidase cbb3-type, subunit III	-1,2
<i>fixG</i> (SMa1211)	FixG iron sulfur membrane protein/Cytochrome c oxidase cbb3 type, accessory protein FixG	-1,15
<b>Operon <i>fixK2-N2-O2-Q2-P2</i></b>		
<i>fixK2</i> (SMa0762)	FixK2 transcriptional regulator	-2,42
SMa0763	FixM flavoprotein oxidoreductase	-2,49
<i>fixQ2</i> (SMa0767)	FixQ2 nitrogen fixation protein/Cbb3-type cytochrome oxidase component	-1,96
<b>Operon <i>fixL-J-T1-K1-M</i></b>		
<i>fixK1</i> (SMa1225)	FixK1 transcriptional regulator	-1,7
<i>fixM</i> (SMa1223)	FixM flavoprotein oxidoreductase	-2,86
SMa1151	Putative FixM flavoprotein oxidoreductase	-1,96
<i>napF</i> (SMa1240)	NapF component of periplasmic nitrate reductase/anaerobic respiration	-1,82
SMa1170	Probable Cytochrome c	-1,29
<i>cycA</i> (SMb21367)	Putative cytochrome c class I protein, probably cytochrome c4	-1,65
<i>etfA1</i> (SMc00728)	Putative electron transfer flavoprotein alpha-subunit alpha-ETF flavoprotein	-1,01
<i>ccmD</i> (SMc03850)	Putative heme exporter D (cytochrome C-type biogenesis protein) transmembrane	-1,34

### *Metabolismo central de intermediarios*

<i>phoX</i> (SMc02634)	Monomeric alkaline phosphatase/phosphorus metabolism	-2,03
SMc02148	Putative polyphosphate kinase 2	-1,27
<b>Operon <i>nosRZDFYLX</i></b>		
<i>nosZ</i> (SMa1182)	NosZ nitrous oxide reductase/nitrogen metabolism	-1,39
<i>nosD</i> (SMa1183)	NosD nitrous oxidase accessory protein/nitrogen metabolism	-2,13

### *Metabolismo de macromoléculas*

SMa1126	Putative protease, transmembrane protein	-2,8
<i>degP4</i> (SMa1128)	DegP4 protease like protein	-1,03

**Reguladores transcripcionales**

SMc02030	Transcription regulator LysR family	-1,08
<i>fixT3</i> (SMc03254)	Putative antikinase	-1,09
<i>nrdR</i> (SMc01771)	Putative transcriptional repressor	-1,07
SMa1207	Transcriptional regulator, CAP/Crp family	-1,53

**Miscelánea**

<i>flaA</i> (SMc03037)	Flagellin A/Mobility chemotaxis	-1,05
SMb21053	Putative membrane-anchored glycosyltransferase/surface polysaccharides biosynthesis	-1,35
<i>cyaP</i> (SMa1103)	Adenylate/guanylate cyclase	-2,01
SMa1149	Universal stress protein (Usp)/response to stress	-2,76
SMa1147	Universal stress protein (Usp)/response to stress	-2,55
SMa1158	Universal stress protein (Usp)/response to stress	-1,8

**PROTEÍNAS HIPOTÉTICAS****Proteínas transmembrana probables**

SMc01986	Hypothetical transmembrane protein	-1,97
SMc00317	Putative transporter/auxin efflux carrier	-1,89
SMb20139	Conserved hypothetical transmembrane protein	-1,66
SMa1097	Hypothetical transmembrane protein	-1,41

**Otras proteínas**

SMc04339	Hypothetical protein	-2,52
SMc04310	Hypothetical protein	-1,9
SMc01747	Hypothetical protein	-1,78
SMc04345	Hypothetical protein	-1,67
SMc00391	Hypothetical protein	-1,34
SMc01600	Conserved hypothetical protein	-1,19
SMc00472	Conserved hypothetical protein	-1,14
SMc04337	Conserved hypothetical protein	-1,06
SMc03251	Conserved hypothetical protein	-1,05
SMa1082	Hypothetical protein	-2,13
SMa1154	Conserved hypothetical protein	-2
SMa1050	Hypothetical protein	-1,45
SMa0633	Hypothetical protein	-1,29
SMa1176	Hypothetical protein	-1,24
SMa1077	Nex18 Symbiotically induced conserved protein	-1,24
SMa1052	Conserved hypothetical protein	-1,02

**TRANSCRITOS INDUCIDOS EN EL MUTANTE 1021 $\Delta$ hfq****TRANSPORTE****Transportadores ABC L-aminoácidos/péptidos**

<i>aapJ</i> (SMc02118)*	Extracellular solute-binding protein	1,7
<i>aapQ</i> (SMc02119)	Inner membrane component	1,42
<i>aapP</i> (SMc02121)*	ATP-binding protein	1,95

**Transportadores ABC Spermidina/putrescina**

SMc01966	Extracellular solute-binding protein	1,06
SMc01965	ATP-binding protein	6,39



SMc01963	Permease	1,56
<b>Transporte de alta afinidad de aminoácidos de cadena ramificada</b>		
<i>livH</i> (SMc01951)	Inner-membrane translocator	4,99
<i>livM</i> (SMc01950)	Inner-membrane translocator	1,6
<i>livG</i> (SMc01949)	ATP-binding protein	1,91
<i>livF</i> (SMc01948)	ATP-binding protein	2,55
SMc01947	Conserved hypothetical transmembrane protein	3,39
<i>livK</i> (SMc01946)*	Leu/Ile/Val-binding protein	1,88
<b>Transportadores ABC Uracil/uridina</b>		
SMc01823	ATP-binding protein	1,74
SMc01824	Conserved hypothetical protein/GCN5-related N-acetyltransferase	1,07
SMc01825	Inner membrane component	1
SMc01827	Periplasmic solute-binding protein	2,99
<b>Transportadores ABC Putrescina/agmatina</b>		
SMc01652	Solute-binding protein	1,04
SMc01653	ATP-binding protein	1,26
<b>Transportadores ABC Tetrapéptidos/tripéptidos</b>		
<i>oppB</i> (SMb21197)	Inner membrane component/Permease	1,98
<i>oppC</i> (SMb21198)	Inner membrane component/Permease	1,82
<b>Otros genes de transporte</b>		
SMc02603	Permease	1,02
SMc00140	Putative amino-acid binding periplasmic protein	2,35
SMc04127	ATP-binding protein/Not classified ABC transporter	1,31
SMc01597	Putative amino-acid permease	1,27
SMc02259*	Putative periplasmic binding ABC transporter	1,55
<i>lolC</i> (SMc01935)	Lipoprotein-releasing system transmembrane protein/transport of large molecules	1,9
SMb21572	ABC transporter/periplasmic solute-binding protein precursor/amino acids	1,44
SMb20605	ABC transporter/periplasmic solute-binding protein	1,07
SMb21151	Putative sugar amine ABC transporter/periplasmic solute-binding protein	1,23
SMb20504	Putative ABC transporter periplasmic sugar-binding protein	1,45
SMb20442	Put. Trap-type pentose monosaccharide transp. system/peripl. solute-binding protein	1,03
SMc01962	Conserved hypothetical protein/nitrogen compound metabolic process acting on C-N	1,45
SMc01961	Conserved hypothetical protein/bleomycin resistance protein/dioxygenase	2,19
SMc01960	NADPH-dependent FMN reductase	2,35

---

## METABOLISMO DE PEQUEÑAS MOLÉCULAS

---

<b>Sistema de escisión de glicina</b>		
<i>gcvP</i> (SMc02049)	Glycine dehydrogenase	2,06
<i>gcvH</i> (SMc02048)	Glycine cleavage H-protein	2,02
<i>gcvT</i> (SMc02047)	Glycine cleavage T-protein	3,32
<b>Otros genes relacionados con el metabolismo de pequeñas moléculas</b>		
SMc01930	Putative methylmalonyl-CoA epimerase	3,26
SMc01815	Dihydropyrimidine dehydrogenase	2,21
SMc00115	Putative glyoxalase/bleomycin resistance protein/dioxygenase	1,35

SMc03773	Putative acetyltransferase	1,21
<i>cyaF7</i> (SMb20300)	Putative adenylate cyclase/global regulatory functions	1
SMb21202	Glyoxalase/bleomycin resistance protein/dioxygenase	1,09
SMa0374	Glyoxalase/bleomycin resistance protein/dioxygenase	1,06
SMa0376	Isochorismatase hydrolase/Cysteine hydrolase superfamily	1,7
SMa1038	Multicopper oxidase	1,14
SMa0380	Putative hydrolase/phosphonate metabolism	1,58
SMa1041	Copper binding protein/electron transport	1,32

---

**OTROS PROCESOS CELULARES**


---

**Reguladores transcripcionales**

<i>ohrR</i> (SMc01945)	MarR transcriptional regulator (detoxification)	1,87
------------------------	-------------------------------------------------	------

**Metabolismo macromoléculas**

<i>miaA</i> (SMc01435)	Probable tRNA delta2-isopentenylpyrophosphate transferase/tRNA modification	1,02
<i>degP1</i> (SMc02365)	Probable serine protease/degradation of proteins, peptides, glycopeptides	1,11
SMc00414	Putative phospholipid N-methyltransferase	1,69
SMc01929	RNaseJ	2,96
SMc03796	Endoribonuclease L-PSP/Cleaves mRNAs and inhibits traslation	1,56
SMa0426	Putative ATP-dependent DNA-helicase	1,07

**Metabolismo central de intermediarios**

SMc01814	Probable glutamate synthase small chain/central intermediary metabolism	2,33
<i>dht</i> (SMc01821)	Putative D-hydantoinase (dihydropyrimidinase)/central intermediary metabolism	1,1
<i>xdhA2</i> (SMb21011)	Xanthine dehydrogenase/electron transport	1,46

**Elementos estructurales**

<i>rplS</i> (SMc03863)	50S Ribosomal protein	1,09
<i>rpmA</i> (SMc03772)	50S Ribosomal protein L27	2,2

**Miscelánea**

<i>gst8</i> (SMc03882)	Putative glutathione S-transferase	1,03
<i>wrbA1</i> (SMc00943)	Flavoprotein WrbA	1,1
<i>wrbA3</i> (SMa1935)	Flavoprotein WrbA	1,24
<i>cvpA</i> (SMc00555)	Putative colicin V production homolog transmembrane protein	1,11
SMc04451	Putative chloramphenicol 3-O phosphotransferase/Drug-analog sensitivity and resistance	1,99
SMb20534	Xylose isomerase-like protein	1,3

---

**PROTEÍNAS HIPOTÉTICAS**


---

**Proteínas transmembrana probables**

SMc01931	Conserved hypothetical transmembrane protein	2,59
SMc01763	Hypothetical transmembrane protein	1,08
SMc02174	Hypothetical transmembrane protein	1,06
SMb20724	Conserved hypothetical exported protein	1,01
SMa1507	Conserved hypothetical protein/basic membrane lipoprotein	1,06

**Otras proteínas**

SMc01957	Conserved hypothetical protein	2,43
SMc00617	Conserved hypothetical protein	1,07

SMc04164	Conserved hypothetical protein	1,03
SMa0592	Conserved hypothetical protein	1,61
SMa0674	Hypothetical protein	1,55
SMa1043	Conserved hypothetical protein	1,41

\*Genes identificados por análisis proteómico en la cepa 2011 3.4

## **PROTEÍNAS DIFERENCIALMENTE EXPRESADAS EN *S. MELILOTI* 2011 3.4**

**Tabla II-2. Proteínas diferencialmente acumuladas en *S. meliloti* 2011 y 2011-3.4 revelado por análisis de geles 2D.**

<i>Proteína candidata<sup>a</sup></i>	<i>Función<sup>c</sup></i>
<b>GroEL2 (SMa0744)</b>	Chaperonin Cpn60/TCP-1
<b>GrpE (SMc01142)</b>	Probable heat shock protein/ GrpE nucleotide exchange factor
<b>IbpA (SMc04040)</b>	Probable heat shock protein/HSP20-like chaperone
<b>IbpA (SMB20712)*</b>	Putative myo-inositol ABC transporter
<b>AraA (SMB20895)</b>	ABC transporter, sugar uptake
<b>FrcB (SMc02171)*</b>	Putative fructose ABC transporter
<b>AgIE (SMc03061)*</b>	ABC transporter, $\alpha$ -glucosides
<b>IoIE (SMc00433)*</b>	Putative myo-inositol catabolism protein
<b>IoID (SMc01166)*</b>	Putative malonic semialdehyde oxidative decarboxylase
<b>AcsA1 (SMc04093)</b>	Probable acetyl-coenzyme A synthetase
<b>Tig (SMc02050)</b>	Probable trigger factor
<b>SMc01242</b>	Conserved hypothetical signal peptide protein
GroEL5 (SMB21566)	Chaperonin Cpn60/TCP-1/Putative heat sock protein
SMc02259*	Putative amino acid ABC transporter
DppA1 (SMc00786)	Putative amino acid or peptide ABC transporter
DppA2 (SMc01525)	Putative dipeptide binding periplasmic protein
LivK (SMc01946)*	Putative leucine-specific binding protein precursor
AapJ (SMc02118)*	Probable general L-amino acid ABC transporter
AapP (SMc02121)*	Probable general L-amino acid ABC transporter
SMc02378	Putative glycine betaine ABC transporter
ChoX (SMc02737)	Putative choline ABC transporter
Ocd (SMB21494)	Putative ornithine cyclodeaminase
ArgII (SMc03091)	Probable arginase
PckA (SMc02562)	Phosphoenolpyruvate carboxykinase
AhcY (SMc02755)	Probable adenosylhomocysteinase
SMc04385	Putative aldehyde dehydrogenase transmembrane protein
SMc00242	Hypothetical signal peptide protein
Efp (SMc00357)	Probable elongation factor P
SMc02884	Probable membrane lipoprotein
SMc00777	Conserved hypothetical protein
SMc02354	Conserved hypothetical protein
SMc02503	Conserved hypothetical protein
<b>SMc02911</b>	Conserved hypothetical protein

<sup>a</sup> Nomenclatura propuesta en <http://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi> Las proteínas reprimidas en el mutante se indican en negrita.

<sup>b</sup> Clasificación funcional según KEGG (<http://www.genome.jp/kegg/>).

\*Los transcritos correspondientes fueron encontrados también diferencialmente regulados en el mutante *S. meliloti* 1021 $\Delta$ hfq.

## GENES COMPARTIDOS POR ANÁLISIS TRANSCRIPTÓMICOS/ PROTEÓMICOS Y RNA-Seq DE Hfq<sup>FLAG</sup>.

**Tabla II-3. Genes compartidos por los análisis transcriptómicos, proteómicos y RNA-Seq.**

Gen	Función	Gen	Función
<i>SMa0633</i>	hypothetical protein	<i>SMc01814</i>	putative oxidoreductase
<i>SMa0744</i>	chaperonin GroEL;	<i>SMc01815</i>	dihydropyrimidine dehydrogenase
<i>SMa0762</i>	transcriptional regulator FixK	<i>SMc01821 (dhp)</i>	phenylhydantoinase
<i>SMa1038</i>	putative copper-containing oxidase	<i>SMc01827</i>	hypothetical protein
<i>SMa1041</i>	copper oxidase	<i>SMc01929</i>	hypothetical protein
<i>SMa1043</i>	hypothetical protein	<i>SMc01946 (livK)</i>	putative leucine-specific binding protein precursor
<i>SMa1077</i>	Nex18 symbiotically induced protein	<i>SMc01947</i>	hypothetical protein
<i>SMa1082</i>	hypothetical protein	<i>SMc01951 (livH)</i>	high-affinity branched-chain amino acid transport permease
<i>SMa1128 (degP4)</i>	DegP4 protease like protein	<i>SMc01966</i>	ABC transporter protein
<i>SMa1149</i>	hypothetical protein	<i>SMc02050 (tig)</i>	spermidine/putrescine ABC transporter periplasmic protein
<i>SMa1154</i>	hypothetical protein	<i>SMc02118 (appJ)</i>	trigger factor; Involved in protein export
<i>SMa1176</i>	hypothetical protein	<i>SMc02119 (appQ)</i>	general L-amino acid-binding periplasmic ABC transporter protein
<i>SMa1182 (nosZ)</i>	nitrous-oxide reductase	<i>SMc02121 (appP)</i>	general L-amino acid transport permease ABC transporter protein
<i>SMa1183 (nosD)</i>	NosD nitrous oxidase accessory protein	<i>SMc02148</i>	ATP-binding ABC transporter protein
<i>SMa1225</i>	transcriptional regulator FixK	<i>SMc02148</i>	hypothetical protein
<i>SMa1243 (azu1)</i>	Azu1 pseudoazurin	<i>SMc02259</i>	putative periplasmic binding ABC transporter protein
<i>SMa1266 (hemN)</i>	coproporphyrinogen III oxidase	<i>SMc02378</i>	putative copper-containing oxidase
<i>SMa1296 (adhA)</i>	Alcohol dehydrogenase, Zn-dependent class III	<i>SMc02514</i>	putative periplasmic binding ABC transporter protein
<i>SMa1507</i>	hypothetical protein	<i>SMc02519</i>	ABC transporter ATP-binding protein
<i>SMB20714 (iatP)</i>	myo-inositol ABC transporter	<i>SMc02689</i>	aldehyde hidrolase
<i>SMB20724</i>	hypothetical protein	<i>SMc02737 (choX)</i>	putative glycine betaine-binding ABC transporter protein
<i>SMB21011 (xihA2)</i>	xanthine dehydrogenase	<i>SMc02755 (ahcY)</i>	S-adenosyl-L-homocysteine hydrolase
<i>SMB21151</i>	Putative sugar amine ABC transporter, periplasmic solute-binding protein	<i>SMc02884</i>	putative lipoprotein precursor
<i>SMB21566 (groL5)</i>	chaperonin GroEL	<i>SMc02911</i>	hypothetical protein
<i>SMB21572</i>	hypothetical protein	<i>SMc03037</i>	flagellin A protein
<i>SMc00115</i>	putative amino-acid binding periplasmic protein	<i>SMc03061 (aglE)</i>	alpha-glucoside ABC transporter
<i>SMc00140</i>	putative signal peptide protein	<i>SMc03205 (purU1)</i>	periplasmic-binding protein
<i>SMc00242</i>	elongation factor P; Involved in peptide bond synthesis	<i>SMc03251</i>	formyltetrahydrofolate deformylase
		<i>SMc03253</i>	hypothetical protein
			putative L-proline 3-hydroxylase protein

SMc00357 ( <i>efp</i> )	hypothetical protein	SMc03772 ( <i>rpmA</i> )	50S ribosomal protein L27
SMc00472 ( <i>dppA1</i> )	hypothetical protein	SMc03858 ( <i>pheAa</i> )	chorismate mutase
SMc00777	periplasmic dipeptide-binding protein	SMc03863 ( <i>rplS</i> )	50S ribosomal protein L19
SMc00786	chaperonin GroEL	SMc03882 ( <i>gst8</i> )	putative glutathione S-transferase protein
SMc00943 ( <i>wrbA1</i> )	TrpR binding protein WrbA	SMc04040 ( <i>ibpA</i> )	heat shock protein
SMc00982	putative dioxygenase protein	SMc04093 ( <i>acsA2</i> )	acetyl-CoA synthetase
SMc01048 ( <i>hfq</i> )	RNA-binding protein Hfq	SMc04127	ABC transporter ATP-binding protein
SMc01525 ( <i>dppA2</i> )	putative dipeptide binding periplasmic protein	SMc04164	hypothetical protein
SMc01597	putative amino-acid permease protein	SMc04330 ( <i>mttB</i> )	putative trimethylamine methyltransferase protein
SMc01600	hypothetical protein	SMc04342	methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase
SMc01652	ABC transporter	SMc04346 ( <i>ilvC</i> )	ketol-acid reductoisomerase
SMc01770 ( <i>glyA</i> )	serine hydroxymethyltransferase	SMc04439	putative glycine betaine transport ATP-binding ABC transporter protein

## OLIGONUCLEÓTIDOS UTILIZADOS

Tabla II-4. Oligonucleótidos utilizados en este trabajo.

Nombre	Aplicación	Secuencia
Hfq_Fw	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-TCCGCAACGGGCTCTTCTAT-3'
Hfq_Rv	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-TGCCCGCGTTCGTATAGCCT-3'
hfqforw1	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-AAAAGAATTCATCTCTTTCTCAACACGGTCC-3'
hfqrev2	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-AAAAGGTACCCAGCTTCCTCATTCTCGAACA-3'
hfqforw3	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-AAAAGAATTCGTCCGCAAGCAAAAGATTTC-3'
hfqrev4	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-AAAAGGTACCCGTGCTTGTAGACGAGCTGA-3'
hfqi_1	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-GTGTGTAAGCTTCGGCCGCTTCTTTCTTA-3'
hfqi_2	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-CCGTGTAAGCTTCGGGATGTGAGGAACAT-3'
5HfqMut	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-TCTTCATCACCGCTGCTACC-3'
3HfqMut	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-AACGATCATGCCGTGAACGA-3'
5Hfq_C	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-GGCATATCGGCGATAACG-3'
3Hfq_C	Creación mutante 1021 $\Delta$ <i>hfq</i>	5'-CACGCTTGGTAATGTTC-3'
Hfqphus-i Fw	Fusión Hfq:eGFP	5'-ACCACAAGATCTGGAGGCAGCTTCTCA-3'
Hfqphus-i Rv	Fusión Hfq:eGFP	5'-CCGACTAGATCTGGATCCTCCGGATCTT-3'

3Hfq_C fus	Fusión Hfq:eGFP	5'-CATCCCGTCAGGAGGCAGCTT-3'
ChimericHFQFus	Fusión Hfq:eGFP	CGAAGATCCGGAGGATCCCATCCCGTCAGGAGGCAGCTT-3'
ChimericGFPFus	Fusión Hfq:eGFP	5'-GGTTCGGTAGCATGGTGAGCAAGGGCCGAGGAG-3'

## PLÁSMIDOS UTILIZADOS.

Tabla II-5. Construcciones utilizadas en este trabajo.

Nombre	Descripción
pGEM <i>hfq</i>	pGEM-T con la región genómica y ORF de <i>hfq</i>
pGEM $\Delta$ <i>hfq</i>	pGEM-T con la región genómica y sin el ORF de <i>hfq</i>
pk18 $\Delta$ <i>hfq</i>	pK18mobsacB con región promotora <i>hfq</i> y sin el ORF
pGEMHfq	pGEM-T con la secuencia codificante <i>hfq</i>
pJBHfq	pJB3Tc19 con la secuencia codificante <i>hfq</i>
pGEM <i>hfq</i> -tga	pGEM-T con la región genómica y el ORF <i>hfq</i> sin el codón de terminación de la traducción
pGEMeGFP	pGEM-T con el ORF de la eGFP
pBBs-eGFP	pBBR1MCS2 que expresa la GFP bajo el promotor syn
pGEMHfq:GFP	pGEM-T que expresa la fusión traduccional Hfq:GFP desde el promotor <i>hfq</i>
pJBHfq:GFP	pJB3Tc19 que expresa la fusión traduccional Hfq:GFP desde el promotor <i>hfq</i>



# **CAPÍTULO 3**

**LAS FAMILIAS DE sRNAs  $\alpha$ r15 Y  $\alpha$ r7:**

**CARACTERIZACIÓN DE**

**Smr15C1, Smr15C2 Y Smr7C**





## III-1. INTRODUCCIÓN

Aunque quizá no los más abundantes en número entre todos los tipos de sRNAs expresados por las bacterias, los *trans*-sRNAs son los más investigados y mejor conocidos a nivel funcional. Este tipo de transcritos son codificados entre ORFs (RIGs), en genes de localización preferentemente cromosómica que se expresan diferencialmente, por lo que su actividad se asocia a respuestas fisiológicas específicas ante cambios en el medio externo. Su mecanismo de acción mayoritario está basado en apareamientos, en algún caso imperfectos, con series cortas y generalmente discontinuas de nucleótidos complementarios localizados en las regiones 5'-UTR de sus mRNAs diana (codificados en *trans*), influyendo sobre la traducción y/o estabilidad de éstos. Si bien la hibridación entre secuencias complementarias es un mecanismo primigenio y universal de regulación génica post-transcripcional, los *trans*-sRNAs parecen haber estado sujetos a una rápida evolución, de forma que la conservación de la secuencia nucleotídica y estructura secundaria de cada ribo-regulador suele estar limitada a especies bacterianas filogenéticamente próximas. De la misma forma, la interacción limitada y poco extensa entre el *trans*-sRNA y su diana justifica también la casi nula correlación entre la localización de sus *loci* respectivos en los genomas bacterianos y la capacidad de estos ribo-reguladores para la regulación de múltiples mRNAs en la mayoría de los casos.

Por tanto, el estudio de su distribución y grado de conservación en bacterias, la caracterización de su expresión diferencial y de los factores que contribuyen a ésta y, fundamentalmente, la identificación de su (sus) mRNA diana y de los elementos que intervienen en la interacción con éste son aspectos esenciales para entender la función de los *trans*-sRNAs en el control de la fisiología bacteriana.

### III-1.1. REDUNDANCIA GENÓMICA Y FUNCIONAL DE LOS *TRANS*-sRNAs

Es relativamente frecuente que el análisis de la conservación en bacterias de los *trans*-sRNAs utilizando sus modelos de co-varianza revele la presencia de varias copias de los mismos en genomas individuales. Es el caso de ribo-reguladores bien caracterizados, ya referidos en la introducción general, como OmrA/OmrB, Qrr1-4 y GlmY/GlmZ, o de otros sRNAs de más reciente descubrimiento y, por tanto, todavía no caracterizados a nivel funcional como Smr15C1-2 y Smr14C1-3, Smr14A1-2 y Smr14B, identificados en *S. meliloti* en este trabajo. Aunque las ventajas eco-fisiológicas de estas reiteraciones únicamente se entienden en determinados casos, se pueden anticipar diversos escenarios funcionales para los *trans*-sRNA parálogos. Primero, la presencia de varias copias del mismo sRNA en un mismo genoma puede ser redundante para amplificar la respuesta ante variaciones sutiles del estímulo

externo. Es el caso de los Qrr de *V. cholerae*, donde la acumulación de un único sRNA Qrr es suficiente para reprimir el *quorum sensing*, pero es necesaria la delección de las cuatro copias del mismo para activar constitutivamente esta respuesta (Lenz *et al.*, 2004). En este caso los perfiles de expresión de todos los parálogos son idénticos, posibilitando así la compensación de función en caso de pérdida de uno o varios de ellos (Svenningsen *et al.*, 2009).

En segundo lugar, la repetición de sRNAs puede resultar en un efecto aditivo, como es el caso de los reguladores Qrr identificados en *V. harveyi* (Tu and Bassler, 2007). En esta bacteria, los promotores de los cinco genes *qrr* están diferencialmente regulados, lo que sugiere que responden ante diferentes indicadores metabólicos para integrar diversas señales medioambientales, de forma que la delección de cada sRNA Qrr afecta en diferente medida el *quorum sensing*. Es sorprendente, en cualquier caso, que a pesar de ser componentes de sistemas de *quorum sensing* idénticos, en especies muy relacionadas del género *Vibrio*, los sRNAs Qrr operan de acuerdo a mecanismos diferentes en cada bacteria. Aunque se desconoce la razón de esto, este hecho ilustra de nuevo la versatilidad y variedad de matices de la ribo-regulación y sus mecanismos.

Finalmente, una tercera posibilidad es que sRNAs duplicados o redundantes actúen de forma independiente uno de otro. Esto puede ocurrir de muchas maneras y lo más probable es que cada sRNA regule conjuntos diferentes, más o menos solapados, de mRNAs. No obstante, el caso más sorprendente de divergencia en los mecanismos de acción de dos *trans*-sRNAs similares en secuencia nucleotídica y estructura se ha descrito en *E. coli* para GlmY y GlmZ (Urban and Vogel, 2008). Estos dos sRNAs actúan de forma jerárquica en la misma cascada de regulación utilizando mecanismos completamente diferentes para activar la traducción del mensajero del gen *glmS*, que codifica un enzima esencial del metabolismo de amino-azúcares. Como ya se mencionó en la introducción general GlmZ es un ejemplo de *trans*-sRNA que activa de forma directa la traducción de su mRNA diana, GlmS, por apareamiento de bases en la región 5'-UTR de éste que elimina horquillas que bloquean el RBS. La acumulación intracelular de GlmZ depende a su vez de la expresión de GlmY, que actúa por encima de éste en esta cascada contribuyendo indirectamente a la activación de la traducción de GlmS mediante el secuestro de YhbJ, una proteína de función desconocida que inactiva GlmZ.

### III-1.2. MECANISMOS DE INTERACCIÓN TRANS-sRNA-mRNA: ELEMENTOS IMPLICADOS

#### III-1.2.1. ELEMENTOS DEL RNA IMPLICADOS EN EL APAREAMIENTO DE BASES

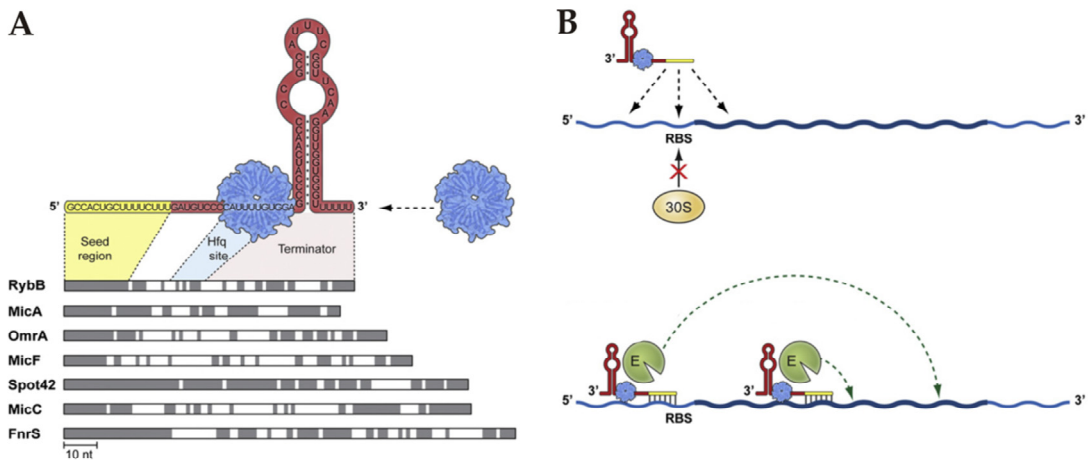
Los primeros estudios de sRNAs antisentido codificados en plásmidos, fagos o transposones establecieron que determinados elementos estructurales y las interacciones en varias etapas son cruciales en el mecanismo de apareamiento de bases en los RNAs bacterianos (Wagner *et al.*, 2002; Brantl, 2007). A pesar de una complementariedad perfecta con regiones de cientos de nucleótidos o más en sus mRNA diana, los RNA antisentido reconocen éstos mediante contactos rápidos y de alta afinidad que implican unos pocos nucleótidos expuestos en los bucles de las horquillas estructurales del ribo-regulador, el mRNA o ambos. Tras este contacto inicial se establecen nuevos apareamientos que, no obstante, requieren con frecuencia reordenamientos estructurales en el RNA (Han *et al.*, 2010).

Este mecanismo de apareamiento de bases se ha demostrado también para algunos *trans*-sRNAs de función conocida, como RNA III, RsaE y SprD de *S. aureus*. Estos ribo-reguladores reconocen normalmente secuencias ricas en citosinas que son accesibles en las regiones desapareadas de los bucles estructurales (Geissmann *et al.*, 2009; Bohn *et al.*, 2010; Chabelskaya *et al.*, 2010).

Sin embargo, en muchos otros *trans*-sRNAs que actúan mediante apareamiento de bases, la región requerida para la interacción con el mRNA está en tramos de cadena sencilla (Peer and Margalit, 2011). De hecho, estos sRNAs parecen modulares en enterobacterias (Fig. III-1A). En ellos se identifica un extremo 3' estructurado seguido de una cola poli-U, que promueve la terminación de la transcripción independiente de Rho y protege al sRNA de las endonucleasas-3'; un segundo dominio de unión a Hfq y un tercer tramo, de nucleótidos desapareados y generalmente muy conservados, en la región 5' del transcrito, que sería el utilizado para la interacción con el mRNA diana y, por tanto, el funcionalmente esencial. De hecho experimentos con quimeras de ribo-reguladores generados por fusión de estas regiones 5' de los sRNAs de *Salmonella* RybB o MicC a andamiajes estructurales de sRNAs no relacionados, revelaron que no se altera la capacidad de estos sRNAs para el reconocimiento de sus mRNA diana respectivos (Pfeiffer *et al.*, 2009; Papenfort *et al.*, 2010). Sin embargo, la longitud óptima y la composición nucleotídica de los elementos de interacción con la diana en los *trans*-sRNAs bacterianos no han sido totalmente definidas. En algunos casos (e.g. sRNAs SgrS y RybB) interacciones de 6 o 7 nucleótidos con el mRNA son suficientes para una ribo-regulación efectiva, pero es

evidente que estos requerimientos pueden variar entre las distintas interacciones sRNA-mRNA.

Por lo que se refiere al mRNA diana, los estudios iniciales sugieren que la interacción con el *trans*-sRNA ocurre principalmente en regiones que solapan completamente con el RBS o son inmediatamente adyacentes a este. Sin embargo, otros estudios más recientes en *E. coli* y *Salmonella* han revelado que algunos sRNAs puede interaccionar hasta 70 nt por encima o 15 por abajo del codón de iniciación AUG afectando también positiva o negativamente la traducción (Bouvier *et al.*, 2008; Frohlich and Vogel, 2009). Los sRNAs que afectan preferentemente la estabilidad del mensajero pueden tener en éste otros sitios de interacción, incluyendo regiones dentro de la propia secuencia codificante (Pfeiffer *et al.*, 2009) (Fig. III-1B).



**Figura III-1. Propiedades generales de los *trans*-sRNAs que actúan mediante apareamiento de bases.** A) Diagrama que muestra la estructura modular de los sRNAs y su relación con Hfq. Las regiones conservadas de los sRNAs enterobacterianos se muestran en gris, frecuentemente relacionadas con la zona de apareamiento. B) Esquema que representa los posibles sitios de unión en la región 5'-UTR del mRNA de los sRNAs que bloquean al ribosoma (parte superior) y representación de las vías de apareamiento que pueden desencadenar la degradación por RNasas (E, RNasaE) (parte inferior). Tomado de Storz *et al.*, 2011.

El impacto de las estructuras del sRNA, de los apareamientos no canónicos y de las secuencias situadas fuera de la zona de complementariedad sRNA-mRNA en la eficiencia de la ribo-regulación no ha sido estudiado con detalle todavía. No obstante, es previsible que también sean elementos que contribuyan a la identificación por parte del sRNA de sus mRNA diana específicos, que usualmente coexisten en la célula con otros miles de transcritos no ribo-regulados.

### III-1.2.2. PAPEL DE LAS CHAPERONAS DE RNA

La actividad de los *trans*-sRNAs que operan mediante apareamiento de bases depende de la presencia de la chaperona de RNA Hfq en bacterias Gram-negativas. (Vogel and Luisi, 2011). Sin embargo, los detalles de cómo Hfq se une a los sRNAs y afecta la ribo-regulación son escasamente conocidos.

Algunos experimentos *in vitro* sugieren que Hfq generalmente se une a regiones de cadena sencilla ricas en A/U, frecuentemente localizadas en las cercanías de una horquilla estructural, como reminiscencia de los sitios de unión de proteínas Sm y LSm eucariotas. Sin embargo, la predicción de motivos específicos de unión a Hfq en sRNAs es poco fiable debido a la falta de conservación y la ubicuidad de las secuencias ricas en A/U. Además, otras evidencias indican que la cola poli-U de los extremos 3' de los sRNAs también puede ser reconocida por esta chaperona (Otaka *et al.*, 2011; Sauer and Weichenrieder, 2011). Estudios estructurales y genéticos sugieren que las caras proximal y distal del anillo hexamérico de Hfq expuestas a posibles interacciones podrían establecer contactos simultáneos con varias moléculas de RNA. Sin embargo, se desconoce si son ambas caras las que unen el complejo sRNA-mRNA o si cada molécula se asocia a una cara de la proteína específicamente, y los posibles sitios de contacto con el mRNA no están tampoco documentados. No obstante, todas las evidencias indican que Hfq facilita la formación del dúplex sRNA-mRNA incrementando la cinética de anillamiento de los nucleótidos complementarios o promoviendo los reordenamientos estructurales necesarios en alguna de las moléculas (Fender *et al.*, 2010; Hopkins *et al.*, 2011; Hwang *et al.*, 2011; Soper *et al.*, 2010). Además de facilitar el apareamiento de bases, Hfq parece modular los niveles intracelulares del sRNA (Aiba, 2007; Brennan and Link, 2007; Valentin-Hansen *et al.*, 2004). La mayoría de los sRNAs expresados en *E. coli* son menos estables en ausencia de Hfq, posiblemente porque la proteína los protege de la degradación mientras no ocurre la interacción con el mRNA.

La contribución de Hfq a la regulación génica post-transcripcional mediada por *trans*-sRNAs ha sido casi exclusivamente estudiada en *E. coli* y enterobacterias filogenéticamente relacionadas como *Salmonella* fundamentalmente. Más recientemente ha sido también documentada la participación en la ribo-regulación de esta chaperona de RNA en otras bacterias Gram-negativas como la espiroqueta *Borrelia burgdorferi* (Lybecker *et al.*, 2010) e incluso en alguna Gram-positiva como *Listeria monocytogenes*, aunque ambas expresan proteínas Hfq muy divergentes de las de enterobacterias (Nielsen *et al.*, 2010). Curiosamente, no es tampoco infrecuente encontrar bacterias como *Burkholderia cenocepacia* (Ramos *et al.*, 2011) con múltiples

copias en su genoma del gen que codifica Hfq, lo que sugiere una posible diversificación funcional en estos casos.

Obviamente, existen otras proteínas bacterianas que pueden contribuir a la función de los sRNAs que actúan mediante apareamiento de bases. Éstas podrían actuar como chaperonas de RNA, además de Hfq o en su ausencia, en especies en las que no existen homólogos de esta proteína. De esta forma, se ha sugerido muy recientemente un papel como chaperona de RNA para la proteína de *E. coli* ProQ, que posee un dominio C-terminal que comparte características estructurales con Hfq (Chaulk *et al.*, 2011). Otras proteínas sin homología alguna con Hfq como YbeI, muy conservada en bacterias y con analogías estructurales a algún dominio de la proteína Argonauta de eucariotas, tiene un impacto en la expresión génica en *S. meliloti* similar al de Hfq, por lo que su función como RNA chaperona debe ser estudiada (Pandey *et al.*, 2011). En cualquier caso, la necesidad de una chaperona de RNA en este proceso no tiene por qué ser universal. Por ejemplo, la represión de OmpA por el RNA Vrr en *V. cholerae* es menos severa pero sigue ocurriendo en ausencia de Hfq. Quizá podría especularse que la dependencia de Hfq sería tanto menor cuanto más extensa y perfecta fuera la complementariedad entre sRNA y mRNA y/o más alta fuera la concentración intracelular del ribo-regulador.

### III-1.3. ESTRATEGIAS PARA LA IDENTIFICACIÓN DEL(LOS) mRNA DIANA

La identidad del(los) mRNAs regulado(s) directamente por un *trans*-sRNA es quizá el dato más informativo sobre la función de éste. El problema de la identificación del mensajero diana de un sRNA subyace a la propia naturaleza de su mecanismo de acción de forma que el hecho de que la complementariedad entre ambos en las zonas de interacción sea discontinua, con la participación frecuentemente de elementos atípicos (apareamientos no canónicos) e interacciones terciarias, dificulta enormemente esta tarea.

Dada la complejidad de este tipo de interacciones, la identificación de la diana pasa por la combinación de varias estrategias, a saber:

- **Aproximaciones bioinformáticas.** La disponibilidad de la secuencia genómica del organismo en cuestión es requisito indispensable para poder utilizar una aproximación bioinformática en la búsqueda de la diana. Cada vez van apareciendo nuevos algoritmos que predicen la interacción entre sRNA y mRNA, añadiendo parámetros físicos y energéticos relevantes en ésta. Existen programas como IntaRNA (Busch *et al.*, 2008), RNAplex (Tafer and Hofacker, 2008) o TargetRNA

(Tjaden *et al.*, 2006) que generan bancos de candidatos utilizando algoritmos basados en la complementariedad de bases, en la accesibilidad en sRNA y mRNA de las posibles secuencias complementarias y en estructuras secundarias favorecidas por la interacción. Pero dado el desconocimiento de otros elementos que intervienen en las interacciones cualquier aproximación *in silico* al problema suele generar grandes listas de candidatos que, no obstante, son un punto de partida útil para las subsiguientes verificaciones experimentales.

- **Tecnologías de alto rendimiento.** Están documentados en la bibliografía dos diseños experimentales que involucran estas tecnologías (Vögel and Wagner, 2007): *i*) sobreexpresión del sRNA biotinilado para la retención en columna de biotina de los complejos sRNA-mRNA y la secuenciación masiva (RNA-Seq) de la sub-población de mRNAs; *ii*) comparación de los perfiles transcriptómicos (*microarrays* o RNA-Seq) y/o proteómicos de las cepas silvestres o que no expresan el sRNA en estudio (mutantes por delección) y las que sobreexpresan de forma inducida o constitutiva el sRNA, para la identificación posterior de los mRNA diana entre los transcritos/proteínas diferencialmente regulados.

En cualquier caso, las interacciones sRNA-mRNA predichas *in silico* y las inferidas por la utilización de tecnologías de alto rendimiento según las estrategias descritas necesitan de su validación mediante técnicas genético-moleculares independientes. Existen multitud de ensayos *in vitro* para testar la formación de los híbridos sRNA-mRNA que operan durante la ribo-regulación. Sin embargo, dado que estos sistemas son artificiales y en ellos no se tienen en cuenta muchos factores que afectan a esta interacción (e.g. la asistencia de proteínas como Hfq), la significación biológica de sus resultados se pone con frecuencia en entredicho. Por ello se han diseñado también sistemas *in vivo* para la verificación de apareamientos particulares y su efecto sobre la traducción del mRNA (Urban and Vogel, 2007). Se trata de ensayos genéticos que consisten en la co-expresión en la célula, desde plásmidos independientes compatibles, del sRNA y de una fusión traduccional de la región 5' del supuesto mRNA diana a un gen reportero que suele ser GFP, evaluando la fluorescencia de la fusión en presencia y ausencia (control) del sRNA.

Entre los sRNAs predichos *in silico* y verificados experimentalmente en el Capítulo 1, siete no han sido descritos previamente en bacterias y son catalogables como *trans*-sRNAs. En este Capítulo se presentan algunos abordajes *in silico*, genéticos y genómicos para una primera aproximación a la función de tres de ellos; Smr15C1/15C2, dos sRNAs de secuencia y estructura homólogas codificados por



genes contiguos en la misma RIG de *S. meliloti* 1021, y Smr7C, un sRNA cuyo gen está presente en copia única en el genoma de esta bacteria.

## III-2. DISEÑO EXPERIMENTAL

### III-2.1. ESTRUCTURA Y DISTRIBUCIÓN DE LOS sRNAs Smr7C, Smr15C1 Y Smr15C2

#### III-2.1.1. MAPEO DE EXTREMOS 3' BASADO EN LA CIRCULARIZACIÓN DE MOLÉCULAS DE RNA

La determinación experimental de los extremos 3' de los transcritos Smr7C, Smr15C1 y Smr15C2 se llevó a cabo mediante el método de Vogel and Hess, 2001. Se tomaron 8  $\mu$ g de RNA de bacterias de *S. meliloti* 1021 crecidas en TY hasta D. O. 0,6 (fase logarítmica) o 2,4 (fase estacionaria), y se trataron con 10 U de TAP (Epicentre) durante 30 min a 37°C, para convertir los grupos trifosfato de los transcritos primarios en extremos monofosfato. Inmediatamente después de la extracción orgánica, el RNA se trató con 40 U de T4 RNA ligasa (NEB) a 17°C durante toda la noche, para circularizar los transcritos. Después de una nueva extracción orgánica y precipitación con EtOH, 2  $\mu$ g del RNA auto-ligado se convirtieron en cDNA utilizando 100 pmol de *random* hexámeros y 200 U de SuperScript II (Invitrogen) en una reacción de 20  $\mu$ l y a 42°C. Después de la inactivación del enzima (5 min a 85°C), las muestras se trataron con 1 U de RNasaH (Roche), 30 min a 37°C. Se tomó un volumen de 1  $\mu$ l para amplificar con una Taq polimerasa y una pareja de cebadores diseñados para amplificar la región de unión de los extremos 5' y 3' de cada transcrito Smr (5-7Ccir/3-7Ccir, para Smr7C; 5-15C1cir/3-15C1cir, para Smr15C1; 5-15C2cir/3-15C2cir, para Smr15C2, ver apéndice capítulo III). Los productos de PCR se separaron electroforéticamente en geles de agarosa al 3 % y los fragmentos del tamaño esperado se clonaron en pGEM®-T easy (Promega) y se secuenciaron con los oligonucleótidos universales T7 y SP6.

#### III-2.1.2. PREDICCIÓN Y REPRESENTACIÓN DE ESTRUCTURAS SECUNDARIAS DE RNAs

Para la predicción de estructuras secundarias de las secuencias de los sRNAs determinadas experimentalmente en este trabajo se utilizó el algoritmo **RNAfold** (Gruber *et al.*, 2008; <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). Esta herramienta permite plegar estructuras de RNA en base a parámetros energéticos y representar gráficamente estas estructuras. Para la mayoría de los casos se utilizaron los parámetros por defecto. Para la representación gráfica se utilizó además la

herramienta **VARNA** (Darty *et al.*, 2009), que permite el diseño y la edición de estas estructuras secundarias.

### III-2.1.3. CONSERVACIÓN EN BACTERIAS: MODELOS DE CO-VARIANZA

La búsqueda de homólogos de los sRNAs en estudio en otros grupos bacterianos se basó en modelos de co-varianza generados para cada familia de sRNAs (modelos Infernal v. 10.0; <http://www.sanger.ac.uk/Software/Rfam>). Este análisis se realizó en colaboración con la Dra. Coral del Val (Escuela Técnica Superior de Ingenieros Informáticos, Universidad de Granada). La generación de estos modelos se detalla en el apartado I-2.1.2. del capítulo I de esta tesis doctoral. Los candidatos predichos por Infernal en cada genoma bacteriano fueron asignados a una familia dada si su valor de  $E$  fue  $10^{-3}$  o inferior. Los miembros de cada familia fueron agrupados mediante WPGMA basado en las distancias de los alineamientos y el número óptimo de grupos en cada árbol se calculó finalmente mediante el programa **RNAclust** ([www.bioinf.uni-leipzig.de/~kristin/Software/RNAclust/](http://www.bioinf.uni-leipzig.de/~kristin/Software/RNAclust/)). El formato Stockholm se utilizó para representar el alineamiento correspondiente a cada familia.

## III-2.2. ESTUDIO DE LA EXPRESIÓN DIFERENCIAL DE LOS sRNAs Smr

### III-2.2.1. ANÁLISIS DE REGIONES PROMOTORAS

Para la predicción de las cajas promotoras correspondientes a la subunidad  $\sigma^{70}$  de la RNA polimerasa se utilizaron los consensos de las cajas -35/-20 CTTGAC-n17-CTATAT, que se ha demostrado que se encuentran ampliamente conservadas en muchos géneros del grupo de las  $\alpha$ -proteobacteria (MacLellan *et al.*, 2006).

Para identificar sitios de unión de otros factores de transcripción conocidos se utilizó como referencia el archivo FASTA con las secuencias depositadas en RegPredict (<http://regpredict.lbl.gov/regpredict/help.html>; Novichkov *et al.*, 2010), y las matrices (Position Weight Matrices; PSWM) de RegulonDB (<http://regulondb.ccg.unam.mx>; Gama-Castro *et al.*, 2011). Se generaron matrices para cada factor de transcripción desde las secuencias de RegPredict utilizando el programa Consensus/Patser, eligiendo la mejor matriz final para motivos con longitudes comprendidas entre 14-30 pb y con un valor de  $E$  inferior a  $10^{-10}$  para cada matriz, (ver "Thresholded consensus" en <http://gps-tools2.its.yale.edu>). Además, se realizó la búsqueda de motivos conservados desconocidos utilizando [http://meme.sdsc.edu/meme4\\_6\\_1/intro.html](http://meme.sdsc.edu/meme4_6_1/intro.html) (Bailey and Elkan, 1994).

### III-2.2.2. CONDICIONES DE CULTIVO

Para la caracterización de la expresión de los genes *smr15C1*, *smr15C2* y *smr7C* en *S. meliloti* 1021 se eligieron condiciones que imitaban estreses tanto abióticos como bióticos. Para cada serie de experimentos se hicieron preinóculos de las bacterias en medio completo TY sin estrés que luego se inocularon en las condiciones de ensayo. Estas condiciones fueron elegidas en base a criterios bibliográficos y experimentales, a saber:

**1. Expresión en fase de crecimiento:** para establecer de forma rigurosa los puntos de recogida de muestras a lo largo del crecimiento de *S. meliloti* en medio completo TY, se realizó una curva del mismo determinando la densidad óptica a 600 nm y la cantidad de viables (ufc/ml) en intervalos de 2 h durante un tiempo de 60 h post-inoculación, en cultivos aireados y en agitación continua. La expresión de los sRNAs se determinó en cultivos de D.O. 0,3, 0,6 (fase logarítmica), 1,2 (fase estacionaria temprana) y 2,4 (fase estacionaria tardía) (Fig. III-2).

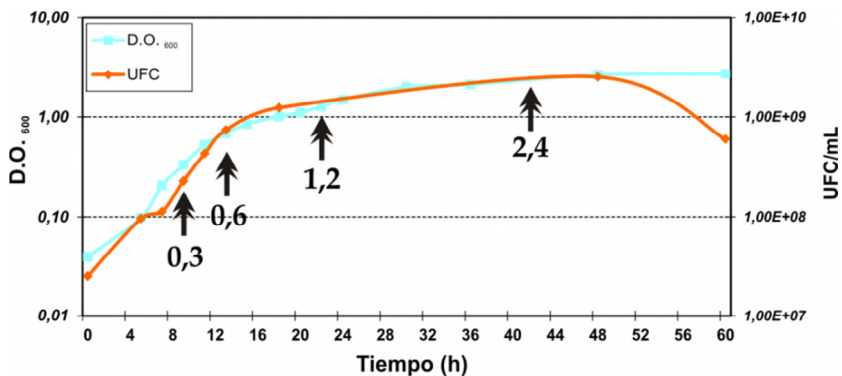


Fig. III-2. Monitorización del crecimiento de *S. meliloti* 1021. Variación de la D.O. y del número de viables (ufc/ml) a lo largo del tiempo en cultivos de 1021 en TY crecidos a 28 °C. Los resultados que se muestran son media de dos réplicas biológicas (inóculos) con tres repeticiones cada una. Las flechas dobles indican los puntos de recogida de muestras para la extracción de RNA total en las cinéticas de expresión de los genes *smr*, indicando la D.O. a la que se recogieron dichos cultivos.

**2. Estrés salino.** Se utilizaron dos condiciones: i) salinidad moderada, compatible con una supervivencia mayoritaria de plantas de *M. sativa* durante el establecimiento de la simbiosis (Fernández-López, datos experimentales) que fue imitada por adición de 50 mM de NaCl al medio completo TY, y ii) shock salino, aplicado por adición de 400 mM de NaCl a cultivos bacterianos en fase logarítmica (D.O. 0,6) en medio TY e incubación de las bacterias durante 1 h adicional. Esta concentración de sal y el tiempo de incubación de las bacterias en su presencia son las condiciones a las que se

han observado un mayor número de cambios en el transcriptoma de *S. meliloti* en respuesta al estrés salino (Domínguez-Ferreras *et al.*, 2006).

**3. Estrés oxidativo.** Para imitar condiciones de estrés oxidativo se añadió 1 mM de  $H_2O_2$  a los cultivos en TY y D.O 0,6 dejando incubar 1 h más. Esta concentración de  $H_2O_2$  se puede considerar subletal y su efecto en la adaptación de las bacterias al estrés oxidativo fue determinado experimentalmente. Para ello se analizó la capacidad de supervivencia después de sometidas a un choque letal de 20 mM de  $H_2O_2$  de bacterias previamente cultivadas en presencia de 1 mM de  $H_2O_2$  en relación a bacterias control sin pretratar. Las bacterias pre-adaptadas presentaron menor mortalidad (58,06%) que las que no habían sido sometidas al pretratamiento (93,32%) en los primeros 15 min después del choque oxidativo. En tiempos sucesivos las bacterias recuperaron cierta capacidad de división, que fue también mayor en las crecidas previamente en presencia de 10 mM de  $H_2O_2$ .

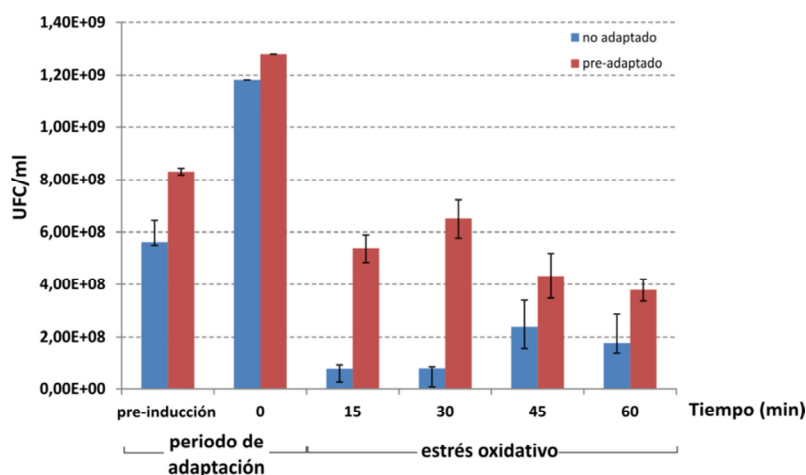


Fig. III-3. Adaptación al estrés oxidativo en *S. meliloti* 1021. Recuento de viables (ufc/ml) de dos cultivos independientes de *S. meliloti*, uno pre-adaptado 1 h con 1mM de  $H_2O_2$  y otro sin adaptación. Ambos se sometieron a shock oxidativo con 20 mM  $H_2O_2$  (t=0) y se monitorizó la supervivencia en intervalos de 15 min hasta 60 min. Cada resultado es media de dos réplicas biológicas con 3 repeticiones cada una.

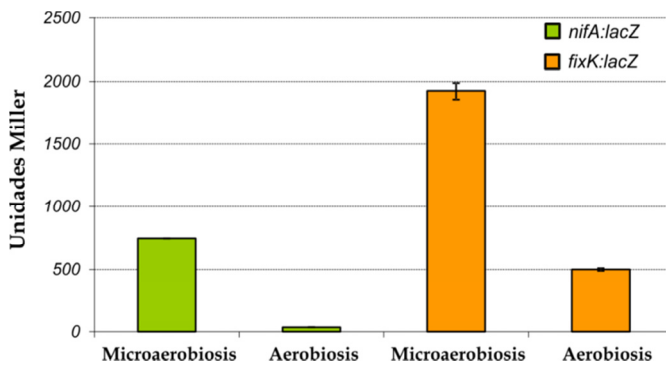
**4. Estrés por carencia de hierro.** La condición de carencia de hierro en el medio de cultivo se imitó mediante el crecimiento de *S. meliloti* en un cultivo en TY suplementado con 300  $\mu$ M de 2,2'-Dipiridil (Sigma). Esta concentración es capaz por sí misma de captar el hierro presente y generar condiciones de hambre de hierro, muy relevantes en la patogénesis de bacterias intracelulares de organismos superiores (Kitphati *et al.*, 2007).

**5. Estrés por etanol.** Dado que muchos sRNAs actúan a nivel de porinas de la membrana externa o transportadores, se generó un estrés de membrana con la

incubación de bacterias en medio TY suplementado con 2% de etanol (Valverde and Haas, 2008).

**6. Estrés por pH ácido.** Para generar esta condición se tomó un cultivo de *S. meliloti* en TY hasta que alcanzó la D.O. de 0,5-0,6, y seguidamente se lavaron las células y se sustituyó este medio por medio TY tamponado a pH 5,6 con 20 mM MES, incubándose en agitación durante 1 h.

**7. Baja tensión de oxígeno (microaerobiosis).** Para la generación de condiciones de baja tensión de oxígeno, los cultivos crecidos hasta D.O. 0,6 en medio TY se burbujearon con una mezcla de 2% Oxígeno/98% Argón durante 10 min, incubándose durante 4 h más en esa atmósfera. Para verificar que estas condiciones inducían marcadores de microaerobiosis conocidos se hicieron experimentos de actividad  $\beta$ -galactosidasa con bacterias de *S. meliloti* 2011 portadoras de fusiones transcripcionales de los promotores de los genes *fixK* y *nifA* (pMF57 y pCHK57, respectivamente; Ditta *et al.*, 1987; Foussard *et al.*, 1997) al gen reportero *lacZ*. Los resultados de estos experimentos corroboraron que, efectivamente, nuestras condiciones de microaerobiosis eran capaces de inducir la expresión de dichos genes (Fig. III-4).



**Fig. III-4. Actividad de los promotores de *nifA* y *fixK* en microaerobiosis.** Ensayo Miller de la actividad  $\beta$ -galactosidasa de las fusiones con *lacZ* de los promotores *nifA* y *fixK* en aerobiosis y microaerobiosis.

**8. Inducción por exudados radiculares y flavonoides.** Se partió de cultivos de *S. meliloti* 1021 en medio TY hasta D.O. 0,6 y las bacterias recogidas por centrifugación se lavaron 2 veces con el medio de crecimiento de plantas Rigaud y Puppo preparándose inóculos de  $10^6$  células que se sometieron a los siguientes tratamientos: *i*) incubación de las bacterias en medio Rigaud y Puppo durante 14 h (condición control), *ii*) incubación durante 7 días en presencia de plantas de *M. sativa* previamente crecidas en medio Rigaud y Puppo en tubos Falcon (condiciones según Fernández-López y López-Contreras, sin publicar), *iii*) incubación durante 14 h en medio Rigaud y Puppo suplementado con 20 mM de luteolina (Sigma).

**9.Endosimbiosis:** se recogieron nódulos maduros (30 días) de plantas de *M.sativa* cultivadas en jarras Leonard e inoculadas con *S. meliloti* 1021.

Las bacterias sometidas a los diferentes tratamientos se recogieron rápidamente para la extracción del RNA total que fue sometido a análisis Northern con sondas específicas para los sRNAs Smr15C1, Smr15C2 y Smr7C. Las señales de hibridación obtenidas se cuantificaron mediante el programa Quantity One® (Bio-rad) y se normalizaron respecto a la del RNA ribosómico 5S en cada condición. Una vez normalizadas, las intensidades se relativizaron respecto a la señal en fase logarítmica de crecimiento, considerada condición de referencia.

### III-2.3. GENERACIÓN DE MUTANTES EN LOS GENES *smr*

Como abordaje genético para el estudio de la función de los sRNAs Smr15C1, Smr15C2 y Smr7C se diseñaron dos tipos de mutantes; por delección y sobreexpresión de los correspondientes genes.

#### III-2.3.1. MUTANTES POR DELECCIÓN

Para la eliminación de la(s) copia(s) cromosómica(s) de los diferentes *loci smr* y su sustitución por un marcador de resistencia a Eritromicina (Ery<sup>r</sup>) se utilizó un método basado en la doble recombinación homóloga mediante el plásmido suicida pK18*mobsacB*. Las etapas para la generación de estos mutantes se esquematizan en la Fig. III-5.

Primeramente, se amplificaron utilizando como molde DNA total de *S. meliloti* 1021 las regiones genómicas que contienen los genes a deleccionar (~1 kb por encima y por debajo de cada gen) con parejas de cebadores que incorporan las dianas de restricción *Xba*I y *Sph*I en los extremos 5' y 3' de cada amplicón, respectivamente (i.e. 5-7C/3-7C, para Smr7C y 5-15C1/3-15C2, para la región genómica de Smr15C1/C2). Los fragmentos resultantes (2.113 pb para *smr7C* y 2.210 para *smr15C1-smr15C2*) se clonaron en pGEM-T Easy, dando lugar a los plásmidos pGEMgsmr7C y pGEMgsmr15C1/C2. Estas construcciones se utilizaron como molde para una nueva PCR utilizando oligonucleótidos divergentes flanqueantes de cada gen a deleccionar (i.e. 5-7C-i/3-7C-i, para *smr7C*; 5-15C1-i/3-15C1-i, para *smr15C1*, 5-15C2-i/3-15C2-i, para *smr15C2* y 5-15C1-i/3-C15C2-i, para la doble delección de *smr15C1* y *smr15C2*) que incorporaban en sus extremos 5' la diana de restricción *Kpn*I más 6 nt adicionales. Los amplicones resultantes (con la secuencia íntegra de los plásmidos y las dos regiones genómicas adyacentes al gen de interés) se sometieron a digestión con *Kpn*I y se auto-ligaron con 1 U de DNA ligasa, para generar los plásmidos pGEMΔ*smr7C*, pGEMΔ*smr15C1*, pGEMΔ*smr15C2* y pGEMΔ*smr15C1/C2*). En estas

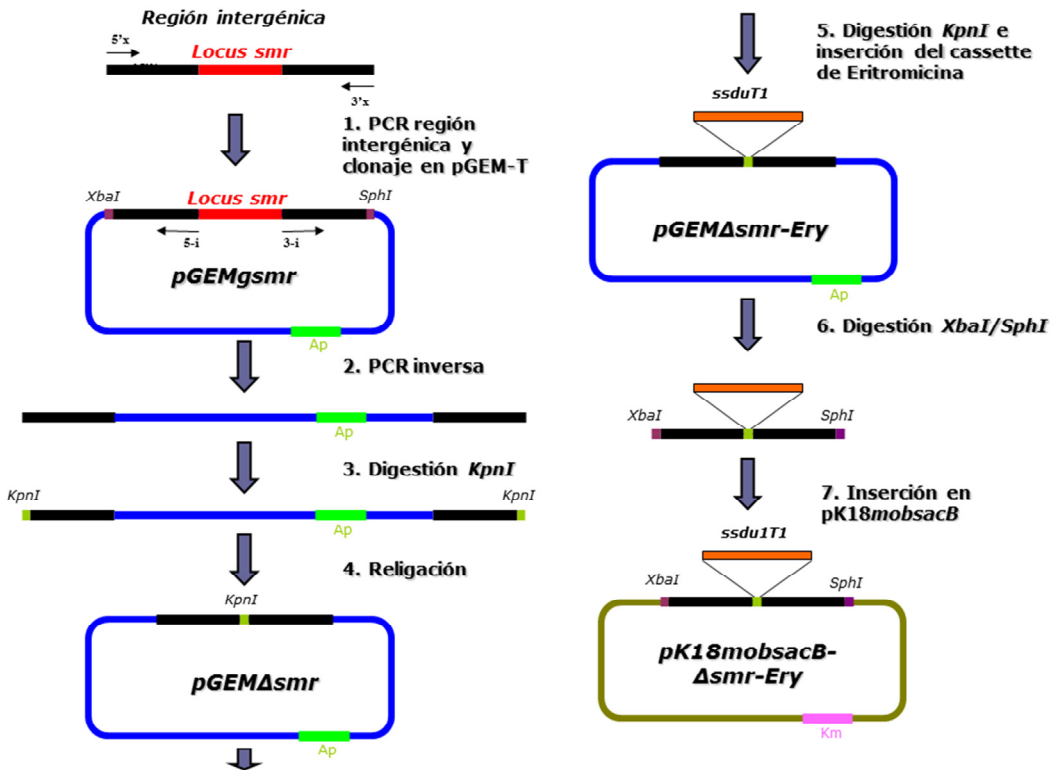


Fig. III-5. Estrategia seguida para la mutación de los genes *smr*.

construcciones se insertó el marcador de resistencia para seleccionar la mutación posteriormente. En nuestro caso se utilizó el **cassette de resistencia a eritromicina** *ssDUTI* diseñado en nuestro laboratorio (número de patente ES-201130782). Este *cassette*, de 195 nt, está integrado por el promotor sintético syn (Giacomini *et al.*, 1994) seguido de un motivo Shine-Dalgarno, la secuencia codificante del tetrapéptido que confiere resistencia a la eritromicina y el terminador de la transcripción T1. Este *cassette* se amplificó desde pGEMssDUTI con la pareja de cebadores 5-Ery-Kpn/3-Ery-Kpn que incorporan a ambos extremos la diana de restricción *KpnI*. El fragmento generado se clonó en el sitio *KpnI* de los plásmidos pGEMΔ*smr7C*, pGEMΔ*smr15C1*, pGEMΔ*smr15C2* y pGEMΔ*smr15C1/C2*, dando lugar a pGEMΔ*smr7C*-Ery, pGEMΔ*smr15C1*-Ery, pGEMΔ*smr15C2*-Ery y pGEMΔ*smr15C1/C2*-Ery, respectivamente. Finalmente, los insertos de estos plásmidos (con un tamaño de 2.134 pb para Δ*smr7C*, 2.204 pb para Δ*smr15C1*, 2.166 pb para Δ*smr15C2* y 1.908 pb para Δ*smr15C1*-Δ*smr15C2*), se recuperaron mediante restricción *XbaI*-*SphI* y se clonaron en pK18*mobsacB*, dando lugar a pK18Δ*smr7C*-Ery, pK18Δ*smr15C1*-Ery, pK18Δ*smr15C2*-Ery y pK18Δ*smr15C1/C2*-Ery. Todas las amplificaciones por PCR se realizaron con la polimerasa corregidora de errores Phusion® DNA polimerase (New England Biolabs,

NEB) y los amplicones siempre se verificaron por secuenciación. El intercambio alélico se llevó a cabo tal como se describe en Materiales y Métodos. Las distintas etapas del procedimiento se esquematizan en la Fig. III-5.

Una vez obtenidas las colonias que presumiblemente contenían cada mutación se hicieron verificaciones de las mismas mediante PCR de colonia (con las parejas de oligonucleótidos 7Csec/7Csec-i, 15C1sec/15C1sec-i, 15C2sec/15C2sec-i y 15C1sec/15C2sec-i) Southern blot para descartar posibles reordenamientos genómicos y Northern blot, para testar la ausencia de los codificados por los *loci smr* deletionados en cada caso.

### III-2.3.2. MUTANTES POR SOBREENPRESIÓN

Para la sobreexpresión constitutiva de los transcritos Smr se diseñaron dos tipos de construcciones, una basada en el promotor sintético *syn* (Giacomini *et al.*, 1994) en el vector pBBR1MCS2 y otra en el promotor de la lactosa (*lac*) en el plásmido pSRKKm (Ditta *et al.*, 1980) (Fig. III-6A y B).

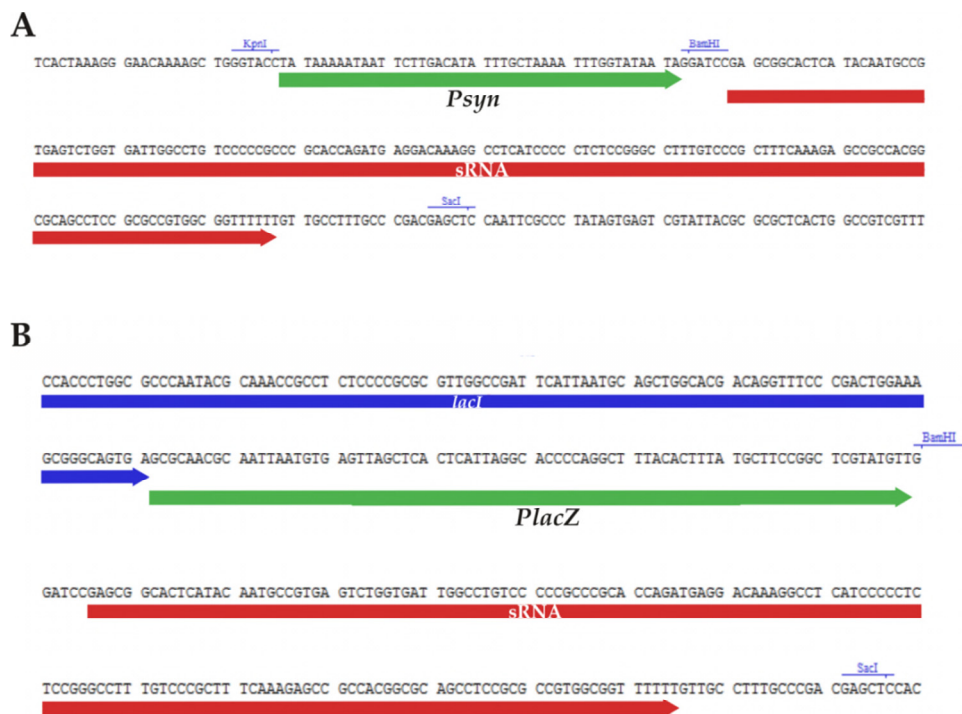


Fig. III-6. Promotores utilizados en la sobreexpresión de genes *smr*. **A)** Secuencia nucleotídica de las construcciones basadas en pBBSYN. **B)** Secuencia nucleotídica de las construcciones basadas en pSRKKm, en donde se indica la posición del gen *lacI* y el promotor *PlacZ*. La posición del *sRNA* en ambas construcciones también se indica, junto con los sitios de restricción más relevantes.



Para las construcciones en pBBR1MCS2 la caja promotora se generó mediante el anillamiento de los oligonucleótidos totalmente complementarios, synF y synR, que dejan en el híbrido extremos cohesivos con los sitios de restricción *KpnI* y *BamHI*. Este fragmento se insertó en pBBR1MCS2 como *KpnI-BamHI*, dando lugar a pBBSYN. La secuencia correspondiente a cada gen *smr* se amplificó desde DNA genómico de *S. meliloti* 1021 con *Taq* polimerasa, utilizando parejas de cebadores específicas que añaden los sitios *BamHI* y *SacI* a los extremos de los productos de PCR (*Smr7C* Fw/Rv, *Smr15C1* Fw/Rv y *Smr15C2* Fw/Rv). Los fragmentos generados se clonaron en pGEM-T Easy, se verificaron por secuenciación y se recuperaron mediante restricción *BamHI-SacI* para clonarse finalmente en el plásmido pBBSYN, dando lugar a pBBSYN*smr7C*, pBBSYN*smr15C1* y pBBSYN*smr15C2*. En todos los casos, la distancia entre el promotor y el inicio de transcripción del sRNA fue de 7 nt.

Para la expresión de los genes *smr* desde el promotor *lac* se utilizó el plásmido pSRKkm descrito por Khan *et al.*, 2008. En primer lugar se llevó a cabo la eliminación del sitio *BamHI* del *polilinker* de este plásmido mediante digestión con la enzima, sellado de los extremos 5' protuberantes mediante la adición de *Klenow* y religación con DNA ligasa, dando lugar a pSRK\*. La modificación se verificó mediante secuenciación con el cebador secSRK. Posteriormente, se realizó una PCR inversa sobre pSRK\* con los oligonucleótidos FwSRK/RvSRK que incorporan la diana *BamHI* en cada extremo del amplicón. RvSRK hibrida en el promotor de *lacZ* eliminando la secuencia correspondiente al operador al que se une el represor *LacI* y por tanto dotando al sistema de una expresión constitutiva. El producto de PCR se digirió con la enzima *BamHI* y se religó para generar el plásmido pSRK\_C. La secuencia de los sRNAs de interés se extrajo de sus respectivos clones en pGEM-T como fragmentos *BamHI-SacI* que se clonaron en pSRK\_C para dar lugar a pSRK*smr7C*, pSRK*smr15C1* y pSRK*smr15C2*. En todos los casos, las construcciones generadas se verificaron por secuenciación Sanger con el cebador secSRK.

Los plásmidos pBBR1MCS2 y pSRK\_C (controles) así como los derivados de los mismos que expresan los sRNAs *Smr* en estudio se movilizaron mediante cruces triparentales a *S. meliloti* 1021 y a los mutantes obtenidos por delección de cada gen *smr*.

### III-2.4. IDENTIFICACIÓN DE mRNAs DIANA.

#### III-2.4.1. PREDICCIÓN BIOINFORMÁTICA: INTARNA

Para la predicción bioinformática de las posibles dianas de los sRNAs en estudio en el genoma de *S. meliloti* 1021 se utilizó la interfaz web del programa IntaRNA

(Busch *et al.*, 2008; Richter *et al.*, 2010; <http://rna.informatik.uni-freiburg.de:8080/IntaRNA.jsp>). Para la generación de las listas priorizadas de candidatos IntaRNA combina esencialmente dos parámetros calculados *in silico*; la energía de hibridación sRNA-mRNA y la accesibilidad para el apareamiento de las secuencias complementarias en cada molécula. En nuestro análisis los *inputs* del programa fueron las secuencias completas de los sRNAs Smr7C, Smr15C1 y Smr15C2 determinadas experimentalmente en este trabajo y un fichero FASTA con todas las secuencias nucleotídicas comprendidas entre la posición -250 y +150 respecto al codón de inicio anotado para cada uno de los ORFs del genoma de *S. meliloti* 1021. Se asumió, por tanto, la interacción sRNA-mRNA en la 5'-UTR del mensajero diana, que opera mayoritariamente en la ribo-regulación en bacterias. Como parámetro adicional para la generación de las predicciones se impuso un apareamiento mínimo sRNA-mRNA de 8 nt consecutivos en ambas moléculas.

#### III-2.4.2. ENSAYO DE DOBLE PLÁSMIDO *IN VIVO*

La validación experimental *in vivo* de los mRNA diana candidatos para los transcritos Smr se llevó a cabo siguiendo el método propuesto en Urban and Vogel, 2007 utilizado con éxito en enterobacterias para este fin. El método está basado en la co-expresión desde plásmidos compatibles en la misma célula del sRNA y de una fusión traduccional de la región 5' del posible mRNA diana a GFP como sistema reportero.

Para ello se diseñaron dos tipos de plásmidos, uno de bajo número de copias (derivado de pJB3Tc19), que expresa la fusión traduccional de la región 5' (i.e. 5'-UTR más los primeros codones) del mRNA candidato con el extremo N-terminal de eGFP desde el promotor constitutivo *syn*, y un segundo tipo de plásmido que expresa el sRNA desde el promotor *lac* de pSRK\_C (Fig. III-7).

A su vez, los **plásmidos derivados de pJB3Tc19** fueron de dos tipos; uno diseñado para la expresión de regiones 5'-UTRs de mRNAs monocistrónicos o de los primeros en un operón (pR\_eGFP) y otro que imita una unidad transcripcional dicistrónica y posibilita la expresión de regiones 5'-UTRs internas en mRNAs policistrónicos (pRO\_eGFP).

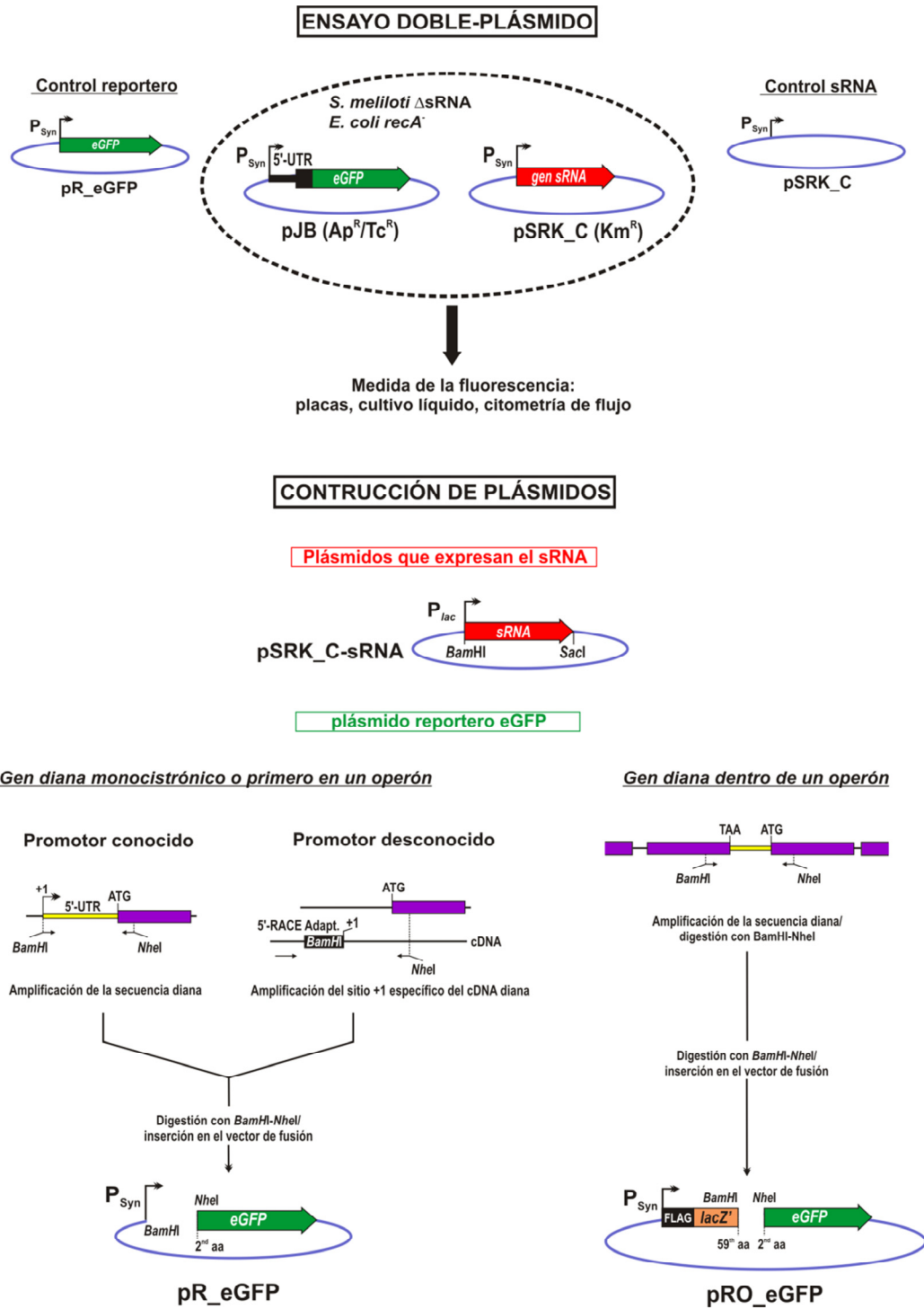


Fig. III-7. Construcciones genéticas para la verificación de las interacciones sRNA-mRNA *in vivo* en *S. meliloti*

Para la generación de **pR\_eGFP**, se amplificó la unidad Psyn-eGFP desde el plásmido pBBsyn-eGFP (utilizado como control de localización de Hfq en el capítulo 2) con la pareja de cebadores 5SynHindIII/3'-GFP EcoRI, que añaden los sitios HindIII y EcoRI a los extremos del amplicón. Este fragmento se clonó en pGEM-T Easy, dando lugar a pGEMPsyn-eGFP, que se verificó mediante secuenciación. Este fragmento se extrajo como HindIII-EcoRI y se insertó en pJB3Tc19 dando lugar al control de expresión pJB\_eGFP. Posteriormente, para eliminar la región 5'-UTR de la GFP y sustituirla por la de los mRNA candidatos a testar se realizó una PCR inversa desde pGEMPsyn-eGFP con los oligonucleótidos Syn-i (que añade un sitio BamHI) y GFP-i2 (que añade BamHI y NheI). El producto de PCR se digirió con BamHI y se autoligó, para generar pGEM\_GFP-ΔUTR. Este plásmido contiene el promotor syn como fragmento HindIII-BamHI, y la región codificante de eGFP flanqueada por los sitios NheI (que codifica el segundo y tercer aminoácido de la eGFP)/ EcoRI. Dado que la presencia de ondas de transcripción que provienen del plásmido podrían dar lugar a falsos resultados positivos, se introdujo un terminador de transcripción independiente de Rho (T1) delante del promotor syn como fragmento SacI-HindIII, generado por el anillamiento de los cebadores complementarios RhoIT\_1 y RhoIT\_2, dando lugar a pGEM\_GFP-ΔUTR-T. El módulo T1-Psyn-GFP se extrajo como fragmento SacI-EcoRI y se clonó en pJB3Tc19, generando pR-eGFP.

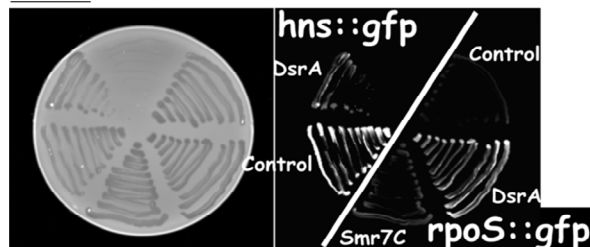
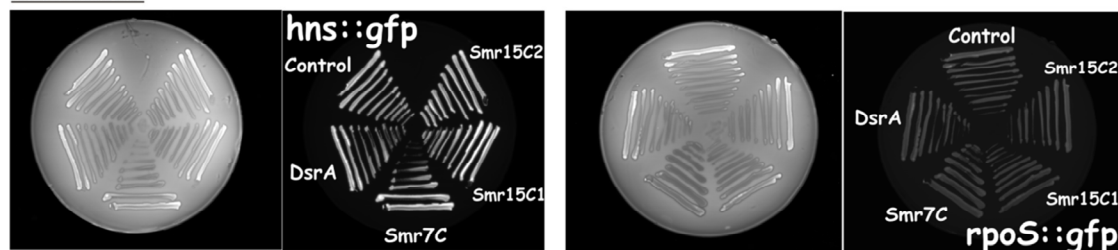
La generación de **pRO\_GFP** comenzó con la amplificación del fragmento FLAG-LacZ' desde *E. coli* DH5α con la pareja de cebadores FlacZ/LacZ-59. El primero es un oligonucleótido de 66 residuos que contiene un sitio BglII, el motivo Shine-Dalgarno para la traducción del péptido FLAG, toda la secuencia codificante de éste y el comienzo de lacZ (hasta el codón 59). El segundo cebador hibrida a partir del codón 59 hacia el ATG de lacZ y añade los sitios consecutivos BamHI y NheI al extremo 3' del fragmento. Todo este fragmento se insertó en el plásmido pGEM\_GFP-ΔUTR-T digerido con BamHI-NheI, dando lugar a pGEM\_GFP\_O-T. El fragmento SacI-EcoRI recuperado de este plásmido se clonó finalmente en pJB3Tc19 generando pRO\_eGFP.

Las regiones 5'-UTR de los mRNA candidatos se clonaron en el plásmido pR\_eGFP como fragmentos BamHI-NheI amplificados desde el sitio de inicio de transcripción de los mRNAs monocistrónicos candidatos cuya identidad en cada caso fue facilitada por la Dra. Anke Becker (Philipps Universität Marburg) de acuerdo a resultados no publicados de RNA-Seq de RNA total de *S. meliloti* 2011 obtenido en diversas condiciones biológicas. En pRO\_eGFP se clonaron fragmentos BamHI-NheI correspondientes a regiones entre ORFs en operones de forma que en cada caso se creó una fusión traduccional al extremo C-terminal del módulo FLAG-LacZ' y al N-terminal de eGFP (Fig. III-7, sección inferior derecha).

En estos ensayos los plásmidos que expresan los sRNAs de forma constitutiva fueron los derivados de pSRK\_C descritos en epígrafes anteriores.

Los ensayos fueron realizados en los mutantes de *S. meliloti* 1021 que no expresan los sRNAs en estudio (i.e. *S. meliloti*  $\Delta smr7C$  y *S. meliloti*  $\Delta smr15C1/C2$ ) a los que se transfirieron los derivados de pR\_eGFP/pRO\_eGFP y pSRK\_C correspondientes mediante cruces triparentales sucesivos. Como control negativo en todos los casos se utilizó el vector vacío pSRK\_C. Los dobles transformantes se crecieron en placas de TY y las bacterias se observaron en escáner de fluorescencia PharosXS (Bio-rad) a una longitud de onda de 530 nm.

Para validar el sistema se hicieron los correspondientes ensayos de doble plásmido testando las interacciones conocidas del sRNA DsrA con los mRNAs que codifican el regulador transcripcional asociado a nucleoide Hns y el factor  $\sigma^{38}$  o RpoS. Está bien documentado que la acumulación intracelular de DsrA inhibe la traducción de Hns mientras que, por el contrario, posibilita la de RpoS al eliminar horquillas inhibitorias en la región 5'-UTR de este mensajero (Urban and Vogel, 2007). Células de *E. coli* DH5 $\alpha$  y *S. meliloti* 1021 fueron co-transformadas independientemente con los plásmidos pR\_Hns o pR\_RpoS que expresan las fusiones traduccionales Hns::eGFP y RpoS::eGFP, respectivamente, y el plásmido pSRK\_DsrA que expresa constitutivamente el sRNA. Como controles negativos de los ensayos se utilizaron el plásmido vacío pSRK\_C y pSRK\_Smr7C, pSRK\_Smr15C1 y pSRK\_Smr15C2, que expresan los sRNAs heterólogos Smr7C, Smr15C1 y Smr15C2 de *S. meliloti*. Los resultados obtenidos (Fig. III-8) revelaron que, efectivamente, en células de *E. coli* la expresión de DsrA provoca una disminución de la fluorescencia provocada por la transcripción y traducción de la fusión Hns::eGFP, que es evidente en presencia de los plásmidos que no expresan el sRNA. Por el contrario, la fluorescencia de la fusión RpoS::eGFP es apenas perceptible en los ensayos control, confirmando el posible efecto inhibitorio de la región 5'-UTR de RpoS en la traducción, mientras que aumenta visiblemente cuando se co-expresa con DsrA. Cuando las mismas combinaciones de plásmidos se testaron en *S. meliloti* (Fig. III-8) no se observó ningún efecto de la presencia de DsrA sobre la fluorescencia basal de cada una de las fusiones, probablemente debido a que la interacción requiera factores del huésped. La expresión de los sRNAs Smr7C, Smr15C1 y Smr15C2 no alteró el patrón de fluorescencia en ningún caso.

*E. coli**S. meliloti*

**Fig. III-8. Validación del ensayo de doble plásmido *in vivo*.** Verificaciones en células de *E. coli* (superior) y *S. meliloti* (inferior) de la interacción del sRNA DsrA sobre las regiones 5'-UTR de sus mRNAs diana Hns y RpoS en fusiones traduccionales con la GFP. El control corresponde al plásmido pSRK\_C. Cada panel corresponde a la imagen en luz blanca (izquierda) y a la fluorescencia a 530 nm (derecha).

### III-3. RESULTADOS

Este capítulo de resultados está estructurado en dos apartados; en el primero se describe la caracterización funcional de los sRNAs Smr15C1 y Smr15C2 y en el segundo la de Smr7C, abordada, no obstante, mediante metodologías similares en todos los casos.

#### III-3.1. LA FAMILIA DE sRNAs $\alpha 15$ : CARACTERIZACIÓN DE Smr15C1 Y Smr15C2

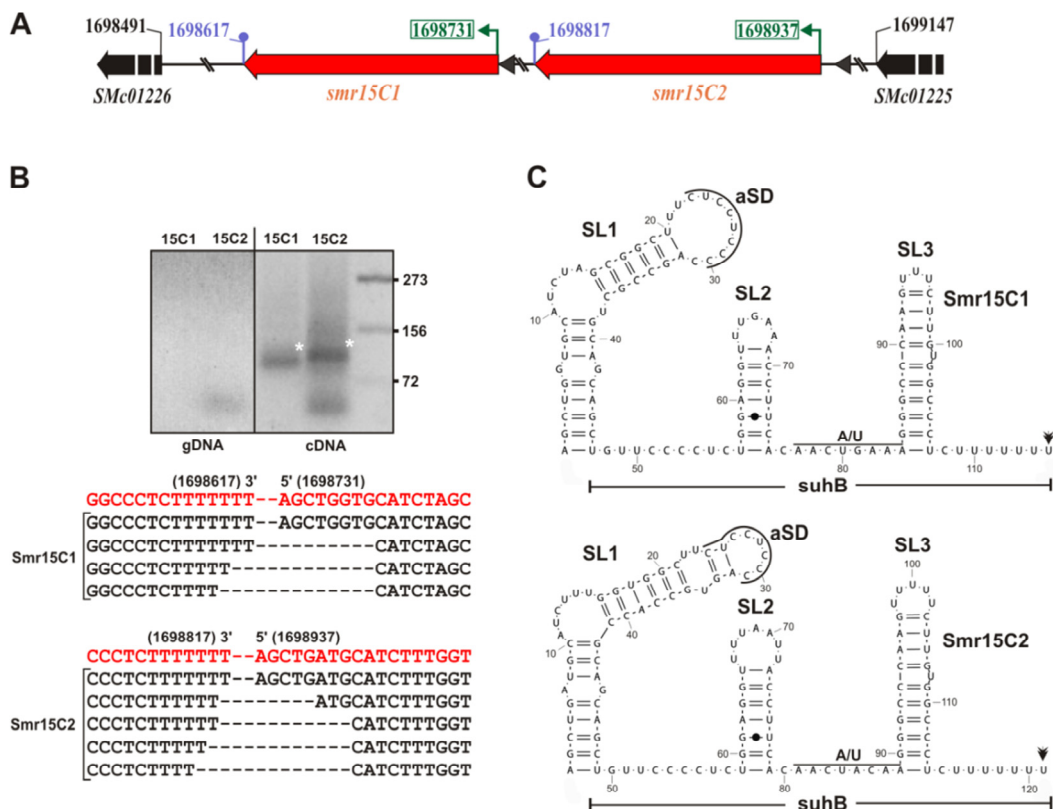
##### III-3.1.1. MAPEO DE LOS EXTREMOS 3' Y CARACTERIZACIÓN *IN SILICO* DE LOS sRNAs Smr15C1 Y Smr15C2

Como ya se ha descrito en el Capítulo 1, los genes *smr15C1* y *smc15C2* se encuentran localizados en tándem en la misma RIG en el genoma de *S. meliloti* 1021 (Fig. III-9A). Su transcripción ocurre en la misma dirección y la secuencia nucleotídica primaria de los transcritos Smr15C1 y Smr15C2 muestra una alta homología (84 % de identidad).

La poliadenilación de los extremos 3' de las moléculas de RNA es un proceso relativamente frecuente y bien documentado en bacterias, que también ocurre en los sRNAs, favoreciendo la degradación de éstos por las ribonucleasas celulares (Urban

and Vogel, 2008). Dado que los extremos 3' de Smr15C1 y Smr15C2 no estaban mapeados experimentalmente y las implicaciones de una posible poliadenilación de estos transcritos, se procedió a la determinación de los mismos. Para ello se circularizaron las moléculas de muestras de RNA total de *S. meliloti* 1021 extraído en fase logarítmica y en fase estacionaria de crecimiento (D.O<sub>600</sub> 0,6 y 2,4 respectivamente) antes de ser sometidas a transcripción reversa. La amplificación del cDNA circular obtenido en cada caso utilizando cebadores divergentes específicos para Smr15C1 (RNA fase estacionaria) y Smr15C2 (RNA fase logarítmica) generó bandas mayoritarias del tamaño esperado para ambos transcritos (115 pb y 121 pb, respectivamente) que fueron resueltas en geles de agarosa al 3% (Fig. III-9B). La amplificación de DNA genómico de *S. meliloti* 1021 con las mismas parejas de cebadores no generó bandas específicas del tamaño esperado. Ambos productos de PCR fueron clonados y la secuenciación de varios clones independientes en cada caso no reveló signos de poliadenilación en ninguno de los sRNAs, mapeando sus extremos 3' en cualquiera de las cuatro últimas Uridinas de los terminadores independientes de Rho predichos en ambos transcritos. En algunos casos estos extremos aparecieron asociados a moléculas más cortas de lo esperado, con extremos 5' diferentes a los determinados experimentalmente mediante RACE 5', que podrían corresponder a especies degradadas de ambos transcritos, en cualquier caso nunca detectadas en las hibridaciones Northern y, por tanto, minoritarias.

La simulación del plegamiento de las secuencias nucleotídicas de Smr15C1 y Smr15C2 con RNAfold predice en ambos casos una estructura secundaria con tres horquillas bien diferenciadas (Fig. III-9C): i) la primera horquilla (**SL1**), comprende un tallo largo que posee un bucle abierto terminal en Smr15C1 y otro, más pequeño, en Smr15C2. En ambos bucles es evidente el motivo anti-Shine Dalgarno (aSD) "CUCCUCCC"; ii) la segunda horquilla (**SL2**) es similar en los dos transcritos y de longitud menor que la anterior, dejando únicamente desapareados 5 nt en Smr15C1 y 7 en Smr15C2; iii) la última horquilla (**SL3**) corresponde en los dos transcritos al terminador de la transcripción independiente de Rho. Además de los bucles de cada horquilla las zonas desapareadas en ambas moléculas son tres. La comprendida entre SL1 y SL2 es de 9 nt en los dos transcritos con predominio de uridinas y citosinas. La localizada entre SL2 y el terminador de transcripción tiene una longitud de 8 nt en ambos casos y podría constituir un motivo rico en A/U de asociación con la chaperona Hfq. Finalmente, la última región desapareada en ambos transcritos es la cola poli-U característica de los terminadores independiente de Rho respectivos, que es otro motivo descrito de unión a Hfq en los sRNAs bacterianos (Storz *et al.*, 2011).



**Fig. III-9. Los sRNAs Smr15C1/C2 de *S. meliloti*.** **A)** RIG donde se localizan los genes *smr15C1/C2*, indicándose los genes flanqueantes y las coordenadas de los inicios de transcripción mapeados mediante RACE 5' (encuadrados en verde) y los extremos 3' (última U del terminador independiente de Rho). **B)** Gel RACE 3' de los productos amplificados sobre DNA genómico (izda.) y cDNA circular (derecha), indicando las bandas de tamaño esperado con asteriscos. Secuencia de los clones (inferior) secuenciados después del clonaje de los productos de PCR anteriores, indicando en rojo la anotación correspondiente al genoma de *S. meliloti*. **C)** Estructuras secundarias predichas por VARNA para los sRNAs Smr15C1/C2, indicando la fracción conservada de la familia *suhB* y la última U de la cola poli-U del terminador (flecha doble). Abreviaturas: SL, dominio o stem-loop; aSD, anti-Shine Dalgarno; A/U, posible sitio de unión a la chaperona Hfq.

Cuando la base de datos Rfam fue interrogada con el modelo estructural de Smr15C1 y Smr15C2 aparecieron sRNAs bacterianos con homología restringida a los motivos SL2 y SL3, que aparecen agrupados en la familia de sRNAs denominada *suhB* (modelo de co-varianza RF00519) (Fig. III-9C) (<http://rfam.sanger.ac.uk/family/RF00519>). Sin embargo, no se encontraron homólogos correspondientes a los transcritos completos en esta base de datos.



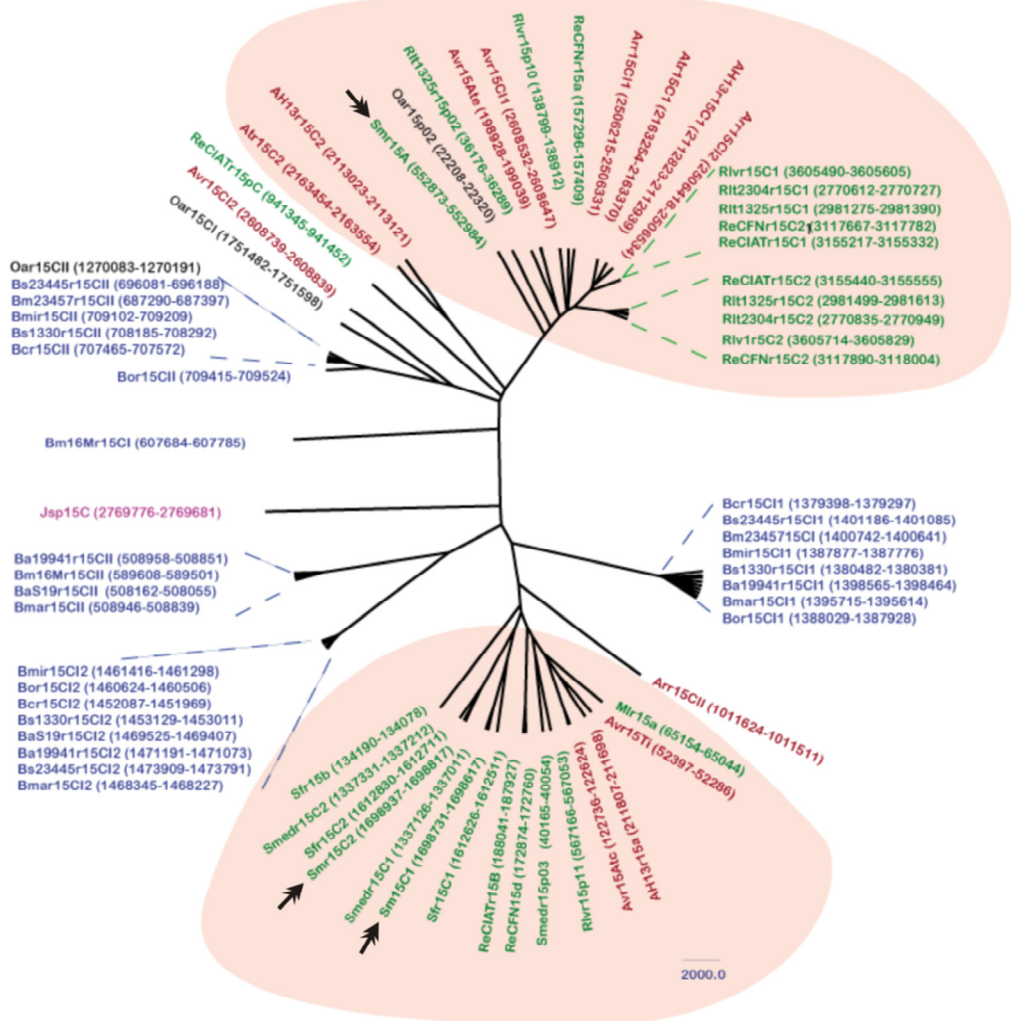
### **III-3.1.1.1. Conservación y distribución de Smr15C1 y Smr15C2 en el orden Rhizobiales**

La identificación de homólogos estructurales de Smr15C1 y Smr15C2 en otros grupos bacterianos se llevó a cabo mediante la utilización del modelo de co-varianza (MC) generado con Infernal 1.0 (Nawrocki *et al.*, 2009) a partir de alineamientos de secuencias con más del 78% de identidad (identificadas mediante BLASTN), que fueron curados manualmente. El alineamiento estructural (formato Stockholm) para la generación del MC que identifica a los miembros de esta familia de sRNAs, que hemos denominado  $\alpha 15$ , se muestra en la Fig. III-10. De él se deduce una estructura secundaria consenso para los sRNAs  $\alpha 15$  en la que son evidentes los motivos SL1, SL2 y SL3 predichos en los transcritos Smr15C1 y Smr15C2. Esta estructura parece sustentada por un alto grado de co-varianza, particularmente en los nucleótidos complementarios que forman el tallo de la horquilla SL1.

Como ya se describió en el Capítulo 1 este MC identificó una tercera copia de estos transcritos (Smr15A) entre las coordenadas A 552.873 y T 552.984 del plásmido simbiótico pSymA de *S. meliloti* 1021. La secuencia nucleotídica de Smr15A presenta una identidad del 70% con Smr15C1 y del 68% con Smr15C2, siendo la más diferente de las tres, aunque su estructura secundaria es muy similar a la de sus equivalentes cromosómicos. La expresión de Smr15A, sin embargo, fue casi indetectable en células de *S. meliloti* 1021 cultivadas en medio completo TY.

La utilización del modelo  $\alpha 15$  para el rastreo de genomas bacterianos en busca de nuevos miembros de la familia resultó en la identificación de 67 sRNAs con homología estructural a Smr15C1, Smr15C2 y Smr15A, todos ellos en especies del orden Rhizobiales filogenéticamente próximas a *S. meliloti* (i.e. especies de los géneros *Sinorhizobium*, *Rhizobium*, *Mesorhizobium*, *Agrobacterium* y *Brucella*). La nomenclatura propuesta para estos sRNAs es la inicial del género y la especie (añadiéndose el código de la cepa si hubiera más de una), seguida de la letra "r" (en alusión a RNA), del calificativo de la familia (15) y de la localización genómica (e.g. ReCIATr15C2, *R. etli* CIAT r familia 15 copia cromosómica 2). La relación completa de los miembros de la familia  $\alpha 15$  y sus coordenadas en los respectivos genomas puede consultarse en página web [http://en.wikipedia.org/wiki/User:Rczarg/%CE%B1r15\\_RNA](http://en.wikipedia.org/wiki/User:Rczarg/%CE%B1r15_RNA).





**Fig. III-11. Relaciones filogenéticas entre los sRNAs de la familia  $\alpha r15$ .** Arbol filogenético correspondiente a los miembros identificados en *S. meliloti* (marcados con flechas dobles) y predichos en el resto de organismos de la familia  $\alpha r15$ . Leyenda: *Smr15C1* = *S. meliloti* 1021 (NC\_003047), *Smr15C2* = *S. meliloti* 1021 plásmido pSyMA (NC\_003037), *Smr15A* = *S. meliloti* 1021 plásmido pSyMA (NC\_003037), *Smedr15C1* = *S. medicae* WSM419 cromosoma (NC\_009636), *Smedr15C2* = *S. medicae* WSM419 cromosoma (NC\_009636), *Smedr15p03* = *S. medicae* WSM419 plásmido pSMED03 (NC\_009622), *Sfr15C1* = *S. fredii* NGR234 cromosoma (NC\_012587), *Sfr15C2* = *S. fredii* NGR234 cromosoma (NC\_012587), *Sfr15b* = *S. fredii* NGR234 plásmido pNGR234b (NC\_012586), *Atr15C2* = *A. tumefaciens* str. C58 cromosoma circular (NC\_003062), *AH13r15C2* = *A. sp. H13-3* cromosoma (NC\_015183), *Atr13r15C1* = *A. sp. H13-3* cromosoma (NC\_015183), *AH13r15a* = *A. sp. H13-3* plásmido pAspH13-3a (NC\_015184), *ReCIATr15C2* = *R. etli* CIAT 652 (NC\_010994), *ReCIATr15C1* = *R. etli* CIAT 652 (NC\_010994), *ReCIATr15pC* = *R. etli* CIAT 652 plásmido pC (NC\_010997), *ReCIATr15B* = *R. etli* CIAT 652 plásmido pB (NC\_010996), *Arr15C1I* = *A. radiobacter* K84 cromosoma 1 (NC\_011985), *Arr15C1II* = *A. radiobacter* K84 cromosoma 2 (NC\_011983), *Rlt2304r15C2* = *R. leguminosarum* bv. *trifolii* WSM2304 cromosoma (NC\_011369), *Rlt2304r15C1* = *R. leguminosarum* bv. *trifolii* WSM2304 cromosoma (NC\_011369), *Avr15C2* = *A. vitis* S4 cromosoma 1 (NC\_011989), *Avr15C1* = *A. vitis* S4 cromosoma 2 (NC\_011989), *Avr15pAtc* = *A. vitis* S4 plásmido pAtS4c (NC\_011984), *Avr15Ate* = *A. vitis* S4 plásmido pAtS4e (NC\_011981), *Avr15Ti* = *A. vitis* S4 plásmido pTiS4 (NC\_011982), *Rlvr15C2* = *R. leguminosarum* bv. *viciae* 3841 (NC\_008380), *Rlvr15C1* = *R. leguminosarum* bv. *viciae* 3841 (NC\_008380), *Rlvr15p10* = *R. leguminosarum* bv. *viciae* 3841 plásmido pRL10 (NC\_008381), *Rlvr15p11* = *R. leguminosarum* bv. *viciae* 3841 plásmido pRL11 (NC\_008384), *Rlt1325r15C2* = *R. leguminosarum* bv. *trifolii* WSM1325 (NC\_012850), *Rlt1325r15C1* = *R. leguminosarum* bv. *trifolii*

WSM1325 (NC\_012850), Rlt1325r15p02 = *R. leguminosarum* bv. *trifolii* WSM1325 plásmido pR132502 (NC\_012858), ReCFNr15C2 = *R. etli* CFN 42 (NC\_007761), ReCFNr15C1 = *R. etli* CFN 42 (NC\_007761), ReCFNr15d = *R. etli* CFN 42 symbiotic plásmido p42d (NC\_004041), ReCFNr15a = *R. etli* CFN 42 plásmido p42a (NC\_007762), Mlr15a = MesoR. loti MAFF303099 plásmido pMLa (NC\_002679), Bcr15CII = *B. canis* ATCC 23365 cromosoma II (NC\_010104), Bcr15CI2 = *B. canis* ATCC 23365 cromosoma I (NC\_010103), Bcr15CI1 = *B. canis* ATCC 23365 cromosoma I (NC\_010103), Bs23445r15CI2 = *B. suis* ATCC 23445 cromosoma I (NC\_010169), Bs23445r15CI1 = *B. suis* ATCC 23445 cromosoma I (NC\_010169), Bs23445r15CII = *B. suis* ATCC 23445 cromosoma II (NC\_010167), Bm16MCI = *B. melitensis* bv. 1 str. 16M cromosoma I (NC\_003317), Bm16Mr15CII = *B. melitensis* bv. 1 str. 16M cromosoma II (NC\_003318), BaS19r15CII = *B. abortus* S19 cromosoma 2 (NC\_010740), BaS19r15CI2 = *B. abortus* S19 cromosoma 1 (NC\_010742), BaS19r15CI1 = *B. abortus* S19 cromosoma 1 (NC\_010742), Bm23457r15CII = *B. melitensis* ATCC 23457 cromosoma II (NC\_012442), Bm23457r15CI = *B. melitensis* ATCC 23457 cromosoma I (NC\_012441), Bs1330r15CII = *B. suis* 1330 cromosoma II (NC\_004311), Bs1330r15CI2 = *B. suis* 1330 cromosoma I (NC\_004310), Bs1330r15CI1 = *B. suis* 1330 cromosoma I (NC\_004310), Ba19941r15CI2 = *B. abortus* bv. 1 str. 9-941 cromosoma I (NC\_006932), Ba19941r15CI1 = *B. abortus* bv. 1 str. 9-941 cromosoma I (NC\_006932), Ba19941r15CII = *B. abortus* bv. 1 str. 9-941 cromosoma II (NC\_006933), Bmar15CII = *B. melitensis* biovar *Abortus* 2308 cromosoma II (NC\_007624), Bmar15CI2 = *B. melitensis* biovar *Abortus* 2308 cromosoma I (NC\_007618), Bmar15CI1 = *B. melitensis* biovar *Abortus* 2308 cromosoma I (NC\_007618), Bor15CI2 = *B. ovis* ATCC 25840 cromosoma I (NC\_009505), Bor15CI1 = *B. ovis* ATCC 25840 cromosoma I (NC\_009505), Bor15CII = *B. ovis* ATCC 25840 cromosoma II (NC\_009504), Bmir15CII = *B. microti* CCM 4915 cromosoma 2 (NC\_013118), Bmir15CI2 = *B. microti* CCM 4915 cromosoma 1 (NC\_013119), Bmir15CI1 = *B. microti* CCM 4915 cromosoma 1 (NC\_013119), Oar15CI = *O. anthropi* ATCC 49188 cromosoma 1 (NC\_009667), Oar15CII = *O. anthropi* ATCC 49188 cromosoma 2 (NC\_009668).

En estos genomas los genes  $\alpha$ 15 están presentes en al menos tres copias; dos de ellas localizadas en el cromosoma principal de cada especie y el resto en plásmidos o segundos cromosomas (e.g. en especies del género *Brucella*). Al igual que Smr15C1 y Smr15C2 en *S. meliloti* 1021, los 22 miembros de la familia  $\alpha$ 15 más cercanos a éstos (según la puntuación de Infernal 1.0) están codificados en genes cromosómicos dispuestos en tándem en la misma RIG de los siguientes genomas: *S. medicae*, *S. fredii*, *R. leguminosarum* bv. *trifolii* (cepas WSM304 y WSM35), *R. leguminosarum* bv. *viciae*, *R. etli* (cepas CFN 42 y CIAT 652), *A. vitis*, *A. tumefaciens*, *A. radiobacter* y *A. H13*. Además, los dos ORFs que flanquean estos loci sRNA están también conservados en la mayoría de los casos. En especies del género *Brucella*, sin embargo, las dos copias cromosómicas de los genes que codifican representantes de esta familia de sRNAs se localizan en posiciones distantes del replicón principal aunque sus regiones genómicas son también totalmente sinténicas en estos grupos bacterianos.

El agrupamiento de todos los miembros de la familia  $\alpha$ 15 utilizando criterios de homología estructural generó un árbol que correlaciona, en lo referente a las copias cromosómicas de los mismos, con la filogenia del orden inferida de marcadores taxonómicos clásicos (i.e. 16S RNA) (Fig. III-11). Sin embargo, muchos de los sRNAs codificados en plásmidos tienen una localización más dispersa en el árbol. La mayoría de estos loci se localizan en regiones genómicas no sinténicas y están flanqueados por genes que codifican secuencias de inserción (e.g. Smr14A), transposasas, excisionasas y otros elementos genéticos que favorecen los eventos de movilidad y los reordenamientos genómicos.

### III-3.1.2. EXPRESIÓN DE Smr15C1 Y Smr15C2

#### **III-3.1.2.1. Cinética de expresión en medio TY**

Las cinéticas de expresión de Smr15C1 y Smr15C2 en cultivos de *S. meliloti* 1021 en medio completo TY se determinaron mediante la hibridación de RNA total extraído en diferentes etapas de crecimiento con oligonucleótidos para la detección específica de cada transcrito (Fig. III-12A, panel izquierdo). Estos experimentos revelaron una señal de hibridación para Smr15C1 de poca intensidad durante la fase logarítmica que, sin embargo, aumentó paulatinamente con la edad del cultivo hasta alcanzar su máximo en fase estacionaria tardía (D.O. 2,4). Por el contrario, la expresión de Smr15C2 fue máxima en etapas tempranas del crecimiento (D.O. 0,3) y disminuyó progresivamente hasta hacerse casi indetectable en fase estacionaria tardía (D.O. 2,4).

#### **III-3.1.2.2. Expresión en condiciones de estrés**

En una segunda serie de hibridaciones con las mismas sondas se analizó la acumulación de cada transcrito en condiciones biológicas diversas que imitan situaciones de estrés biótico y abiótico relevantes en simbiosis (e.g. exposición a flavonoides o exudados radiculares, microaerobiosis y nódulos maduros) o a las que se conoce que responden sRNAs bacterianos bien caracterizados (e.g. shock osmótico, estrés de membrana o carencia de Fe) (Fig. III-12A, panel derecho).

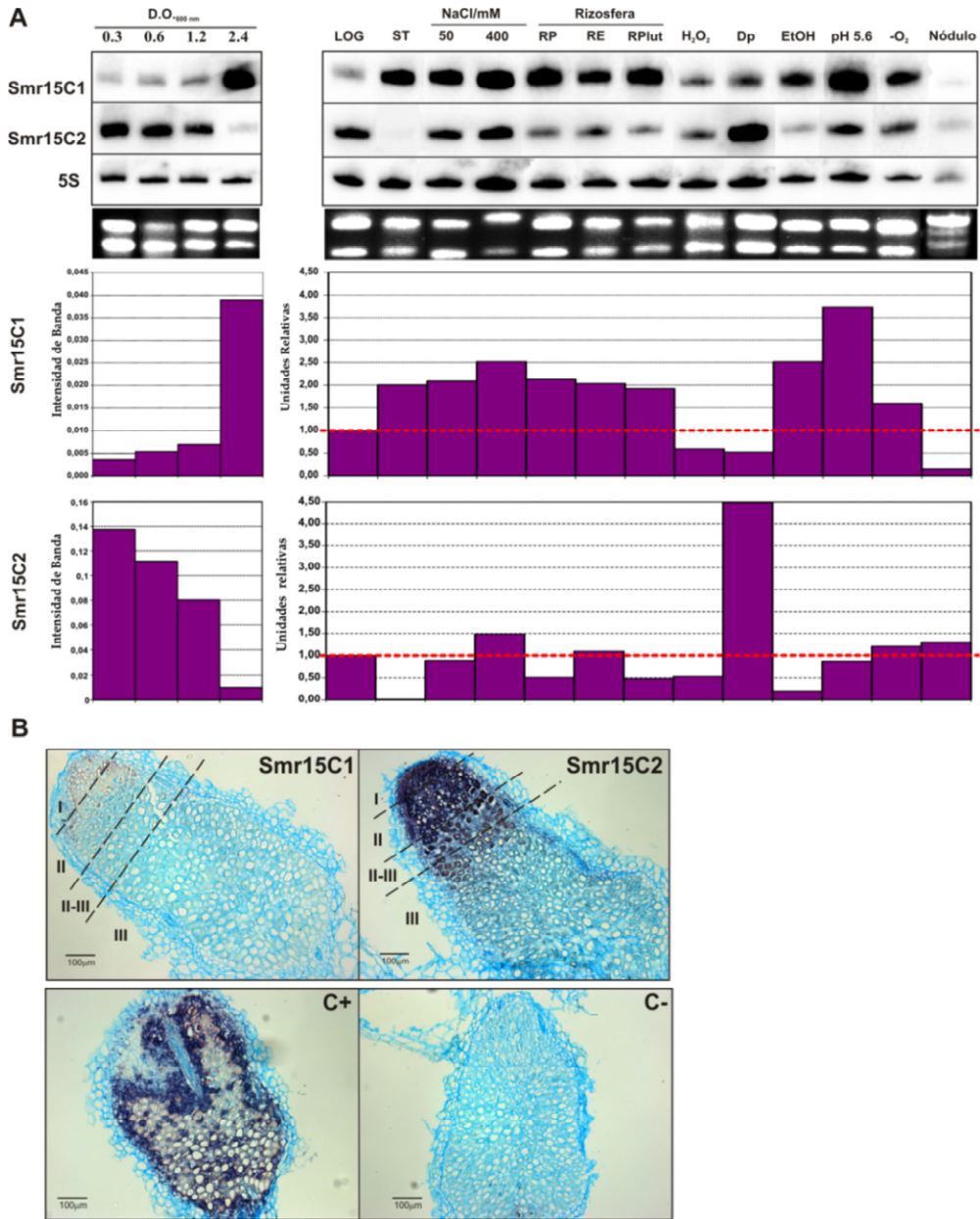
Estas nuevas hibridaciones confirmaron la inducción de Smr15C1 en fase estacionaria de crecimiento (Fig. III-12A parte derecha). La intensidad de la señal de hibridación fue similar o ligeramente mayor que ésta en RNA de bacterias crecidas en presencia de 50 mM de NaCl, sometidas a shock osmótico, estrés de membrana (i.e. tratamiento con EtOH), microaerobiosis y acidez, condición esta última en la que se detectaron los máximos niveles de expresión (>3,5 veces superior a la referencia). La acumulación de Smr15C1 fue también similar a la observada en fase estacionaria cuando las bacterias se incubaron en medio de enraizamiento Rigaud y Puppo. Sin embargo, no se observó efecto alguno de la presencia de plantas de alfalfa (exudados radiculares) o del flavonoide luteolina sobre la expresión de este sRNA, lo que sugiere que su inducción en estas condiciones es posiblemente debida a las propias características del medio Rigaud y Puppo en el que, por la extrema limitación de nutrientes (i.e. carencia de fuentes de C y N), las bacterias pueden sobrevivir en presencia de exudados radiculares pero no dividirse activamente. Por otra parte, en bacterias sometidas a estrés oxidativo o crecidas en medio suplementado con el

quelante de hierro 2,2'-Dipiridil la acumulación de Smr15C1 disminuyó discretamente respecto a la detectada en la condición de referencia (i.e. cultivo en fase logarítmica en TY), mientras que en endosimbiosis no se observó señal de hibridación.

El análisis de la expresión de Smr15C2 en las mismas condiciones confirmó la represión de este transcrito en fase estacionaria tardía de crecimiento. El estrés salino, la acidez, la microaerobiosis y la endosimbiosis (i.e. nódulos maduros) no modificaron sustancialmente la expresión de este sRNA respecto a la observada en bacterias en crecimiento exponencial, mientras que el estrés oxidativo, la adición de etanol al medio y la incubación de las bacterias en medio Rigaud y Puppo resultaron en una discreta disminución de los niveles del transcrito. Ni los exudados radiculares ni la luteolina tuvieron tampoco efectos evidentes sobre la expresión de Smr15C2. Sin embargo, el cultivo en condiciones de limitación de Fe resultó en un claro aumento de la acumulación del transcrito (~4,5 veces respecto a la condición de referencia).

#### **III-3.1.2.3 Expresión en nódulos maduros de alfalfa**

El análisis de la expresión de Smr15C1 y Smr15C2 se completó con experimentos de hibridación *in situ* sobre tejido nodular maduro de *M. sativa* utilizando ribosondas antisentido, sintetizadas *in vitro*, específicas de ambos transcritos. En la Fig. III-12B se muestran cortes de nódulos representativos de las hibridaciones. Estos experimentos revelaron una fuerte señal de hibridación correspondiente a Smr15C2 en el meristemo apical del nódulo (zona I) y en la zona de invasión (II). La intensidad de esta señal va decreciendo en la interzona II-III, donde comienza la diferenciación morfológica de los bacteroides, hasta hacerse imperceptible en los bacteroides maduros que ocupan la zona de fijación de nitrógeno o zona III (Fig. III-12B sección superior). Sin embargo, la ribosonda de Smr15C1 no detectó la presencia del transcrito en ninguna de las zonas del tejido nodular, confirmando así los resultados de las hibridaciones Northern. Como controles positivo y negativo de expresión se hibridaron secciones de nódulos maduros de alfalfa con una sonda para la detección de la anhidrasa carbónica de *Medicago* y una ribosonda sentido de uno de los transcritos Smr, respectivamente (Fig. III-12B, panel inferior).



**Fig. III-12. Expresión de los genes *smr15C1/C2*.** **A)** Cinéticas de expresión en medio TY (panel izquierdo) y acumulación en respuesta a diversos estreses abióticos y bióticos (panel derecho). En la parte inferior se presentan las cuantificaciones de las señales normalizadas con la del RNA 5S. Las líneas punteadas indican la intensidad relativa de la señal de hibridación en la condición de referencia (TY-LOG). La sonda utilizada se indica a la izquierda de cada panel. Leyenda: LOG, fase logarítmica; ST, f. estacionaria; 50 y 400 mM de sal; RP, Rigaud&Puppo; RE, exudados radiculares; RPLut, Rigaud&Puppo con 20  $\mu$ M luteolina; H<sub>2</sub>O<sub>2</sub>, 1 mM agua oxigenada; Dp, 2,2' Dipiridil; -O<sub>2</sub>, microaerobiosis. **B)** Hibridación *in situ* sobre nódulos maduros de *M. sativa* ocupados por *S. meliloti* 1021. La ribosonda utilizada se indica en cada caso. C+, sonda anhidrasa carbónica *M. truncatula*; C-, sonda RNA sentido. La escala de las imágenes se indica en la parte inferior

#### III-3.1.2.4. Análisis de las regiones promotoras

Los experimentos anteriores revelan que los perfiles de expresión de Smr15C1 y Smr15C2 en *S. meliloti* 1021 son prácticamente opuestos, lo que sugiere una regulación diferencial de los mismos.

Todos los *loci*  $\alpha$ 15 tienen cajas promotoras dependientes de  $\sigma^{70}$  reconocibles, con los motivos consenso -35/-10 CTTGAC-n17-CTATAT. Sin embargo, el alineamiento múltiple de las regiones promotoras de los representantes de la familia  $\alpha$ 15 de sRNAs en especies de la familia Rhizobiaceae evidenció zonas conservadas de las mismas hasta aproximadamente la posición -80 respecto al sitio de inicio de transcripción predicho o determinado en su caso (e.g. para los transcritos Smr15C1 y Smr15C2) para cada sRNA (Fig. III-13). Además, este alineamiento discriminó entre dos grupos de promotores, incluyendo el primero los correspondientes a los sRNAs Smr15C1, Sfr15C1 y Smedr15C1 de *S. meliloti*, *S. fredii* y *S. medicae*, respectivamente, y el segundo, el resto de secuencias alineadas. En el primer grupo de promotores, MEME ([http://meme.sdsc.edu/meme4\\_6\\_1/intro.html](http://meme.sdsc.edu/meme4_6_1/intro.html)) identificó una región conservada, aproximadamente entre las posiciones -50 y -73 (enmarcada en verde en la Fig. III-13), cuya secuencia consenso podría ser el motivo de unión predicho para el factor de transcripción de la familia LacI SMb20537 de *S. meliloti*, potencialmente relacionado con la utilización de azúcares. El segundo grupo de promotores presenta una región de 30 nt bien conservada entre las posiciones -46 y -75 (enmarcada en naranja en la Fig. III-13) cuyo consenso podría ser el reconocido por el factor de transcripción SMb20667 (RegTransBase). Este es también un regulador del tipo LacI probablemente relacionado con el metabolismo de dicarboxilato en Rhizobiales que se ha identificado asociado a la expresión de genes como el de la tartrato deshidrogenasa, la succinato semialdehído deshidrogenasa, la 3-hidroxiisobutirato deshidrogenasa y la hidroxipiruvato isomerasa, en *S. meliloti*, y varias especies del género *Rhizobium* (Finan *et al.*, 2001). Este análisis predice, por tanto, que los sRNAs  $\alpha$ 15C1 están diferencialmente regulados en especies del género *Sinorhizobium* respecto a sus homólogos cromosómicos codificados en la misma RIG en éstas y otras especies de la familia Rhizobiaceae. No obstante, la dependencia de Smr15C1 y Smr15C2 de los factores de transcripción SMb20537 y SMb20667, respectivamente, debe ser verificada experimentalmente en *S. meliloti*.



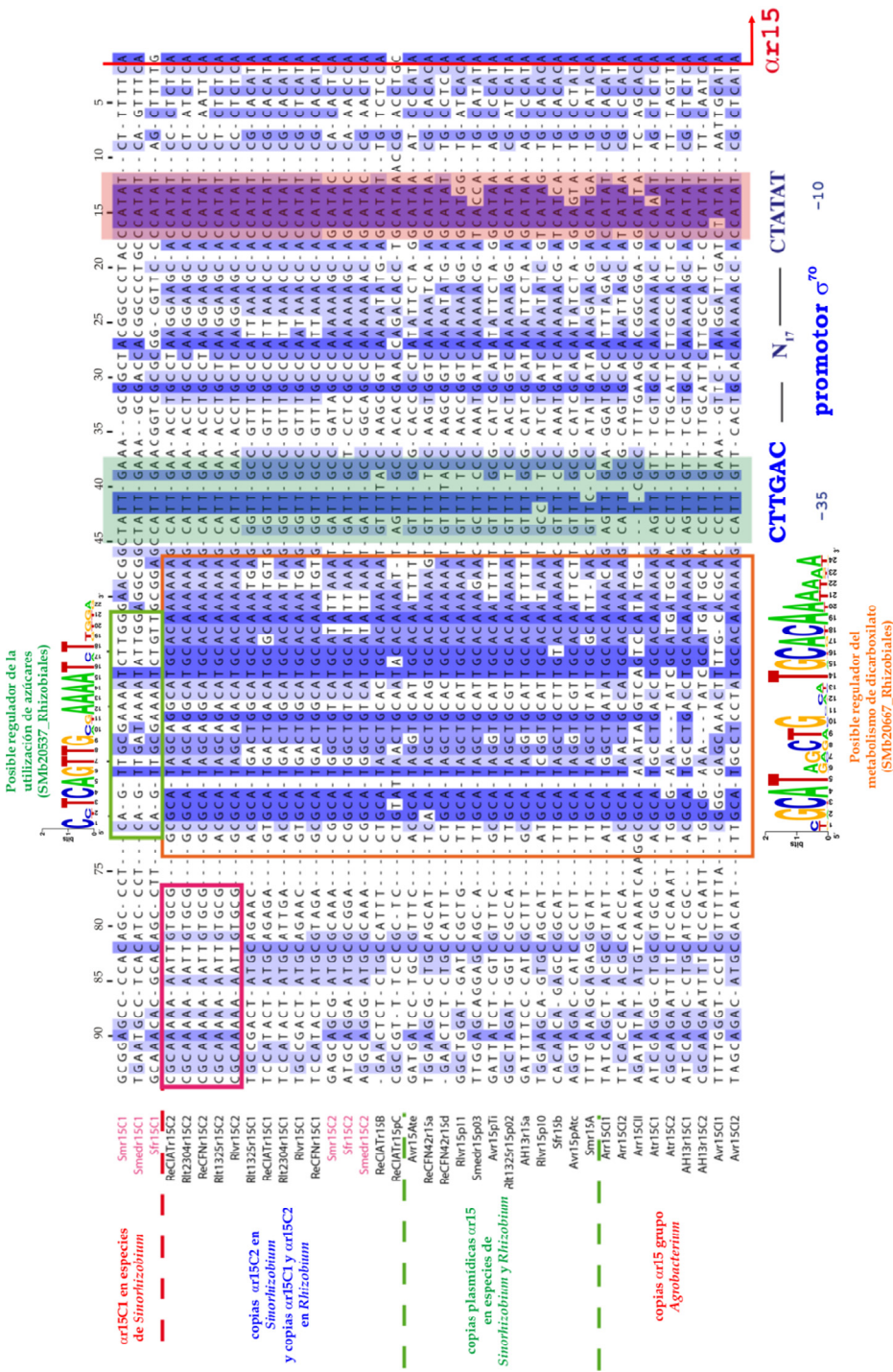


Fig. III-13. Alineamiento de las regiones promotora de los genes *αr15*. Las cajas promotoras -10/-35 de los promotores  $\sigma^{70}$  aparecen marcadas en rosa y verde respectivamente. Los motivos conservados para factores de transcripción se indican encuadrados junto con su secuencia consenso (logo) en cada caso.

### III-3.1.3. APROXIMACIÓN GENÉTICA A LA FUNCIÓN DE Smr15C1 Y Smr15C2

En los siguientes epígrafes se describe la generación y caracterización fenotípica de dos tipos de mutantes en los genes *smr15C1* y *smr15C2*; los obtenidos por **delección** y sustitución posterior de estos *loci* por un marcador de resistencia a Eritromicina y los generados por **sobreexpresión** ectópica de ambos sRNAs desde promotores constitutivos.

#### ***III-3.1.3.1. Generación y verificación experimental de los mutantes smr15C***

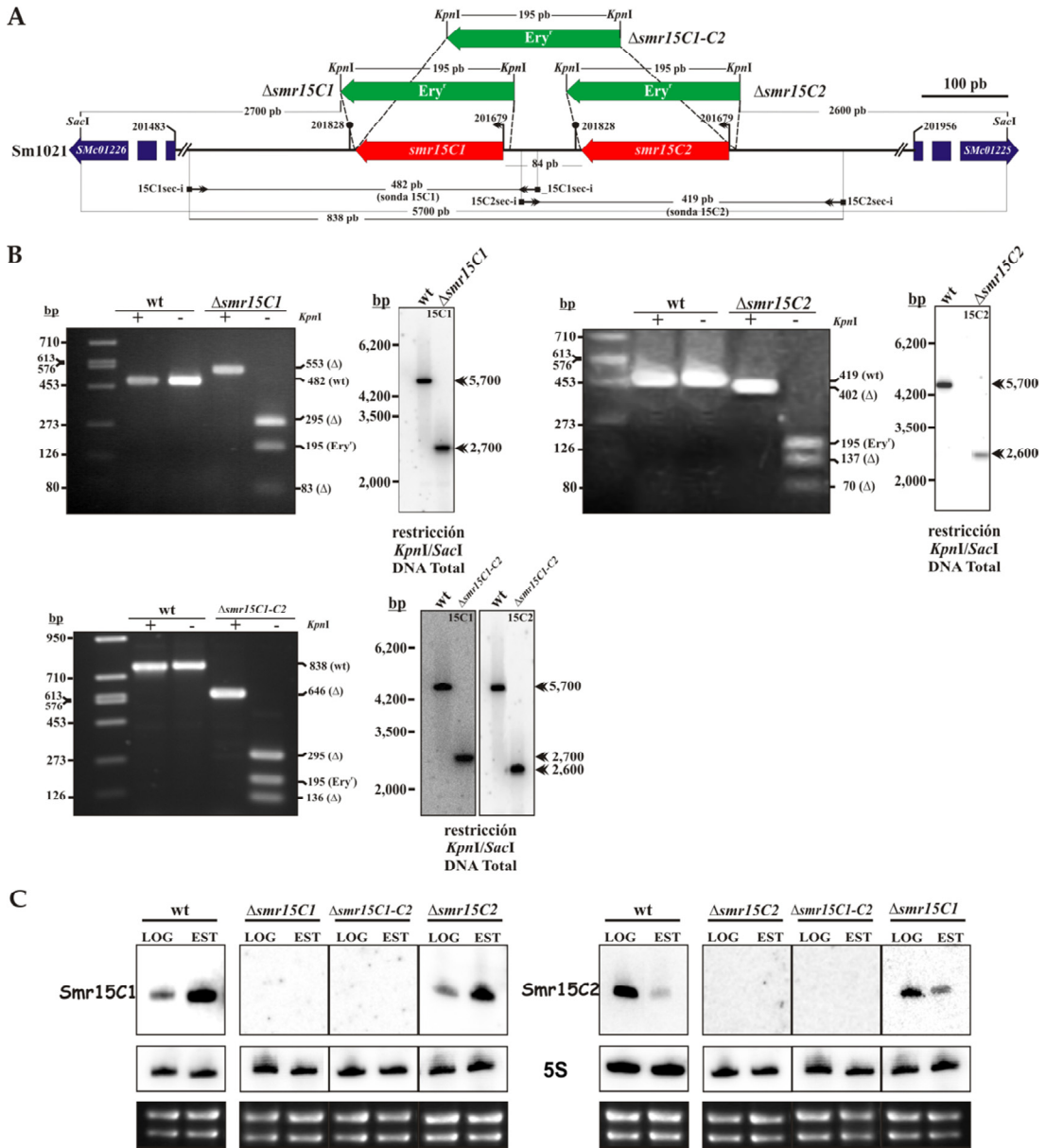
##### III-3.1.3.1.1. Mutantes por delección

Ya que Smr15C1 y Smr15C2 son sRNAs homólogos y quizá funcionalmente relacionados se diseñaron tres tipos de mutantes; dos obtenidos por delección completa e independiente de cada *locus* manteniendo intacto el vecino ( $\Delta smr15C1$  y  $\Delta smr15C2$ ) y un tercero por delección simultánea de ambos genes ( $\Delta smr15C1-C2$ ) (Fig. III-14A). La generación correcta de la mutaciones fue verificada mediante PCR, Southern blot (Fig. III-14B) y Northern blot (Fig. III-14C).

Para las amplificaciones de los alelos silvestre y mutante se utilizaron parejas de cebadores diferentes en cada caso; 15C1sec/15C1sec-i (para verificar  $\Delta smr15C1$ ), 15C2sec/15C2sec-i (para  $\Delta smr15C2$ ) y 15C1sec/15C2sec-i (para  $\Delta smr15C1-C2$ ). En DNA total de *S. meliloti* 1021 estos oligonucleóticos generaron fragmentos de 482, 419 y 838 pb, respectivamente, que no son sensibles a la digestión con *KpnI*, que es el sitio de restricción que flanquea el cassette *Ery<sup>r</sup>* (195 pb) que reemplaza los *loci* delecionados en cada mutante (Fig. III-14B, dos primeros carriles de cada gel de agarosa). La amplificación de los alelos mutados con las mismas combinaciones de oligonucleótidos generó amplicones del tamaño esperado en cada caso, 553 pb ( $\Delta smr15C1$ ), 402 pb ( $\Delta smr15C2$ ) y 646 pb ( $\Delta smr15C1-C2$ ), que al ser digeridos con *KpnI* generaron perfiles de restricción en los que fue identificable la banda correspondiente al marcador *Ery<sup>r</sup>* (Fig. III-14B, dos últimos carriles de cada gel de agarosa).

La hibridación de DNA genómico de la cepa salvaje digerido con *KpnI-SacI*, con las sondas 15C1 y 15C2 cuya secuencia solapa con los *loci smr15C1* y *smr15C2*, respectivamente (Fig. III-14A) generó una única señal correspondiente a un fragmento esperado de 5.700 pb en todos los casos (Fig. III-14B, primeros carriles de cada membrana). En DNA total de los mutantes  $\Delta smr15C1$  y  $\Delta smr15C2$  las sondas 15C1 y 15C2 hibridaron con fragmentos *KpnI-SacI* de 2.700 y 2.600 pb, respectivamente, que también fueron evidentes en hibridaciones independientes

sobre DNA del doble mutante  $\Delta smr15C1$ -C2 (Fig. III-14B, segundos carriles de cada membrana).



**Fig. III-14. Verificación experimental de las mutaciones  $\Delta smr15C$ .** A) Región genómica de *smr15C1/C2* en *S. meliloti* 1021 en donde se indican los mutantes generados ( $\Delta smr15C1$ / $\Delta smr15C2$ / $\Delta smr15C1$ -C2) con la posición del *casette* de eritromicina (Ery) que los sustituye en cada caso, además de información necesaria para la interpretación del resto de los paneles. B) Geles de agarosa correspondientes a la verificación por PCR (paneles situados a la izquierda) y Southern blot (paneles situados a la derecha) de los mutantes anteriores. En la esquina superior derecha de cada Southern se indica la sonda de hibridación. El tamaño y el origen de las bandas se indica en cada caso (wt, cepa salvaje;  $\Delta$ , mutante; Ery, *casette* de eritromicina). C) Hibridación Northern sobre RNA total utilizando sondas para *Smr15C1* (izquierda) y *Smr15C2* (derecha). Las cepas utilizadas se indican en cada caso, además de la condición de crecimiento para la obtención del RNA (LOG, fase exponencial; EST, fase estacionaria).

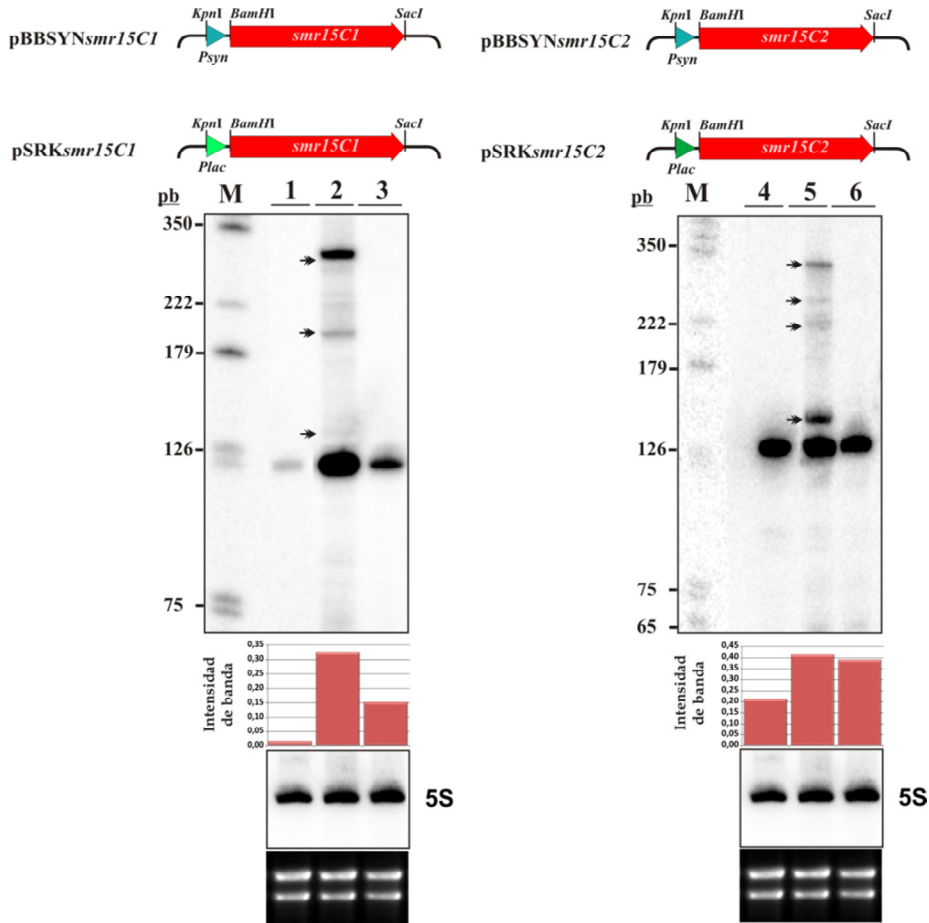
La hibridación de RNA total de cada mutante, obtenido de bacterias cultivadas en medio completo TY hasta las fases exponencial y estacionaria de crecimiento, confirmó la ausencia de los transcritos expresados por los genes deletados en cada caso (Fig. III-14C). Por otra parte, estos mismos experimentos demostraron que el patrón de expresión de estos transcritos no se ve alterado cuando se elimina la copia homóloga (Fig. III-14C, últimos carriles de cada panel) lo que descarta efectos polares de cada mutación y sugiere que no existe una relación jerárquica entre Smr15C1 y Smr15C2.

#### III-3.1.3.1.2. Mutantes por sobreexpresión

Para sobreexpresar los genes *smr15C* desde plásmidos multicopia en *S. meliloti* se diseñaron dos construcciones: pBBSYN*smr15C* y pSRK*smr15C*, basadas en el promotor *syn* y la versión constitutiva del promotor *lac*, según se detalla en el apartado de "Diseño Experimental". Para verificar el rendimiento de ambas construcciones en la transcripción de Smr15C1 y Smr15C2 se compararon los niveles de ambos transcritos en *S. meliloti* 1021 y en los respectivos transconjugantes mediante hibridación Northern sobre RNA total extraído de bacterias crecidas hasta fase exponencial (Smr15C2) y estacionaria de crecimiento (Smr15C1) (Fig. III-15).

En bacterias transformadas con pBBSYN*smr15C1* los niveles de este transcrito aumentaron del orden de 20 veces respecto a los determinados en la cepa silvestre, mientras que con la expresión constitutiva de Smr15C2 mediante el mismo sistema este aumento fue de dos veces respecto a la referencia (Fig. III-15, carriles 2 y 5). Estas hibridaciones, a su vez, mostraron la aparición de una serie de bandas minoritarias correspondientes a transcritos de mayor tamaño que probablemente deriven de ondas de transcripción procedentes del propio plásmido.

La expresión de Smr15C1 desde pSRK se tradujo en un aumento del nivel del sRNA de más de 9 veces, mientras que la de Smr15C2 prácticamente dobló la detectada en la cepa silvestre (Fig. III-15, carriles 3 y 6). Sin embargo, en este caso no se detectaron especies de RNA diferentes a la esperada.



**Fig. III-15. Sobreexpresión constitutiva de Smr15C1 y Smr15C2.** Hibridación Northern sobre RNA total de los mutantes que sobreexpresan *smr15C1* (sección izquierda) y *smr15C2* (sección derecha) utilizando la sonda específica de cada gen. Los carriles en cada caso corresponden a: M, marcador pGEM; 1, cepa silvestre 1021; 2, 1021pBBSYNsmr15C1; 3, 1021pSRKsmr15C1 (panel izquierdo); 4, 1021; 5, 1021pBBSYNsmr15C2; 6, 1021pSRKsmr15C2 (panel izquierdo). En cada panel se indican con flechas dobles las bandas inespecíficas de mayor tamaño. En la parte superior de cada panel se muestran los diagramas correspondientes a las construcciones utilizadas en cada caso, indicando los sitios de restricción más relevantes. En la parte inferior se muestra la cuantificación de las señales obtenidas y normalizadas con la intensidad de la señal de hibridación del RNA 5S.

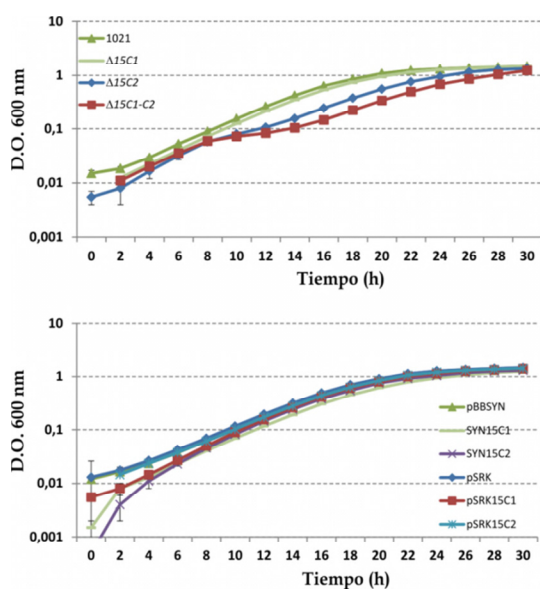
### III-3.1.3.2. Análisis fenotípico de los mutantes *smr15C*

#### III-3.1.3.2.1. Fenotipo en vida libre

Los mutantes *S. meliloti* 1021 $\Delta$ *smr15C1*,  $\Delta$ *smr15C2* y  $\Delta$ *smr15C1-C2* mostraron una morfología de colonia idéntica a la de la cepa silvestre 1021 al crecer en placas de agar con medio completo TY o medio mínimo MM. La monitorización del crecimiento de la cepa salvaje y el mutante  $\Delta$ *smr15C1* en cultivo líquido en medio TY no reveló diferencias significativas entre ambos. Sin embargo el mutante  $\Delta$ *smr15C2* y el doble mutante  $\Delta$ *smr15C1-C2* mostraron un ligero retraso en fase logarítmica (más marcado

en este último), aunque la fase de adaptación (*lag*) y la D.O. final fueron similares a la cepa 1021 (Fig. III-16, sección izquierda).

Los mutantes que sobreexpresan Smr15C1 y Smr15C2 desde pBBSYN y pSRK también fueron sometidos al mismo análisis aunque en este caso las referencias fueron bacterias silvestres transformadas con los plásmidos sin inserto. La observación de la morfología de las colonias control y las que sobreexpresan sendos sRNAs crecidas en placas de TY o MM agar, tampoco evidenció diferencias. La sobreexpresión de Smr15C1 y Smr15C2 mediante cualquiera de los sistemas no tuvo tampoco un efecto evidente sobre la tasa de crecimiento respecto a la observada en las bacterias control en medio TY.

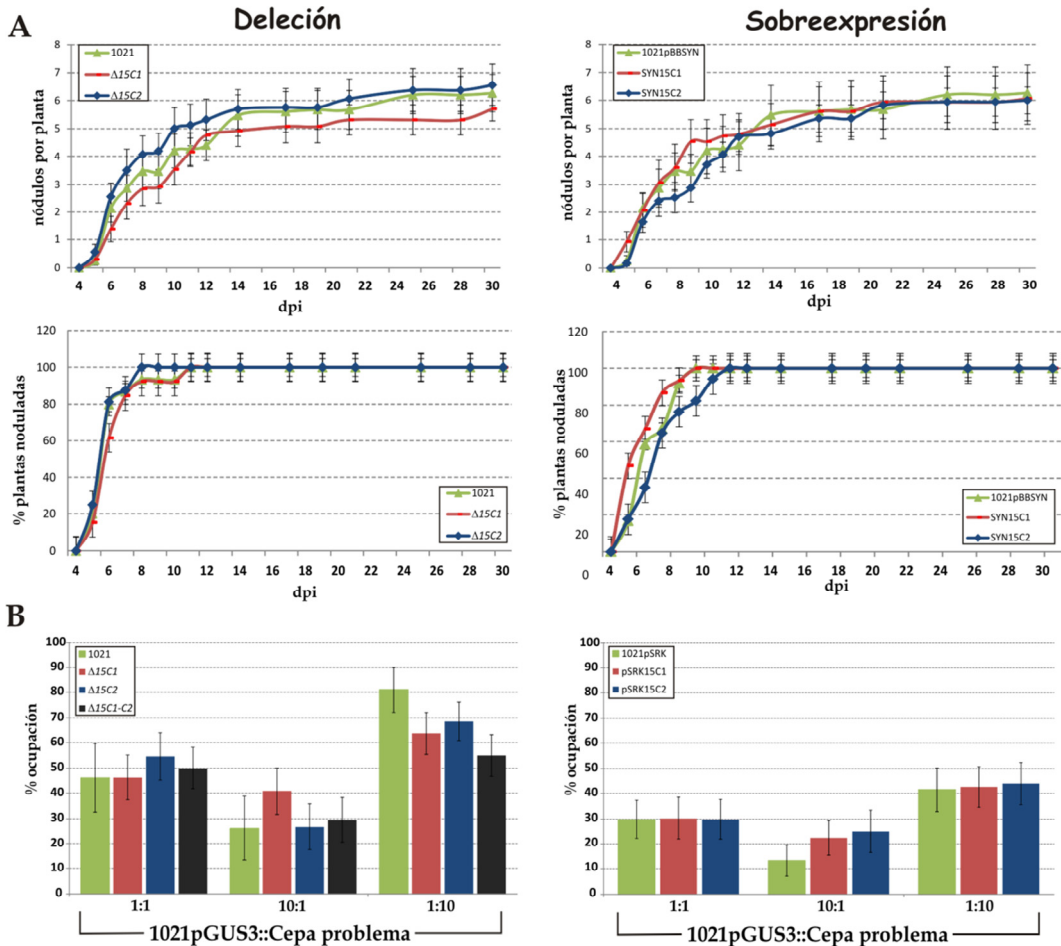


**Fig. III-16. Fenotipo de crecimiento de los mutantes *smr15C*.** Curvas de crecimiento correspondientes a los mutantes por delección (panel superior) y sobreexpresión (panel inferior).

### III-3.1.3.2.2. Fenotipo simbiótico

#### III-3.1.3.2.2.1. Nodulación y competitividad

Para evaluar el comportamiento simbiótico de los mutantes *smr15C* en *S. meliloti* se llevaron a cabo experimentos de nodulación y de competitividad. Para los primeros, se inocularon independientemente sets de 24 plantas crecidas en cultivo hidropónico en tubos de ensayo con suspensiones bacterianas de la cepa salvaje



**Fig. III-17. Cinéticas de nodulación y competitividad de los mutantes en los genes *smr15C*.** Los paneles situados a la izquierda corresponden a los mutantes por delección y los de la derecha a los mutantes por sobreexpresión. **A)** Infectividad y cinéticas de nodulación representadas por el número de nódulos por planta (paneles superiores) y porcentaje de plantas noduladas (paneles inferiores) de los mutantes *smr15C*. **B)** Ensayos de competitividad (porcentaje de ocupación) de nódulos de las cepas silvestre y mutantes en presencia de la reportera pGUS3. La proporción de la cepa problema sobre la control se indica en cada caso.

(1021) y los mutantes simples (delección y sobreexpresión) en cada uno de los genes,. El número de nódulos por planta inducidos por cada cepa (infectividad) y el porcentaje de plantas noduladas (cinética de nodulación) se determinaron a intervalos diarios post-inoculación (dpi) en tres experimentos independientes. En los ensayos con las bacterias que sobreexpresan los sRNAs el control de referencia fue la cepa silvestre transformada con el plásmido pBBSYN vacío. Ni la delección ni la sobreexpresión de *Smr15C1* y *Smr15C2* afectaron significativamente la infectividad y cinética de nodulación de *S. meliloti* 1021 en las condiciones de ensayo, lo que sugiere que estos sRNAs son dispensables para el establecimiento de la simbiosis.

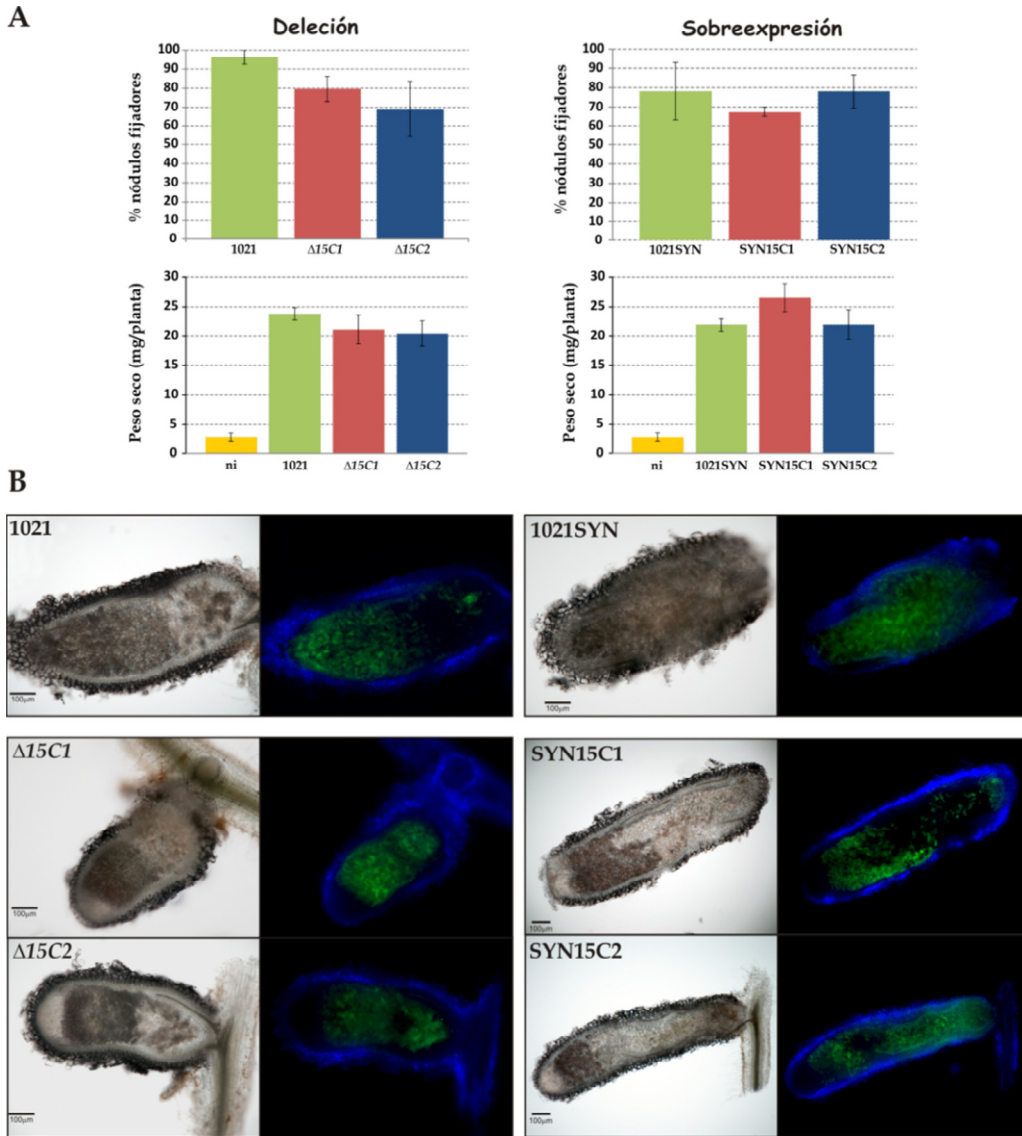
Para los ensayos de competitividad se inocularon sets de 24 plantas de alfalfa crecidas en tubos de ensayo con mezclas, en diferentes proporciones (1:1, 10:1 y 1:10), de la cepa *S. meliloti* 1021 pGUS3, que expresa constitutivamente el gen reportero *gus*, con cada uno de los mutantes en los genes *smr15*. En estos ensayos se incluyó también el doble mutante  $\Delta smr15C1-C2$  mientras que los mutantes por sobreexpresión utilizados fueron los basados en el sistema pSRK. Como referencias de los ensayos con las respectivas cepas mutantes, en series simultáneas de experimentos se co-inoculó la cepa reportera con la cepa silvestre *S. meliloti* 1021 o la cepa 1021pSRK en todas las proporciones mencionadas. Los resultados de estos experimentos se muestran en la Fig. III-17B. El porcentaje de ocupación de nódulos de las cepas mutantes aumentó o disminuyó de acuerdo a su concentración en las mezclas de co-inoculación pero en ningún caso se determinaron diferencias estadísticamente significativas con los porcentajes de ocupación de las respectivas cepas control en cada experimento. Concluimos, por tanto, que la actividad de *Smr15C1* y *Smr15C2* no tiene un aparente impacto en la competencia rizosférica de *S. meliloti* 1021 en condiciones de laboratorio.

#### III-3.1.3.2.2.2. Etapas tardías de la simbiosis

Como primera valoración del impacto de la delección y sobreexpresión de *Smr15C1* y *Smr15C2* en etapas tardías de la simbiosis se procedió al conteo de los nódulos fijadores de nitrógeno (rosáceos) en los sets de 24 plantas inoculadas independientemente con cada cepa mutante y sus respectivas referencias (i.e. *S. meliloti* 1021 para la delección y 1021pBBSYN para la sobreexpresión) 30 dpi. La delección individual de *smr15C1* y *smr15C2* produjo una leve disminución del número de nódulos rosas respecto a la cepa silvestre 1021, aunque en ningún caso estadísticamente significativa (Fig. III-17A, sección superior izquierda). Estos valores correlacionaron, sin embargo, con los del peso seco determinado para las mismas plantas, aunque de nuevo las diferencias no tuvieron significación estadística. La sobreexpresión de ambos sRNAs desde pBBSYN no produjo tampoco variaciones en ninguno de estos dos parámetros respecto a los determinados para la cepa de control, 1021pBBSYN (Fig. III-17A, sección derecha).

La histología de alguno de los nódulos ocupados por las cepas control *S. meliloti* 1021 y 1021pBBSYN y sus respectivos mutantes se visualizó también mediante microscopía de campo claro. Este análisis no reveló diferencias aparentes entre los mutantes y sus respectivas cepas de referencia, mostrando todos los nódulos la





**Fig. III-18. Fenotipos endosimbióticos de los mutantes *smr15C*.** **A)** Porcentaje de nódulos fijadores (paneles superiores) y medida del peso seco (paneles inferiores) de plantas de alfalfa ocupadas por los mutantes por delección (sección izquierda) y sobreexpresión (sección derecha) de los genes *smr15C* (ni, no inoculadas). **B)** Imágenes correspondientes a secciones de nódulos de *M.sativa* obtenidas mediante microscopía de campo claro (imágenes situadas a la izquierda de cada caso) y de fluorescencia utilizando filtros para el verde y el azul (520 y 470 nm, respectivamente). La cepa que ocupa estos nódulos se indica en cada caso. Barra de escala: 100  $\mu$ m.

zonación característica de nódulos indeterminados ocupados homogéneamente por bacteroides autofluorescentes (Fig. III-18B). Las cepas que sobreexpresan los sRNAs inducen nódulos algo más elongados en la zona de senescencia que los de la cepa *S. meliloti* 1021, si bien este discreto fenotipo no es específico de la sobreexpresión de

ninguno de los sRNAs pues fue observado también en la cepa control de estos experimentos (1021 pBBSYN).

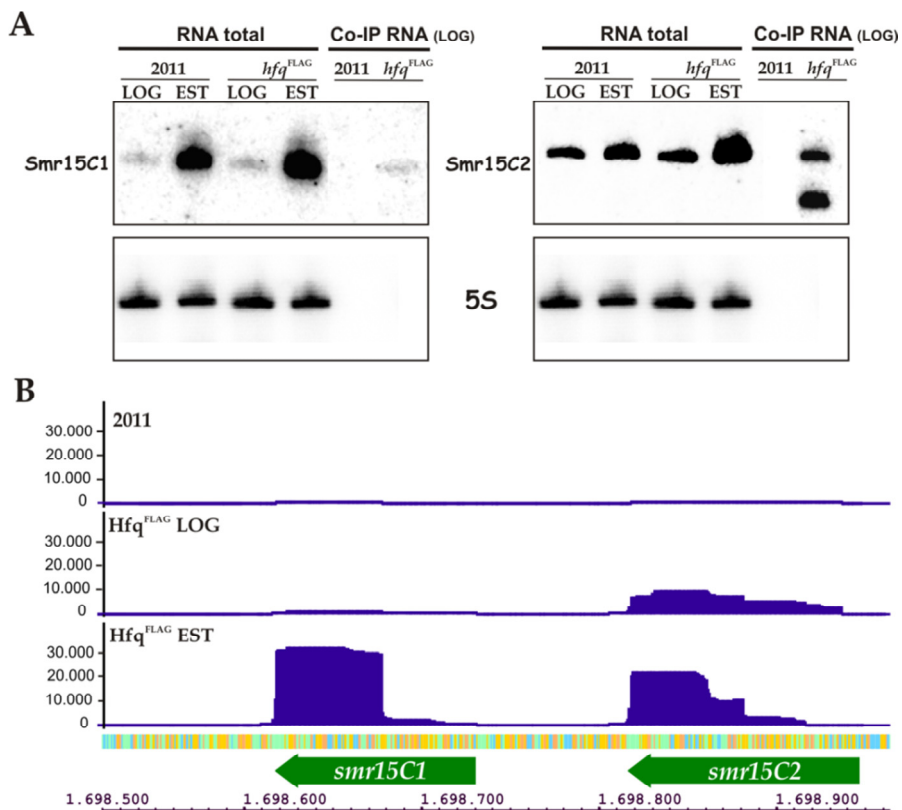
#### III-3.1.4. INTERACCIÓN DE Smr15C1 Y Smr15C2 CON LA CHAPERONA Hfq

La investigación de la posible dependencia de Smr15C1 y Smr15C2 de la proteína Hfq se llevó a cabo mediante dos aproximaciones: análisis de la presencia de estos sRNAs en la subpoblación de transcritos que co-inmunoprecipitaron (CoIP-RNA) con Hfq<sup>FLAG</sup>, como evidencia de su interacción con la chaperona y determinación de la estabilidad de ambos transcritos en el fondo genético  $\Delta hfq$ , como aspecto de la biología de los sRNAs generalmente impactado por la proteína.

Inicialmente el CoIP-RNA preparado a partir de lisados de bacterias en fase exponencial de crecimiento se analizó mediante hibridación Northern. En esta serie de hibridaciones se incluyeron también muestras de RNA total de las cepas *S. meliloti* 2011 y *S. meliloti* 2011 $hfq^{FLAG}$  crecidas en medio TY hasta fase exponencial y estacionaria, como control de la expresión de ambos sRNAs en las mismas (Fig. III-19A). La hibridación con la sonda específica de Smr15C1 reveló que, al igual que en *S. meliloti* 1021, en la cepa 2011 la acumulación máxima de este transcrito también ocurre en fase estacionaria y este patrón de expresión no se vio alterado en su cepa derivada, que expresa la proteína Hfq modificada. El mismo oligonucleótido también detectó Smr15C1 en el CoIP-RNA obtenido de *S. meliloti* 2011 $hfq^{FLAG}$  pero no en el control (i.e. proveniente de lisados de la cepa 2011). La sonda de Smr15C2 reveló también patrones de expresión similares de este transcrito en las cepas 2011 y 2011 $hfq^{FLAG}$ , aunque sorprendentemente diferentes a los observados para el mismo sRNA en *S. meliloti* 1021. En este caso, en el CoIP-RNA se detectaron señales de hibridación correspondientes al transcrito completo Smr15C2 y a otra molécula de RNA de menor tamaño, posiblemente derivada de éste. Estas señales no fueron detectadas en el CoIP-RNA control lo que sugiere también la asociación específica de ambas especies de RNA a Hfq.

La distribución de las lecturas de la secuenciación masiva del CoIP-RNA en la región genómica de Smr15C1 y Smr15C2 confirmó los resultados de las hibridaciones Northern (Fig. III-19B). Los cDNAs derivados de Smr15C1/C2 en el CoIP-RNA obtenido de bacterias en fase exponencial de crecimiento mapearon a lo largo de prácticamente toda la secuencia anotada del transcrito, desde el extremo 5' hasta el terminador de transcripción independiente de Rho, mientras que no se obtuvieron apenas lecturas correspondientes a Smr15C1. Por el contrario, ambos sRNAs aparecieron muy representados en CoIP-RNA de bacterias cultivadas hasta fase estacionaria de crecimiento, momento de su nivel de expresión máximo en esta

bacteria. Sin embargo, en estas condiciones la mayor frecuencia de lecturas correspondió a la región 3' de ambos transcritos, probablemente debido a que Hfq tiene mayor afinidad por la misma. La molécula de RNA derivada de Smr15C2 detectada mediante hibridación Northern en CoIP-RNA podría corresponder, por tanto, a la región 3' de este transcrito, enriquecida en la muestra por su particular afinidad por esta chaperona. Sin embargo, el oligonucleótido utilizado como sonda hibrida específicamente en su región 5', la más divergente en secuencia respecto a Smr15C1, por lo que estos resultados sólo serían explicables por una posible hibridación cruzada no predicha con la región 3' del transcrito expresado por *S. meliloti* 2011.



**Fig. III-19. Asociación de los sRNAs Smr15C1/C2 a la chaperona Hfq.** A) Hibridación Northern sobre RNA total y CoIP-RNA de las cepas sometidas a CoIP con las sondas específicas de Smr15C1 (panel izquierdo) y Smr15C2 (panel derecho). B) Cobertura (número de lecturas) de la región *smr15C1-C2* en las tres librerías de cDNA sometidas a RNA-Seq visualizadas con IGB. La escala inferior corresponde a las coordenadas en el genoma de *S. meliloti* 1021.

Para la determinación de la estabilidad de Smr15C1 y Smr15C2 en ausencia de Hfq, se crecieron bacterias en cultivos independientes de las cepas 1021 y 1021 $\Delta$ *hfq*, en medio TY hasta fase logarítmica (para el análisis de la acumulación de Smr15C2) y fase estacionaria tardía (para el estudio de Smr15C1), añadiendo posteriormente

rifampicina a una concentración de 800 µg/ml para detener toda actividad de transcripción en la célula. A continuación, se recogieron fracciones de ambos cultivos a diferentes tiempos (5, 15, 30 y 45 min), se extrajo el RNA total y se sometió a análisis mediante Northern blot para determinar la cinética de desaparición de ambos transcritos en cada uno de los fondos genéticos (Fig. III-20). El tiempo determinado de vida media (i.e. tiempo en el que la concentración del transcrito va disminuyendo a la mitad) de Smr15C1 y Smr15C2 en *S. meliloti* 1021 fue de 20 y 25 min, respectivamente, mientras que en el mutante 1021 $\Delta$ hfq se redujo drásticamente a 2 y 11 min en cada caso. Estos resultados demuestran que la estabilidad de Smr15C1 y Smr15C2 en ausencia de Hfq está severamente comprometida, lo que constituye una evidencia más de la interacción de ambos transcritos con esta proteína.

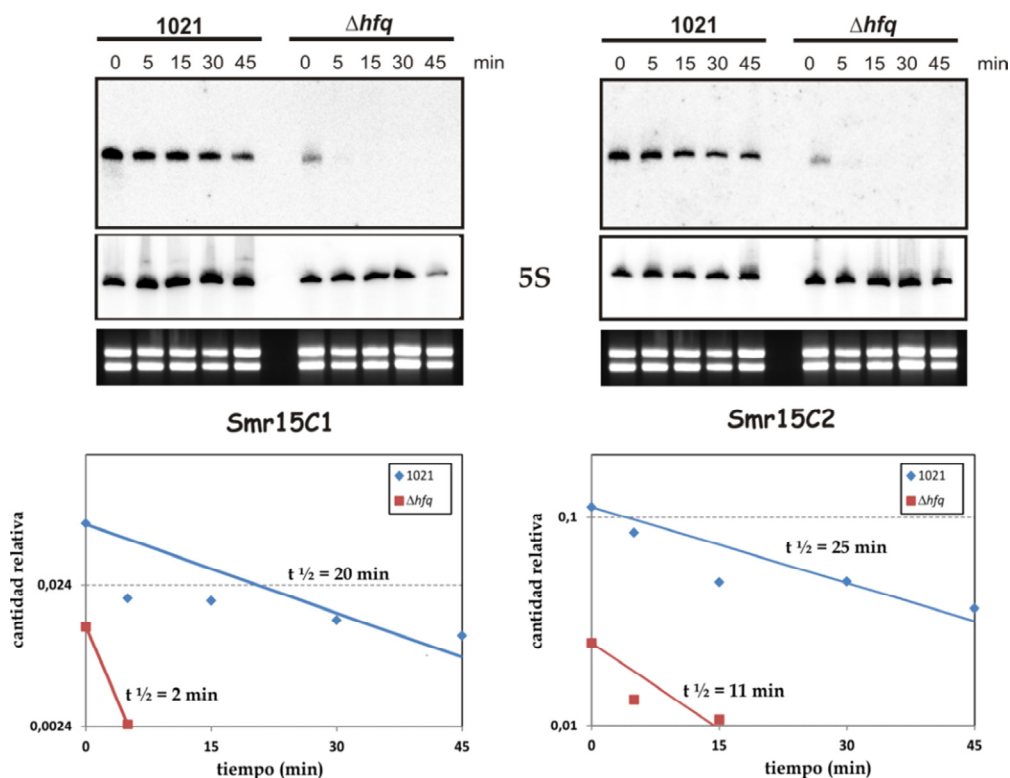


Fig. III-20. Estabilidad de los transcritos Smr15C1/C2 en el fondo genético  $\Delta$ hfq. Los dos paneles superiores representan la hibridación Northern sobre RNA total de la cepa salvaje 1021 y el mutante  $\Delta$ hfq y los inferiores la determinación de las vidas medias ( $t_{1/2}$ ), en min, en los dos fondos genéticos, basada en la cuantificación de cada señal de hibridación en relación a la del RNA 5S. Los tiempos de extracción del RNA total después del tratamiento con Rifampicina se indica en min en la parte superior de cada panel. La sección izquierda representa al sRNA Smr15C1 y la derecha a Smr15C2.

### III-3.1.5. APROXIMACIÓN A LA IDENTIFICACIÓN DE LOS mRNA DIANA DE Smr15C1 Y Smr15C2

La identificación del o los mRNA diana de un *trans*-sRNA es de gran complejidad y generalmente requiere la combinación de varios abordajes como se ha descrito en la sección introductoria de este capítulo. A continuación se describen las primeras aproximaciones a la identificación de las dianas de Smr15C1 y Smr15C2.

#### **III-3.1.5.1. Predicciones con IntaRNA**

Para la predicción de las posibles interacciones entre los sRNAs Smr15C1 y Smr15C2 con sus mRNAs diana se utilizó la herramienta IntaRNA (Busch *et al.*, 2008). La información esencial suministrada al programa fueron las secuencias completas de ambos transcritos y las de los 3 replicones de *S. meliloti* 1021; cromosoma (NC\_003047) y megaplásmidos pSymA (NC\_003037) y pSymB (NC\_003078). La búsqueda de interacciones se limitó a las regiones 5' de todos los genes codificantes de proteínas anotados y más específicamente a las secuencias comprendidas entre las posiciones -250 y +150 respecto al supuesto inicio de traducción AUG de cada mensajero. Esta longitud de secuencia se escogió porque el programa va plegando la molécula de mRNA en ventanas de 150 nt, valorando la accesibilidad de cada motivo para la interacción en el contexto de su estructura secundaria. Como criterio adicional para la generación de las predicciones se impuso el apareamiento de al menos 8 nt contiguos (*seed*) entre sRNA y mRNA, sin restricciones adicionales en cuanto a la longitud de la interacción.

Con estos parámetros IntaRNA generó una lista de 6.182 sitios de interacción posibles para Smr15C1 (3.339 localizados en el cromosoma, 1.281 en pSymA y 1.562 en pSymB), mientras que para Smr15C2 fueron 6.196 (3.345 cromosómicos, 1.287 en pSymA y 1.564 en pSymB). Estas listas se ordenaron de acuerdo a la energía libre del híbrido, dando preferencia a las de menor valor, más favorecidas termodinámicamente. Entre éstas se extrajeron las primeras 50 predicciones correspondientes a cada replicón para cada sRNA. Estas predicciones involucraban mayoritariamente a los nucleótidos desapareados en el bucle de la horquilla SL1 de ambos sRNAs en las interacciones con los mRNAs (motivo  $f_1$ ; Tabla III-1), quizá por la localización en el mismo de un motivo aSD evidente, aunque los nucleótidos que lo flanquean son los más divergentes entre ambas moléculas. Para su análisis posterior se seleccionaron en cada caso las interacciones circunscritas total o parcialmente a 30 nt en el entorno del RBS (i.e. región -15/+15 respecto al codón AUG). Este filtrado redujo los candidatos para Smr15C1 a 69 y a 72 para Smr15C2 (Tabla III-1).

En el caso de Smr15C1, las proteínas codificadas por los mRNAs candidatos aparecen asignadas a las siguientes categorías funcionales en las bases de datos: 34 transportadores ABC (6 en cromosoma, 14 en pSymA y 14 en pSymB), 16 enzimas metabólicas (9 en cromosoma, 1 en pSymA y 6 en pSymB), 5 reguladores transcripcionales (1 en cromosoma y 4 en pSymB), 3 proteínas implicadas en procesos celulares diversos (1 en pSymA y 2 en pSymB) y 11 proteínas hipotéticas (3 en cromosoma, 6 en pSymA y 2 en pSymB). La distribución funcional de las predicciones para Smr15C2 fue similar: 34 transportadores ABC (9 en cromosoma, 11 en pSymA y 14 en pSymB), 22 enzimas metabólicas (7 en cromosoma, 6 en pSymA y 9 en pSymB), 4 factores de transcripción (1 en cromosoma y 3 en pSymB) y 13 proteínas hipotéticas, alguna de localización predicha en membrana (5 en cromosoma, 5 en pSymA y 2 en pSymB).

**Tabla III-1. Predicciones de IntaRNA para Smr15C1/C2 en los tres replicones de *S. meliloti* 1021.**

Replicón	Dominio sRNA	Smr15C1	Smr15C2
<i>Cromosoma</i>	$f_1$ (20,32)	46 (16)	43 (17)
	$f_2$ (63,69)	3 (2)	2 (1)
	$f_1 + f_2$	1 (1)	5 (3)
	mRNA (-15/+15)	19	21
<i>pSymA</i>	$f_1$ (20,32)	44 (20)	42 (20)
	$f_2$ (63,69)	5 (1)	5 (2)
	$f_1 + f_2$	1 (1)	3 (1)
	mRNA (-15/+15)	22	23
<i>pSymB</i>	$f_1$ (20,32)	43 (25)	40 (25)
	$f_2$ (63,69)	2 (1)	3 (1)
	$f_1 + f_2$	5 (2)	7 (2)
	mRNA (-15/+15)	28	28
<b>Predicciones catalogadas</b>		<b>69</b>	<b>72</b>

Este análisis evidenció, por tanto, que los mRNAs predichos por IntaRNA como dianas de Smr15C1 y Smr15C2 codifican mayoritariamente algún componente de transportadores ABC implicados en la adquisición de fuentes de carbono y nitrógeno diversas y en ellos se centró la atención en los estudios posteriores. Como era esperado por la alta homología de secuencia entre estos sRNAs muchos de sus posibles mRNA diana (21) son comunes para ambos transcritos de forma que la lista final se redujo a **46 candidatos** que se relacionan en la Tabla III-2 con indicación de las energías de hibridación predichas en cada caso, las secuencias de sRNA y mRNA implicadas en la interacción, el componente del transportador ABC codificado por el mRNA y el sustrato por el que tendría afinidad cada uno de acuerdo a predicciones o datos experimentales disponibles.

Tabla III-2. Transportadores potencialmente regulados por Smr15C1 y Smr15C2 según las predicciones de IntaRNA.

mRNA candidato	Energía (kcal/mol)	Posición sRNA	Posición mRNA <sup>1</sup>	Componente transportador ABC	Sustrato
<b><u>Smr15C1/Smr15C2</u></b>					
rhaS (SMc02324)	-18,3/-19,3	20-30/22-32	(-15/-5)(-16/-6)	Proteína periplásmica	Rhamnosa
SMc04135	-17,4/-18,1	21-30/22-31	(-13/-4)(-13/-4)	Proteína periplásmica	Amino ácidos/péptidos
SMc02471	-15,2/-16,3	21-30/22-32	(-13/-4)(-14/-4)	Proteína periplásmica	Carbohidratos
ugpB (SMb20416)	17,3/-17,4	21-30/22-31	(-13/-4)(-13/-4)	Proteína periplásmica	Carbohidratos
SMb20374	-16,1/-16,2	22-30/23-31	(-10/-2)(-10/-2)	Proteína periplásmica	Carbohidratos/dicarboxilatos
SMb20503	-15,4/-15,5	20-30/22-31	(-15/-5)(-15/-6)	Proteína periplásmica	Carbohidratos
SMb20295	-14,7/-15,8	23-30/24-32	(-12/-5)/(-13/-5)	Proteína periplásmica	Carbohidratos/dicarboxilatos
SMb20442	-14,7/-16,0	20-30/22-31	(-14/-4)(-14/-5)	Proteína periplásmica	Carbohidratos/dicarboxilatos
SMb20931	-14,2/-17,0	21-30/23-34	(-10/-1)(-13/-2)	Proteína periplásmica	Carbohidratos
SMb20144	-14,2/-14,3	21-30/22-31	(-8/+1)(-8/+1)	Proteína periplásmica	Péptidos
lacF (SMb21653)	-14,1/-14,2	48-57/50-59	(-9/+1)(-9/+1)	Permeasa	Carbohidratos
SMa0271	-20,1/-17,8	21-31/22-31	(-12/-2)(-11/-2)	Permeasa	Carbohidratos
SMa1438	-16,7/-14,0	23-33/24-34	(-14/-4)(-15/-4)	Proteína periplásmica	Péptidos
SMa0800	-16,6/-16,7	21-30/22-31	(-12/-3)(-12/-3)	Permeasa	Espermidina/putrescina
SMa1650	-15,1/-15,3	2-30/22-31	(-13/+18)(-13/-4)	Permeasa	Amino ácidos/péptidos
SMa0495	-14,9/-16,0	48-80/50-86	(-32/+2)(-34/+2)	Proteína periplásmica	Amino ácidos
SMa1339	-14,2/-18,0	21-30/20-33	(-14/-5)(-16/-3)	Permeasa	Amino ácidos/péptidos
SMa0709	-14,2/-13,4	20-29/22-30	(-9/+1)(-9/-1)	Proteína periplásmica	Carbohidratos
SMa0273	-14,1/-13,2	20-29/22-30	(-9/+1)(-9/-1)	Proteína periplásmica	Carbohidratos
matP (SMa0157)	-13,8/-13,9	21-30/22-31	(-10/+1)(-10/+1)	Proteína periplásmica	Malonato
SMa0079	-13,7/-14,3	23-30/24-32	(-15/-8)(-16/-8)	Permeasa	Amino ácidos

<sup>1</sup> Respecto al AUG anotado

**Tabla III-3. Candidatos exclusivos de Smr15C1 y Smr15C2.**

<b>mRNA candidato</b>	<b>Energía (kcal/mol)</b>	<b>Posición sRNA</b>	<b>Posición mRNA<sup>1</sup></b>	<b>Componente transportador ABC</b>	<b>Sustrato</b>
<b><u>Smr15C1</u></b>					
aglF (SMc03062)	-16,2	24-31	-14/-7	Proteína periplásmica	α-glucósidos
aglE (SMc03061)	-15,78	24-31	-13/-6	Proteína periplásmica	α-glucósidos
SMc02417	-14,98	48-74	-32/-6	Proteína periplásmica	Péptidos
SMb21432	-16,2	48-74	-31/-1	Proteína periplásmica	Hierro
SMb21149	-16,0	22-31	-8/+2	Permeasa	Carbohidratos
SMb20181	-15,2	21-31	-11/+1	Proteína periplásmica	Desconocido
SMb21604	-14,4	19-54	-59/-2	Proteína periplásmica	Carbohidratos
SMb20538	-14,1	24-31	-12/-5	Proteína periplásmica	Amino ácidos/péptidos
oppA (SMb21196)	-14,1	21-31	-16/-5	Proteína periplásmica	Amino ácidos/péptidos
SMA1337	-16,5	24-31	-13/-6	Proteína periplásmica	Amino ácidos/péptidos
SMA0522	-16,0	24-32	-14/-6	Proteína periplásmica	Desconocido
SMA1717	-15,0	24-32	-13/-5	Permeasa	Xantina/uracilo
SMA2000	-13,2	21-29	-10/-2	Proteína periplásmica	Carbohidratos
<b><u>Smr15C2</u></b>					
SMc02512	-22,6	2-55	-33/+18	Permeasa	Xantina/Uracilo
thiQ (SMc03871)	-18,0	10-37	-40/-8	Proteína de unión a ATP	Tiamina/espermidina/putrescina
SMc03121	-14,7	22-33	-14/-3	Proteína periplásmica	Amino ácidos/péptidos
SMc02033	-15,0	23-31	-11/-3	Proteína periplásmica	Carbohidratos
SMc00245	-14,2	51-80	-38/-10	Proteína de unión a ATP	Carbohidratos
exsE (SMb20937)	-16,7	9-36	-22/+4	Unión a ATP+permeasa	Carbohidratos
ttpD (SMb20124)	-15,8	22-34	-13/+1	Proteína de unión a ATP	Péptidos
SMb20720	-14,2	23-34	-15/-3	Proteína periplásmica	Carbohidratos
SMb20373	-13,9	22-32	-13/-3	Proteína periplásmica	Dicarboxilatos
agpC (SMb21645)	-13,8	22-32	-18/-8	Permeasa	Péptidos
SMb20661	-13,7	25-33	-14/-6	Proteína de unión a ATP	Carbohidratos
SMA0104	-13,1	42-59	-13/-4	Proteína periplásmica	Amino ácidos/Péptidos



**III-3.1.5.2. Evidencias experimentales**

Para la verificación experimental de estas predicciones se integraron los resultados de tres abordajes; *i*) estudio *in vivo* de una selección de las interacciones sRNA-mRNA predichas por IntaRNA, utilizando el sistema de doble plásmido según el protocolo descrito en el “Diseño Experimental” de este capítulo, *ii*) considerando que Smr15C1 y Smr15C2 son dos sRNAs con gran afinidad por Hfq, entre los mRNAs predichos por IntaRNA se identificaron aquellos diferencialmente expresados en el mutante *S. meliloti* 1021 $\Delta$ hfq (i.e. datos transcriptómicos y proteómicos descritos en el Capítulo 2) y los representados en el CoIP-RNA sometido a secuenciación masiva, y *iii*) dado que la mayoría de los mRNAs predichos por IntaRNA codifican el componente periplásmico de los transportadores ABC se analizó el efecto de la sobreexpresión constitutiva de ambos sRNAs sobre el perfil de proteínas periplásmicas de los mutantes *S. meliloti* 1021 $\Delta$ smr15C1 y 1021 $\Delta$ smr15C2.

La Tabla III-4 recopila todas estas evidencias experimentales, que se describen con más detalle a continuación.

**Tabla III-4. Evidencias experimentales para la verificación de las predicciones de IntaRNA.**

	Transcriptómica / Proteómica Hfq	CoIP-RNA <sup>1</sup>	Ensayo doble plásmido
<b>Smr15C1/Smr15C2</b>			
rhaS (SMc02324)	no	no	-
SMc04135	no	Sal	ND
SMc02471	no	Sal	ND
ugpB (SMb20416)	no	no	ND
SMb20374	no	no	ND
SMb20503	no	no	ND
SMb20295	no	Sal / Est	ND
SMb20442	si (1,03)	Sal / 40°	ND
SMb20931	no	no	ND
SMb20144	si (-1,27)	Sal / 20°	ND
lacF (SMb21653)	no	no	ND
SMA0271	no	no	ND
SMA1438	no	no	ND
SMA0800	no	no	ND
SMA1650	no	no	+(15C2)
SMA0495	no	Sal / 40°	-
SMA1339	no	no	ND
SMA0709	no	no	ND
SMA0273	no	no	ND
matP (SMA0157)	no	Est / Sal	ND
SMA0079	no	no	ND
<b>Smr15C1</b>			
agIF (SMc03062)	si (-1,22)	Log / 20°	ND
agIE (SMc03061)	si (-3,94)	Log / 20°	ND
SMc02417	no	no	-

**Tabla III-4. Evidencias experimentales para la verificación de las predicciones de IntaRNA (continuación).**

SMb21432	no	no	-
SMb21149	no	no	ND
SMb20181	no	Log	ND
SMb21604	no	Log	ND
SMb20538	no	Est	ND
oppA (SMb21196)	no* oppB y oppC (1,98/1,82)	Log / Est	-
SMa1337	no	no	-
SMa0522	no	no	ND
SMa1717	no	no	ND
SMa2000	no	no	ND
<b>Smr15C2</b>			
SMc02512	no	no	ND
thiQ (SMc03871)	no	no	ND
SMc03121	no	Log / Est	-
SMc02033	no	no	ND
SMc00245	no	Log / Est	-
exsE (SMb20937)	no	no	-
ttpD (SMb20124)	no	no	-
SMb20720	no	Log	ND
SMb20373	no	no	ND
agpC (SMb21645)	no	no	ND
SMb20661	no	no	ND
SMa0104	no	Log / Est	ND

<sup>1</sup> Representación de cada mRNA en los bancos de CoIP RNA analizados en este trabajo y/o en tres más obtenidas en otras condiciones biológicas (Becker, A., Schlüter, J.P., sin publicar). Log, fase exponencial; Est, fase estacionaria; 20°C, estrés por bajas temperaturas; 40°C, estrés por altas temperaturas; Sal, shock con 400 mM de NaCl. ND, no disponible. El valor M en datos transcriptómicos se indica entre paréntesis en el caso que corresponda.

#### III-3.1.5.2.1. Ensayo de doble plásmido

Dado el enorme tiempo que consume y lo laborioso de la generación de las construcciones plasmídicas necesarias para el ensayo de las interacciones sRNA-mRNA *in vivo*, se seleccionaron para su estudio mediante el mismo 11 de los 46 mRNAs de la lista priorizada de candidatos generada por IntaRNA. Para esta selección se asumió que las interacciones sRNA-mRNA más favorecidas serían las que involucran a secuencias nucleotídicas más extensas en ambas moléculas. Estos mRNAs se relacionan en la Tabla III-5 junto con los datos más relevantes de las regiones 5' de los mismos clonadas en los plásmidos reporteros pR\_GFP o pRO\_GFP (de ahora en adelante plásmidos diana por simplicidad) y de la fluorescencia de las fusiones en ausencia del sRNA.

Tabla III-5. Características relevantes de los plásmidos diana para Smr15C1/C2 utilizados en el ensayo *in vivo*.

mRNA	Nombre del plásmido	Coordenadas genómicas del inserto <sup>1</sup>	Codón fusionado <sup>2</sup>	Fluorescencia en placa <sup>3</sup>
<i>rhaS</i> (SMc02324)	pR_RhaS::GFP	694991-695084	7	-
<i>SMa1650</i>	pRO_SMa1650::GFP	922707-922799	20	+
<i>SMa0495</i>	pR_SMa0495::GFP	270195-270117	17	+++
<i>SMc02417</i>	pR_SMc02417::GFP	2788857-2789229	26	-
<i>SMb21432</i>	pR_SMb21432::GFP	1362964-1362942	24	-
<i>oppA.1</i> (SMb21196)	pR_OppA.1::GFP	944981-945018	26	+++
<i>oppA.2</i> (SMb21196)	pR_OppA.2::GFP	944769-945018	26	+++
<i>SMa1337</i>	pR_SMa1337::GFP	731200-731337	14	-
<i>SMc03121.1</i>	pR_SMc03121.1::GFP	3215938-3216094	12	+++
<i>SMc03121.2</i>	pR_SMc03121.2::GFP	3216041-3216094	12	+++
<i>SMc00245.1</i>	pR_SMc00245.1::GFP	1813922-1813690	25	+
<i>SMc00245.2</i>	pR_SMc00245.2::GFP	1814006-1813690	25	++
<i>exsE</i> (SMb20937)	pR_ExsE::GFP	1161888-1161991	7	+++
<i>ttpD</i> (SMb20124)	pRO_TtpD::GFP	139885-139986	15	-

<sup>1</sup> Nombre del mRNA (gen) y coordenadas según la anotación genómica de *S. meliloti* 1021.

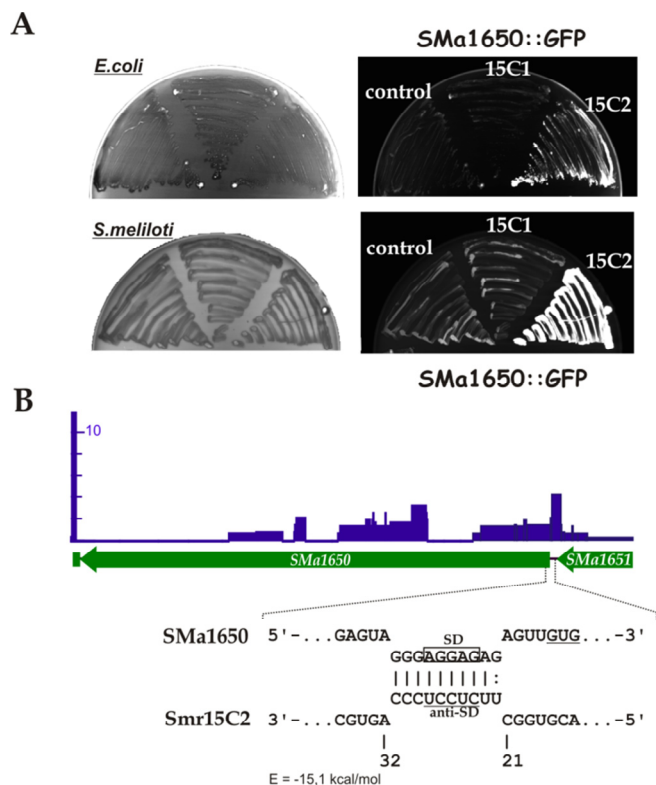
<sup>2</sup> Extremo 5' del fragmento del gen diana en relación al ATG anotado. Los sitios de inicio de la transcripción de cada mRNA fueron facilitados por A. Becker (no publicados).

<sup>3</sup> Intensidad de fluorescencia detectada: +++, muy intensa; ++, intensidad media; +, intensidad baja; -, fluorescencia no detectada

Los ensayos se realizaron en ausencia de expresión de Smr15C1 y Smr15C2 desde sus loci cromosómicos en *E. coli* DH5 $\alpha$  y el mutante *S. meliloti* 1021 $\Delta$ smr15C1-C2. En ambos fondos genéticos cada uno de los plásmidos diana se enfrentó independientemente a pSRK*smr15C1* y pSRK*smr15C2*, que expresan constitutivamente sendos sRNAs, y al control pSRK sin inserto. Cada uno de los dobles transformantes (i.e. 42 por cada cepa ya que de *oppA*, *SMc03121* y *SMc00245* se testaron dos posibles regiones 5' de diferente longitud) se sembró en placas de TY (*S. meliloti*) o LB (*E. coli*) que, una vez crecidas las bacterias, se observaron en escaner de fluorescencia.

No se observaron cambios visibles en la fluorescencia basal de las fusiones a GFP clonadas en los plásmidos diana cuando se co-expresaron con los sRNAs. La única excepción fue pRO\_SMa1650::GFP que mostró una alteración en el patrón de fluorescencia en presencia de pSRK*smr15C2*, resultando en un aumento evidente de la señal en las placas en tres de un total de seis ensayos. El mismo efecto fue también visible en células de *E. coli* DH5 $\alpha$  (Fig. III-21A). Estos resultados sugieren que la traducción del mRNA del gen *SMa1650* es activada específicamente por la expresión de Smr15C2 y que los posibles factores del huésped necesarios para esta activación

(e.g. proteína Hfq) no son particularmente específicos y pueden ser complementados en fondos genéticos de bacterias filogenéticamente distantes como *E. coli*. En cualquier caso, dada la variabilidad de los ensayos, la demostración inequívoca de la interacción SMa1650-Smr15C2 requerirá de la cuantificación del efecto observado en este ensayo (e.g. mediante Western blot con anticuerpos anti-GFP o citometría de flujo para la determinación de la fluorescencia) o de otras aproximaciones experimentales complementarias.



**Fig. III-21. Interacción de Smr15C1/C2 con SMa01650.** A) Ensayo de doble plásmido *in vivo* en *E.coli* (sección superior) y *S. meliloti* (sección inferior) de la región 5'UTR del mRNA SMa1650 con los plásmidos que expresan Smr15C1 y Smr15C2. Control: plásmido vacío. B) Diagrama correspondiente al enriquecimiento de este transcrito en muestras de CoIP-RNA en donde se señala la región en la que se predice la interacción sRNA-mRNA. Los nucleótidos implicados en la interacción y los motivos Shine-Dalgarno (SD), anti-Shine-Dalgarno e inicio de traducción (GUG) también se indican

### III-3.1.5.2.2 Evidencias basadas en la dependencia de Hfq

Es previsible que una proporción importante de los mRNAs/proteínas diferencialmente expresados en el mutante *S. meliloti* 1021 $\Delta$ hfq estén ribo-regulados. Por otra parte, asumiendo que en apareamientos mRNA-sRNA dependientes de Hfq la proteína interacciona con ambas moléculas del híbrido, el mapeo de las lecturas generadas por secuenciación masiva del CoIP-RNA podría aportar evidencias sobre la identidad de los mRNA diana de estos sRNAs así como una “imagen” de los sitios preferentes de interacción de Hfq con los mismos.

*SMa1650* codifica la permeasa de membrana de un transportador ABC probablemente implicado en la adquisición de aminoácidos y péptidos, siendo el segundo gen en el operón tras *SMa1651*, que codifica la proteína periplásmica del sistema. IntaRNA predice una interacción entre Smr15C2 y *SMa1650* que involucra una secuencia de 10 nt en ambas moléculas. En el sRNA la secuencia de apareamiento está en el bucle de la horquilla estructural SL1 (que contiene un motivo aSD) que es complementaria a las posiciones -14 a -5 (respecto a la AUG anotada) en el mRNA de *SMa1650* (Fig. III-21B). Aunque ninguno de los genes de este operón, ni las proteínas codificadas por éstos aparecen diferencialmente expresados en el mutante *S. meliloti*  $\Delta hfq$ , *SMa1651* apareció representado en CoIP-RNA obtenido de bacterias sometidas a estrés salino (A. Becker, J.P. Schlüter, comunicación personal) (Fig. III-21B). La región del transcrito con más cobertura en la secuenciación (i.e. más enriquecida en lecturas respecto al control) fue la secuencia de 13 nt localizada entre los genes *SMa1651* y *SMa1650*, donde se predice la interacción con Smr15C2. El efecto positivo de esta interacción sobre la traducción de *SMa1650* observado en el ensayo *in vivo* sería inicialmente explicable si en ausencia del sRNA fuera una horquilla inhibitoria en la región 5' de *SMa1650* la que impidiera la traducción del mensajero. Sin embargo, la estructura secundaria de la región *SMa1651*-*SMa1650* en el contexto del mRNA policistrónico es difícil de predecir, por lo que la dilucidación de esta cuestión debe ser abordada mediante nuevas aproximaciones.

Otra diana predicha por IntaRNA para Smr15C1 y Smr15C2 fue el mRNA del gen *SMa0495*, que codifica la proteína periplásmica de un transportador ABC potencialmente implicado en el transporte de aminoácidos. Este gen no aparece tampoco diferencialmente expresado en *S. meliloti* 1021 $\Delta hfq$  pero su mensajero está enriquecido en CoIP-RNA preparado a partir de lisados de bacterias sometidas a shock osmótico y térmico (Fig. III-22). La mayoría de lecturas correspondientes a este mRNA mapeó en su región 5', desde el sitio de inicio de transcripción hasta los primeros nucleótidos de la secuencia codificante. La interacción de Smr15C1/Smr15C2 con *SMa0495* predicha por IntaRNA es particularmente extensa e involucra una región de 33 nt en Smr15C1 (posiciones 48-80) y 37 nt en Smr15C2 (posiciones 50-86), que comprende, en ambos transcritos, la segunda horquilla estructural SL2 y las dos regiones de nucleótidos desapareados que la flanquean. En ambos casos se establecerían apareamientos discontinuos con la secuencia nucleotídica localizada entre las posiciones -34 y +2 respecto al codón de inicio de traducción anotado para *SMa1650*, precisamente la zona mayoritariamente cubierta por la secuenciación del CoIP-RNA.

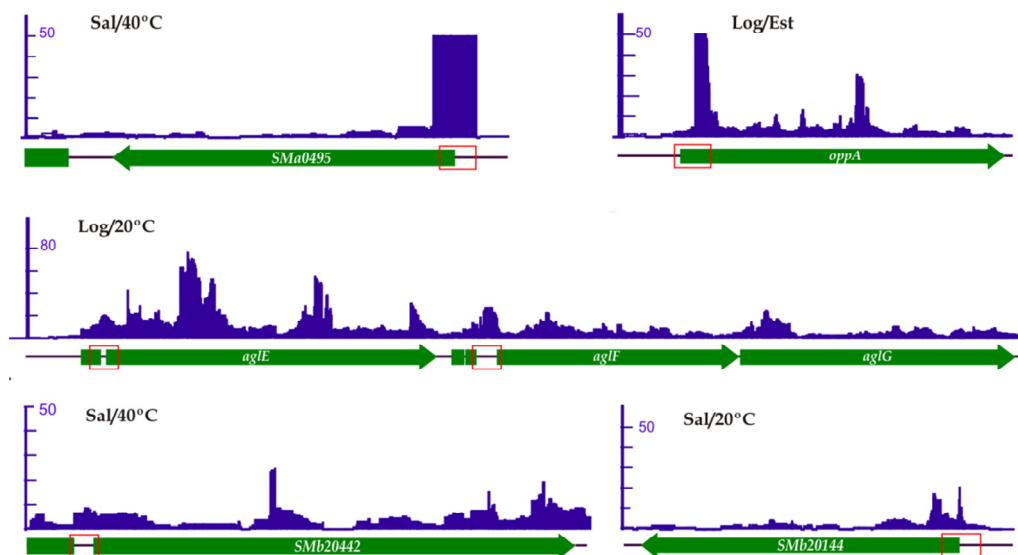


Fig. III-22. Enriquecimiento en librerías de CoIP-RNA de mRNAs diana predichos para Smr15C1/C2. Diferentes mRNAs diana candidatos para Smr15C1/C2. Los ejes verticales representan el enriquecimiento de cada posición nucleotídica en el CoIP-RNA de la cepa *hfq*<sup>FLAG</sup> respecto a la cepa silvestre control. La condición biológica de obtención del CoIP-RNA se indica en cada caso, así como la región de interacción de los sRNAs con cada uno de los mRNAs predicha por IntaRNA (encuadrada en rojo).

En otros cinco mRNAs predichos por IntaRNA como dianas de Smr15C1 y/o Smr15C2 coexisten dos evidencias experimentales que apoyan la predicción; están diferencialmente expresados en *S. meliloti* 1021 $\Delta$ *hfq* y representados en el CoIP-RNA obtenido en diversas condiciones (i.e. shocks osmótico y térmico, bacterias en fase exponencial y estacionaria de crecimiento). Estos mRNAs son los correspondientes a los genes *SMb20144*, *oppA* (*SMb21196*), *aglF* (*SMc03062*), *aglE* (*SMc03061*) y *SMb20442*, que codifican sendas proteínas periplásmicas de transportadores ABC con afinidad por aminoácidos y péptidos (*SMb20144*, *oppA*),  $\alpha$ -glucósidos (*aglEF*) y dicarboxilatos (*SMb20442*). *SMb20442* apareció sobre-representado ( $M=1,03$ ) en el mutante *S. meliloti* 1021 $\Delta$ *hfq*, mientras que *SMb20442*, *aglF* y *aglE* fueron reprimidos (valores de  $M$ , -1,27, -1,22 y -3,94, respectivamente). Sin embargo, *oppA* no apareció entre los genes diferencialmente expresados en ausencia de Hfq en los experimentos presentados en el Capítulo 2 pero sí en los publicados por Gao *et al.* (2010), que describieron una discreta inducción de este gen en el mutante *S. meliloti* 1021 $\Delta$ *hfq* ( $M=0,61$ ). No obstante, los genes que junto a *oppA* especifican este transportador ABC de aminoácidos y péptidos, *oppB* y *oppC*, sí aparecieron significativamente inducidos en nuestros ensayos (valores de  $M$  1,98 y 1,82, respectivamente). IntaRNA predice interacciones termodinámicamente favorecidas para Smr15C1 y Smr15C2 con los

mensajeros de *SMb20442* y *SMb20114* y específicamente de *Smr15C1* con los de *aglF*, *aglE* y *oppA*. En todos los casos los apareamientos involucran secuencias nucleotídicas cortas de 7-10 nt en ambas moléculas; localizados en el bucle de la horquilla estructural SL1 de los sRNAs y solapando siempre el RBS, muy próximos al codón de inicio de la traducción AUG en los mRNA diana (Fig. III-22). La cobertura de estos transcritos por las lecturas de la secuenciación del CoIP-RNA fue heterogénea, si bien entre las regiones más representadas por los cDNAs siempre se encontraron las regiones 5' de cada mRNA potencialmente implicadas en la interacción con *Smr15C1* y/o *Smr15C2* (Fig. III-22).

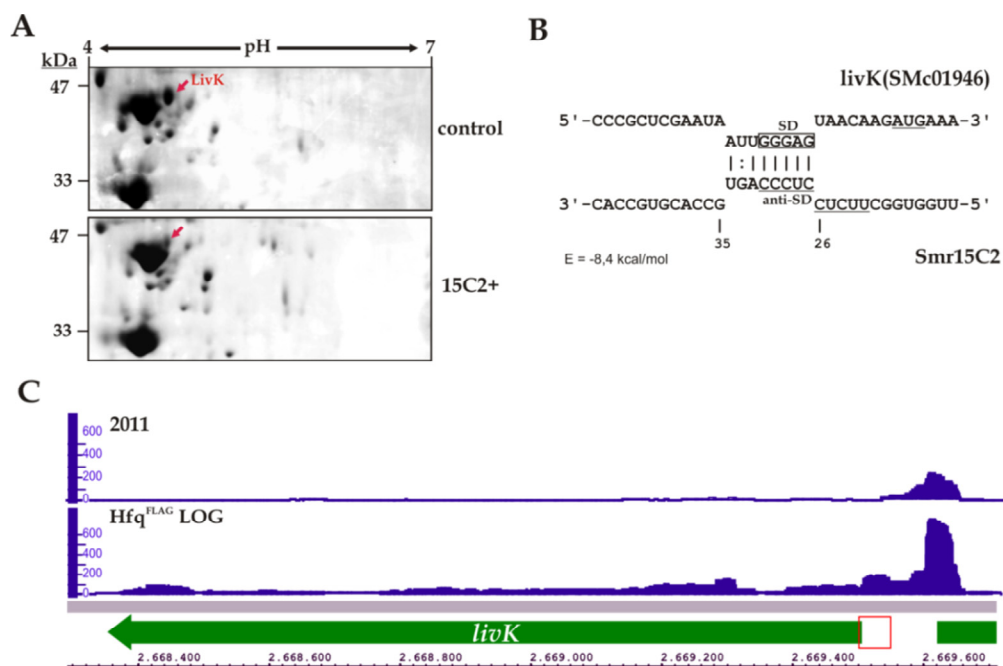
#### III-3.1.5.2.3 Alteración de las proteínas periplásmicas por expresión de *Smr15C1/C2*

Como abordaje experimental independiente para la identificación de mRNAs diana de *Smr15C1* y *Smr15C2* se compararon los perfiles de proteínas periplásmicas de las cepas *S. meliloti* 1021 $\Delta$ *smr15C1* (pSRK<sub>*smr15C1*</sub>) y 1021 $\Delta$ *smr15C2* (pSRK<sub>*smr15C2*</sub>), que expresan constitutivamente cada sRNA desde pSRK en ausencia de sus respectivos *loci* cromosómicos, con los de los mutantes *S. meliloti* 1021 $\Delta$ *smr15C1* y 1021 $\Delta$ *smr15C2* transformados con el plásmido vacío pSRK como control. Cantidades equivalentes de proteínas periplásmicas de todas estas cepas, obtenidas en fase logarítmica de crecimiento en medio TY, se resolvieron inicialmente en geles SDS-PAGE al 10% de acrilamida/bisacrilamida. La comparación de los patrones proteicos no reveló diferencia alguna entre las cepas analizadas, por lo que todas las muestras se resolvieron posteriormente en geles 2D de acrilamida/bisacrilamida al 12%. Las proteínas diferencialmente expresadas entre las cepas control y las que expresan constitutivamente cada uno de los sRNAs en estudio (~5 en cada caso) fueron identificadas mediante MALDI-TOFF. Como hallazgo más relevante de este análisis se encontró la proteína LivK (PM 39 kDa y pI 5) reprimida específicamente por la expresión de *Smr15C2*. La sección más informativa de estos geles se muestra en la Fig. III-23A. LivK, es la proteína periplásmica de unión a sustrato del transportador ABC de leucina, codificado por el operón *livKFGMH*. Todo el mRNA policistrónico de este operón se encontró fuertemente inducido en el mutante *S. meliloti* 1021 $\Delta$ *hfq*, (valores de M 1,88, 2,55, 1,90, 1,60 y 4,99 para los mensajeros de *livK*, *livF*, *livG*, *livM* y *livH*, respectivamente), mientras que en el mutante 2011 $\Delta$ *hfq* la proteína LivK también apareció sobreexpresada respecto a sus niveles en la cepa silvestre. Estos resultados pueden ser explicados por la degradación demostrada de *Smr15C2* en ausencia de *Hfq*.

En nuestra búsqueda inicial IntaRNA no predijo ninguna interacción entre *Smr15C1* y el mRNA de *livK*. Sorprendentemente, el análisis reverso suministrando

al programa la secuencia de Smr15C1 y la de *livK* e imponiendo parámetros sub-óptimos a la interacción, identificó un posible apareamiento entre sRNA y mRNA con energía libre calculada de -8 kcal/mol (Fig. III-23B). Éste implica a ocho nucleótidos en ambas moléculas localizados en la horquilla estructural SL1 entre las posiciones 27 y 34 de Smr15C2 y entre las posiciones -8 y -15 respecto al AUG del mRNA de *livK*, que incluye totalmente el RBS. Esta misma región del mRNA aparece también particularmente enriquecida en CoIP-RNA obtenido de bacterias en fase exponencial de crecimiento (Fig. III-23C).

En conjunto estos resultados sugieren que Smr15C1 tiene un efecto negativo sobre la estabilidad del mensajero de *livK* e inhibe la traducción del mismo por bloqueo de su RBS en un proceso que requiere la participación de Hfq.



**Fig. III-23. Evidencias de la regulación del mRNA de *livK* por Smr15C2.** **A)** Geles 2D de los extractos proteicos periplásmicos correspondientes a la cepa carente de Smr15C2 (control) y la que lo expresa desde el plásmido pSRK (15C2+). La proteína LivK se indica con una flecha. **B)** Interacción entre el mRNA de *livK* y el sRNA Smr15C2 predicha por IntaRNA. Los motivos Shine-Dalgarno, aSD y codón de inicio de la traducción AUG se indican sobre la secuencia. **C)** Número de lecturas correspondientes a la región de *livK* en el CoIP-RNA de la cepa control (2011) y 2011Hfq<sup>FLAG</sup> obtenido en la fase de crecimiento exponencial de las bacterias. La región en donde se localiza la interacción con Smr15C2 se indica encuadrada en rojo.



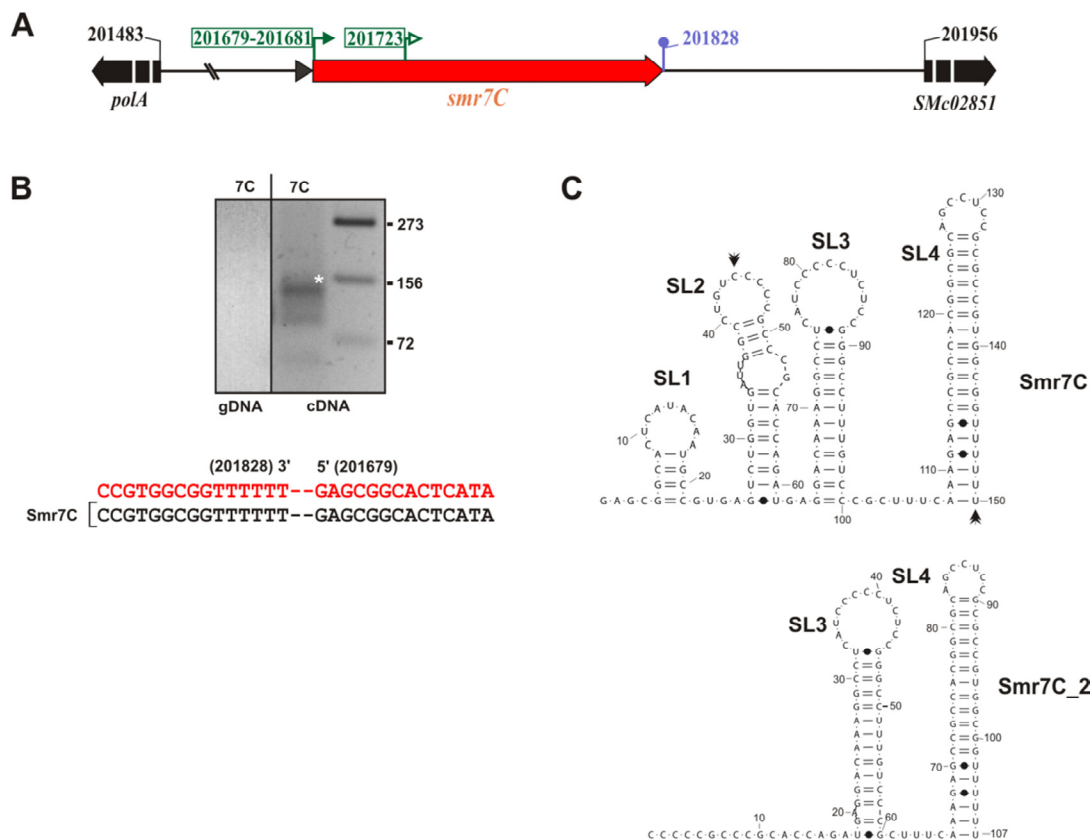
### III-3.2. LA FAMILIA DE sRNAs $\alpha r7$ : CARACTERIZACIÓN DE Smr7C

#### III-3.2.1. MAPEO DEL EXTREMO 3' Y ANÁLISIS *IN SILICO*

En *S. meliloti* 1021, el sRNA Smr7C está codificado en la RIG flanqueada por los genes *polA* y *SMc02851*, que codifican la DNA polimerasa y un regulador transcripcional MarR, respectivamente (Fig. III-24A). En este genoma este *locus* se presenta en copia única localizado en el cromosoma.

El RACE 5' identificó dos posibles inicios de transcripción para Smr7C que mapean en posiciones casi contiguas en el genoma; G 201.679 y G 201.681. Además del transcrito primario las hibridaciones Northern detectan una forma procesada del mismo cuyo extremo 5' fue mapeado en el residuo C 201.723. El extremo 3' de Smr7C se determinó también experimentalmente mediante la circularización del transcrito. La amplificación del cDNA circular con oligonucleótidos divergentes específicos generó una banda mayoritaria del tamaño esperado (150 nt) resuelta en geles de agarosa al 3% que no fue visible cuando se utilizó DNA genómico como molde (Fig. III-24B). Este fragmento fue clonado en pGEM-T Easy y el análisis de la secuencia de varios clones no reveló ningún signo de poliadenilación del extremo 3' del transcrito, mapeando mayoritariamente éste en la posición U 201.818 correspondiente al último residuo de Uridinas del terminador de transcripción independiente de Rho predicho *in silico* para este sRNA. La pareja de cebadores utilizados para la amplificación del DNA circular no permite la detección del extremo 5' de la forma procesada de Smr7C, por lo que el extremo 3' determinado corresponde al transcrito primario, cuyo extremo 5' fue de nuevo mapeado mediante este experimento en el residuo G201.679 del genoma de *S. meliloti* 1021.

El estudio de la estructura secundaria utilizando RNAfold reveló que la forma primaria de Smr7C podría plegar formando cuatro horquillas bien diferenciadas (**SL1**, **SL2**, **SL3** y **SL4**) que dejan tres regiones muy cortas de cadena sencilla situadas, respectivamente, en el extremo 5', entre las horquillas SL1 y SL2 y entre SL3 y SL4, siendo esta última la más extensa con 7 nt (Fig. III-24C). En Smr7C no aparecen secuencias aSD expuestas en los bucles de las horquillas estructurales, ni motivos de unión a Hfq reconocibles en las regiones desapareadas. La última horquilla (SL4) corresponde al terminador de la transcripción independiente de Rho cuya cola poli-U final está totalmente apareada formando parte del tallo de SL4 (Fig. III-24C). La estructura secundaria predicha para la forma procesada del transcrito (107 nt; de ahora en adelante Smr7C\_2) comparte con el transcrito primario las dos últimas horquillas (SL3 y SL4), mientras que su región 5' es una secuencia de 17 nt desapareados (Fig. III-24C).



**Fig. III-24. El sRNA Smr7C de *S. meliloti*.** A) RIG donde se localiza *smr7C*, indicándose los genes flanqueantes y las coordenadas de los inicios de transcripción mapeados mediante RACE 5' (encuadrados en verde) y los extremos 3' (última U del terminador independiente de Rho). B) RACE 3' de los productos de amplificación sobre DNA genómico (izda.) y cDNA circular (derecha), indicando las bandas de tamaño esperado con asteriscos. Secuencia de los clones (inferior) secuenciados tras el clonaje de los productos de PCR anteriores, indicando en rojo la anotación correspondiente al genoma de *S. meliloti*. C) Estructuras secundarias predichas por VARNA para los sRNAs Smr7C y la forma procesada Smr7C\_2, indicando la última U de la cola poli-U del terminador (flecha doble). Abreviaturas: SL, dominio o stem-loop.

Cuando la base de datos Rfam fue interrogada con esta estructura no aparecieron homólogos reconocibles, por lo que Smr7C puede considerarse un sRNA no descrito hasta ahora en bacterias.

### III-3.2.1.1. Conservación y distribución de Smr7C en el orden Rhizobiales

El alineamiento estructural (Stockholm) de los homólogos de Smr7C para la generación mediante Infernal 1.0 del MC que identifica a la familia de sRNAs  $\alpha 7$  se presenta en la Fig. III-25A. Del mismo se infiere una estructura secundaria consenso

sostenida por un alto grado de co-varianza en los nucleótidos que forman los tallos de los motivos SL, a la que se ajusta efectivamente Smr7C (Fig. III-25B).

El modelo  $\alpha 7$  identificó 26 homólogos estructurales, todos ellos codificados por genes cromosómicos de copia única en genomas de  $\alpha$ -proteobacterias del orden Rhizobiales. El árbol derivado del agrupamiento estructural de esta familia de sRNAs se presenta en la Fig. III-26. En el mismo, la denominación de los sRNAs  $\alpha 7$  en cada especie bacteriana en la que están representados sigue una nomenclatura idéntica a la de los miembros de  $\alpha 15$ . Las especies de relación filogenética más próxima codifican también los homólogos más cercanos a Smr7C con valores de  $E$  en Infernal 1.0 más significativos ( $3.66E^{-49}$ - $2.93E^{-12}$ ) i.e. *S. medicae* y *S. fredii*, dos cepas de *R. leguminosarum trifolii* (WSM2304 and WSM1325), dos cepas de *R. etli* CFN 42 y CIAT 652, la cepa de referencia *R. leguminosarum* bv. *viciae* 3841, las especies de *Agrobacterium* (*A. vitis*, *A. tumefaciens*, *A. radiobacter* y *A. H13*) y las especies del género *Mesorhizobium* (*M. loti*, *M. ciceri* y *M. BNC*). El resto de las secuencias identificadas mediante el modelo son estructuralmente más distantes a Smr7C (valores de  $E$  entre  $1.58E^{-10}$  y  $2.88E^{-09}$ ) y se encontraron en los genomas de especies del género *Brucella* (*B. ovis*, *B. canis*, *B. abortus*, *B. microtis*, y algunos biovares de *B. melitensis*) y en *Ochrobactrum anthropi*.

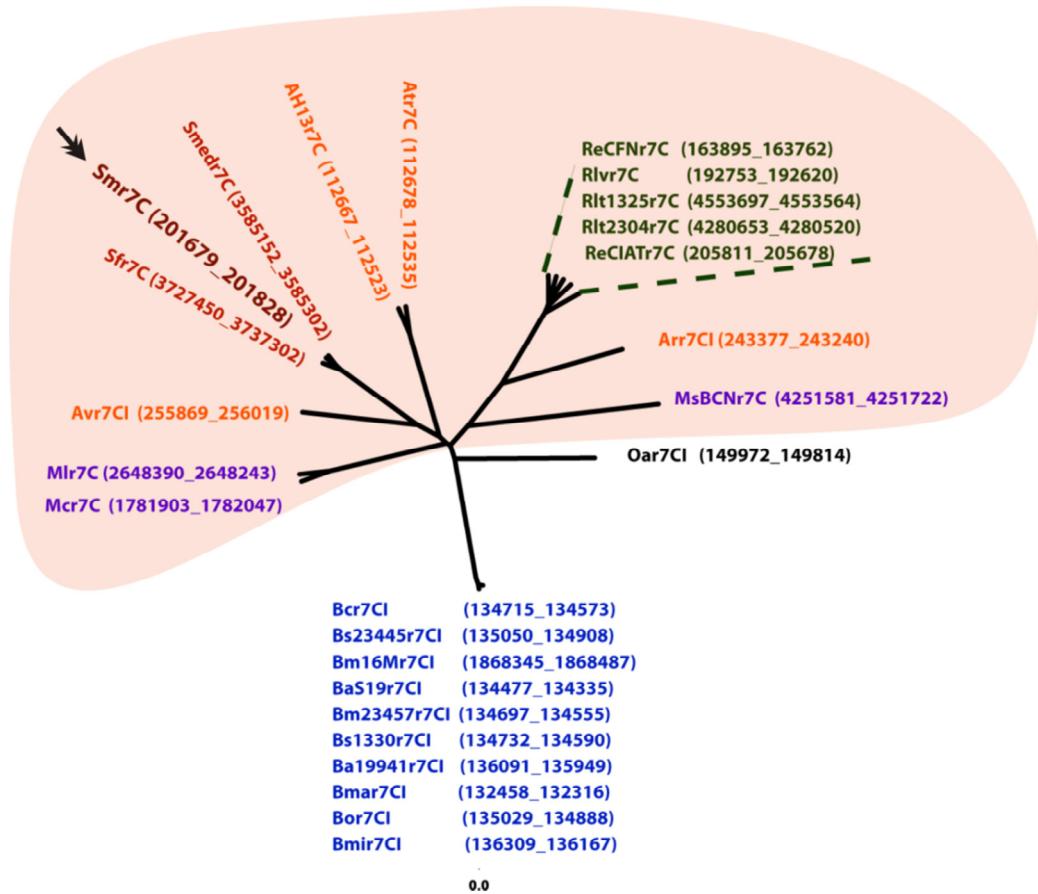
Por otra parte, las regiones genómicas de los sRNAs  $\alpha 7$  son totalmente sinténicas en especies de los géneros *Sinorhizobium*, *Rhizobium* y *Agrobacterium*, en las que los genes flanqueantes del *loci* sRNA cofican en todos los casos una DNA polimerasa y un regulador transcripcional de la familia MarR. La sintenia, sin embargo, es parcial con las mismas regiones genómicas en *Mesorhizobium* y *Brucella*, donde en lugar del regulador MarR, se identificó un gen que codifica una peptidasa.

### III-3.2.2. EXPRESIÓN DE Smr7C

#### **III-3.2.2.1. Cinética de expresión en medio TY**

La hibridación con un oligonucleótido específico para la detección de Smr7C de RNA total de *S. meliloti* 1021 obtenido a D.O. 0,3-0,6-1,2-2,4 de cultivos en medio TY (Fig. III-27A, parte izquierda), reveló una acumulación progresiva del transcrito, que alcanzó su máximo en fase estacionaria temprana (D.O. 1,2) para hacerse casi indetectable en la fase estacionaria tardía de crecimiento (D.O. 2,4). La razón entre las





**Fig. III-26. Relaciones filogenéticas entre los sRNAs de la familia  $\alpha 7$ .** Arbol filogenético correspondiente a los miembros identificados en *S. meliloti* (marcados con flechas dobles) y predichos en el resto de organismos de la familia  $\alpha 7$ . Leyenda: Smedr7C = *S. medicae* WSM419 cromosoma (NC\_009636), Smr7C = *S. meliloti* 1021 (NC\_003047), Sfr7C = *S. fredii* NGR234 cromosoma (NC\_012587), Atr7C = *A. tumefaciens* str. C58 cromosoma circular (NC\_003062), AH13r7C = *A. sp.* H13-3 cromosoma (NC\_015183), ReCIATr7C = *R. etli* CIAT 652 (NC\_010994), Arr7CI = *A. radiobacter* K84 cromosoma 1 (NC\_011985), Rlt2304r7C = *R. leguminosarum* bv. trifolii WSM2304 cromosoma (NC\_011369), Avr7CI = *A. vitis* S4 cromosoma 1 (NC\_011989), Rlvr7C = *R. leguminosarum* bv. viciae 3841 (NC\_008380), Rlt1325r7C = *R. leguminosarum* bv. trifolii WSM1325 (NC\_012850), ReCFNr7C = *R. etli* CFN 42 (NC\_007761), Mlr7C = *M. loti* MAF303099 cromosoma (NC\_002678), Mcr7C = *M. ciceri* biovar biserrulae WSM1271 cromosoma (NC\_014923), Bcr7CI = *B. canis* ATCC 23365 cromosoma I (NC\_010103), Bs23445r7CI = *B. suis* ATCC 23445 cromosoma I (NC\_010169), Bm16Mr7CI = *B. melitensis* bv. 1 str. 16M cromosoma I (NC\_003317), Ba519r7CI = *B. abortus* S19 cromosoma 1 (NC\_010742), Bm23457r7CI = *B. melitensis* ATCC 23457 cromosoma I (NC\_012441), Bs1330r7CI = *B. suis* 1330 cromosoma I (NC\_004310), Ba19941r7CI = *B. abortus* bv. 1 str. 9-941 cromosoma I (NC\_006932), Bmar7CI = *B. melitensis* biovar Abortus 2308 cromosoma I (NC\_007618), Bor7CI = *B. ovis* ATCC 25840 cromosoma I (NC\_009505), Bmir7CI = *B. microti*CCM 4915 cromosoma 1 (NC\_013119), Oar7CI = *O. anthropi* ATCC 49188 cromosoma 1 (NC\_009667), MsBCNr7C = *M. sp.* BNC1 (NC\_008254)

señales de hibridación de las formas Smr7C\_2 y Smr7C del transcrito fue ligeramente superior a uno hasta fase estacionaria temprada donde la acumulación de la forma procesada Smr7C\_2 se estimó en más del doble que la del transcrito primario.

### **III-3.2.2.2. Expresión en condiciones de estrés**

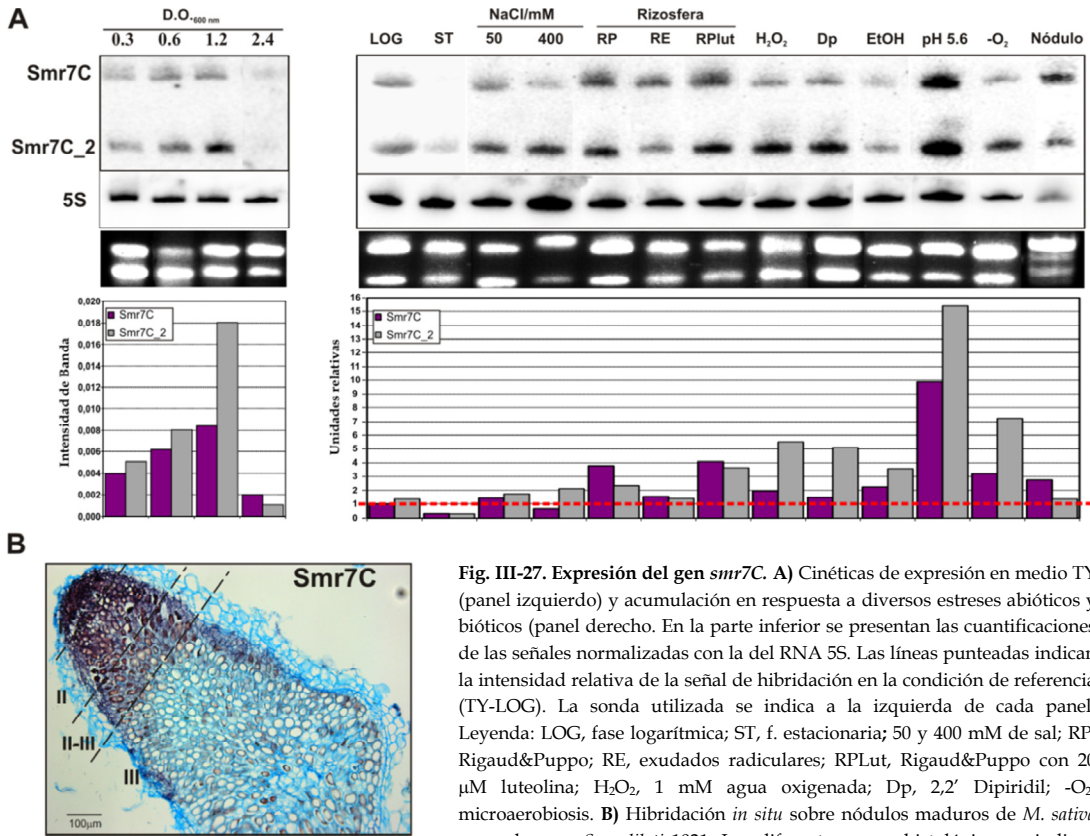
En una segunda hibridación se comparó la expresión de Smr7C en diversas condiciones de estrés biótico y abiótico tomando como referencia la expresión en fase exponencial de crecimiento en medio TY (ver "Diseño Experimental). Este experimento confirmó la casi ausencia del transcrito en fase estacionaria tardía de crecimiento, mientras que la presencia de 50 mM de NaCl en el cultivo no tuvo efecto aparente sobre la expresión. El shock osmótico, el estrés de membrana (EtOH), la endosimbiosis y la incubación de las bacterias en medio Rigaud y Puppo indujeron discretamente la expresión de alguna de las dos formas de Smr7C (2-4 veces) respecto a la referencia, aunque ni la presencia de plantas de alfalfa ni la adición de luteolina alteraron significativamente la acumulación de los transcritos. La acumulación de la forma procesada Smr7C\_2 fue más evidente cuando las bacterias se sometieron a estrés oxidativo, hambre de hierro y microaerobiosis (condiciones en las que la intensidad de la señal de hibridación correspondiente a este transcrito fue siempre superior en más de dos veces a la del transcrito primario), con niveles de inducción del orden de 5-7 veces respecto a la referencia. La máxima expresión de Smr7C ocurrió después de someter a las bacterias a estrés ácido, condición en la que las señales de hibridación fueron 10 (Smr7C) y 15 veces (Smr7C\_2) más intensas que las detectadas en RNA extraído en fase exponencial de crecimiento.

### **III-3.2.2.3. Expresión en nódulos maduros de alfalfa**

La expresión del transcrito Smr7C también se analizó en nódulos maduros de *M. sativa* mediante hibridación *in situ* con una ribosonda complementaria a la secuencia del transcrito primario. Este experimento reveló la presencia de Smr7C en la región apical de nódulo, correspondiente al meristemo (zona I), en la zona de invasión (II) y en la interzona II-III, mientras que en la zona de fijación de nitrógeno o zona III no se detectó señal de hibridación (Fig. III-27B).

### **III-3.2.2.4. Análisis de las regiones promotoras**

El alineamiento múltiple de las secuencias nucleotídicas de las regiones promotoras de los genes  $\alpha 7$  reveló en todos ellos promotores reconocibles dependientes de  $\sigma^{70}$  con motivos consenso -35/-10 CTTGAC-N17-CTATAT y la presencia de un motivo conservado (MEME) de 20 nt entre las posiciones -60 y -80



respecto al sitio de inicio de transcripción predicho para cada sRNA (marcado en naranja en la Fig. III-28). Sin embargo, este motivo no pudo asignarse a ningún factor de transcripción conocido. No obstante, esta secuencia conservada podría contribuir a la expresión diferencial de los miembros de esta familia de sRNAs similar a la observada en Smr7C.

### III-3.2.3. APROXIMACIÓN GENÉTICA A LA FUNCIÓN DE Smr7C

#### III-3.2.3.1. Generación y verificación experimental de los mutantes Smr7C

##### III-3.2.3.1.1. Deleción del gen *smr7C*

La eliminación de la copia cromosómica del gen *smr7C* y su sustitución por un marcador de resistencia a eritromicina (195 pb) se llevó a cabo de igual modo al

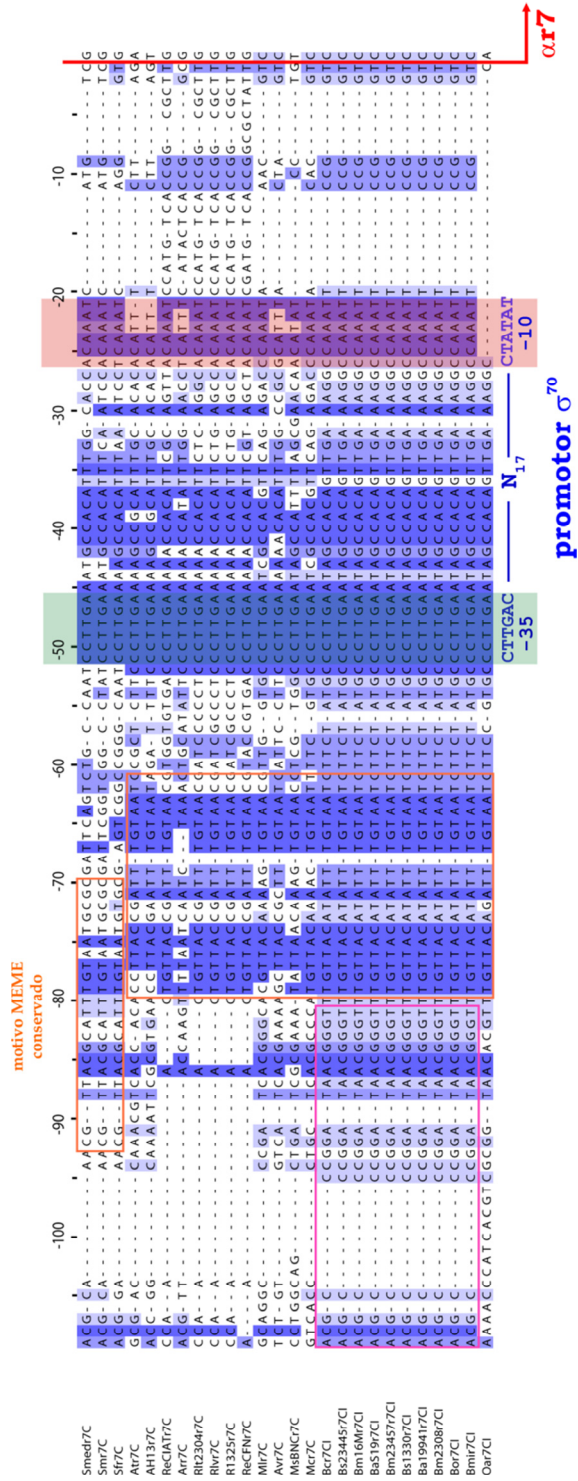
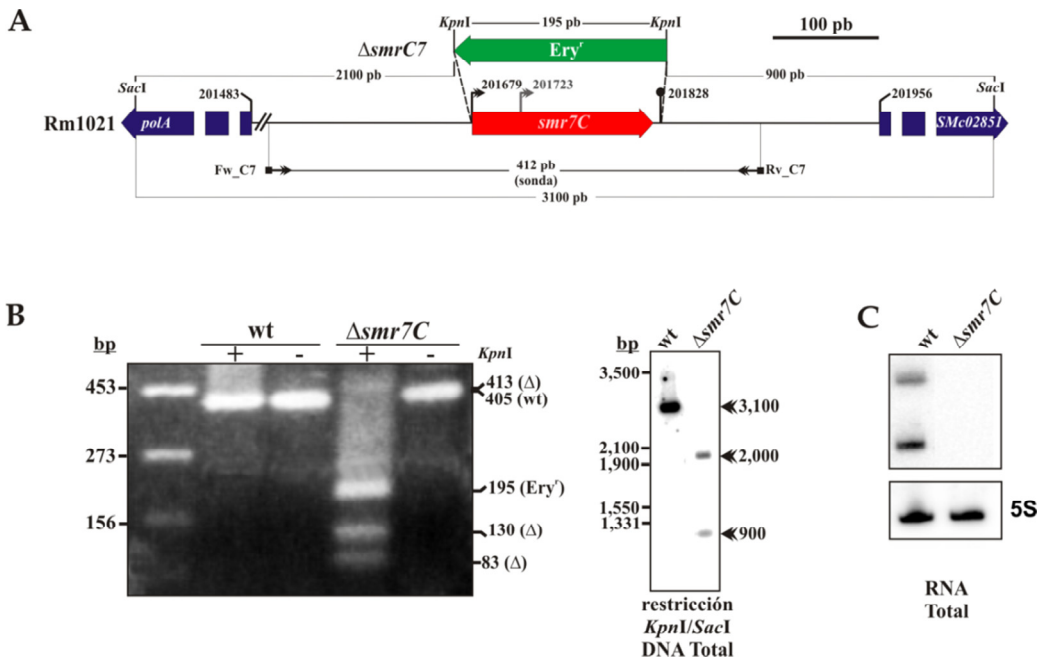


Fig. III-28. Alineamiento de las regiones promotora de los genes  $\sigma^{70}$ . Las cajas promotoras -10/-35 de los promotores  $\sigma^{70}$  aparecen marcadas en rosa y verde respectivamente. Los motivos conservados para factores de transcripción se indican encuadrados junto con su secuencia consenso (logs) en cada caso.



descrito para los genes *smr15C*. La correcta generación de la mutación fue también verificada mediante PCR, Southern y Northern blot (Fig. III-29). La amplificación de DNA genómico en lisados de las cepas salvaje *S. meliloti* 1021 y mutante 1021 $\Delta$ *smr7C* con los cebadores Fw\_C7/Rv\_C7 generó en ambos casos una banda del tamaño esperado (405 pb y 413 pb, respectivamente). El alelo mutado fue sensible a la digestión con *KpnI*, que generó el fragmento de 195 pb correspondiente al marcador y fragmentos genómicos de los tamaños esperados (180 y 83 pb) (Fig. III-29B, sección izquierda).



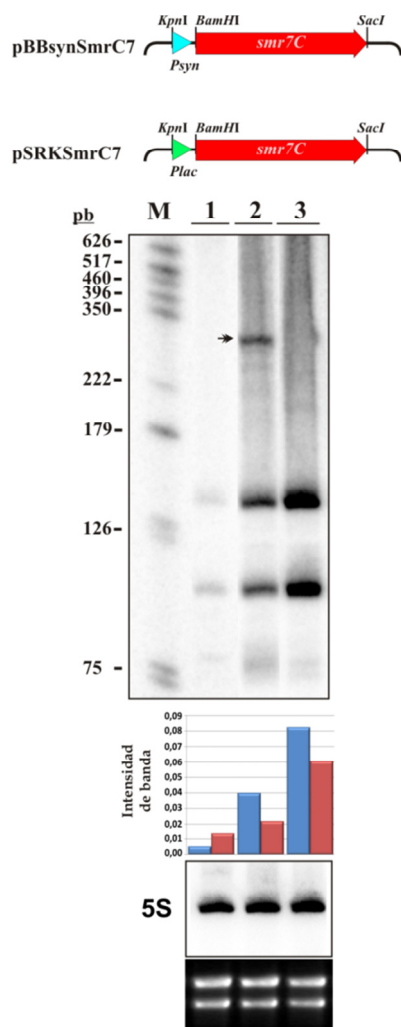
**Fig. III-29. Verificación experimental de la mutación  $\Delta$ *smr7C*.** A) Región genómica de *smr7C* en *S. meliloti* 1021 en donde se indican los mutante generado ( $\Delta$ *smr7C*) con la posición del *cassette* de eritromicina (Ery) que los sustituye en cada caso, además de información necesaria para la interpretación del resto de los paneles. B) Geles de agarosa correspondientes a la verificación por PCR (paneles situados a la izquierda) y Southern blot (paneles situados a la derecha) del mutante anterior. En la esquina superior derecha de cada Southern se indica la sonda de hibridación. El tamaño y el origen de las bandas se indica en cada caso (wt, cepa salvaje;  $\Delta$ , mutante; Ery, *cassette* de eritromicina). C) Hibridación Northern sobre RNA total utilizando la sonda para *Smr7C*.

El DNA genómico de las cepas 1021 y 1021 $\Delta$ *smr7C* se digirió con *KpnI-SacI* y se hibridó con el fragmento generado por amplificación de DNA silvestre con los oligonucleótidos Fw\_C7/Rv\_C7. El patrón de hibridación fue compatible con el esperado en ambos casos detectándose una banda única de 3.100 pb en la cepa 1021 y dos de 2.000 y 900 pb, como consecuencia de la la digestión de ésta con *KpnI*, en el mutante (Fig. III-29B, sección derecha).

Finalmente, el análisis de la expresión de *Smr7C* en muestras de RNA total recogidas en fase logarítmica de crecimiento e hibridadas con una sonda específica reveló la ausencia de cualquier forma del transcrito en el mutante 1021 $\Delta$ *smr7C* (Fig. III-29C).

### III-3.1.3.1.2. Sobreexpresión constitutiva de *Smr7C*

Para sobreexpresar constitutivamente el gen *smr7C* en *S. meliloti* se utilizaron los plásmidos pBBSYN*smr7C* y pSRK*smr7C*. La eficiencia de transcripción del sRNA en ambos casos se analizó mediante hibridación Northern de RNA total obtenido en fase logarítmica de crecimiento en medio TY de la cepa salvaje y de las transformadas con cada uno de los plásmidos.



**Fig. III-30. Sobreexpresión constitutiva de *Smr7C*.** Hibridación Northern sobre RNA total del mutante que sobreexpresa *smr7C* utilizando la sonda específica del gen. Los carriles en cada caso corresponden a: M, marcador pGEM; 1, cepa silvestre 1021; 2, 1021pBBSYN*smr7C*; 3, 1021pSRK*smr7C*. En el panel se indica con flechas dobles las bandas inespecíficas de mayor tamaño. En la parte superior del panel se muestran los diagramas correspondientes a las construcciones utilizadas en cada caso, indicando los sitios de restricción más relevantes. En la parte inferior se muestra la cuantificación de las señales obtenidas y normalizadas con la intensidad de la señal de hibridación del RNA 5S.

En los mutantes 1021pBBSYN*Smr7C* se observó que la forma primaria del transcrito *Smr7C* vio aumentada su expresión del orden de 8 veces con respecto a la expresión silvestre, mientras que la inducción de la forma procesada fue más discreta, detectándose una señal 1,5 veces mayor que en 1021 (Fig. III-30, carril 2). Estas hibridaciones volvieron a mostrar la aparición de bandas minoritarias específicas de mayor tamaño que el del transcrito *Smr7C*, presumiblemente derivadas de ondas de transcripción procedentes del propio plásmido.

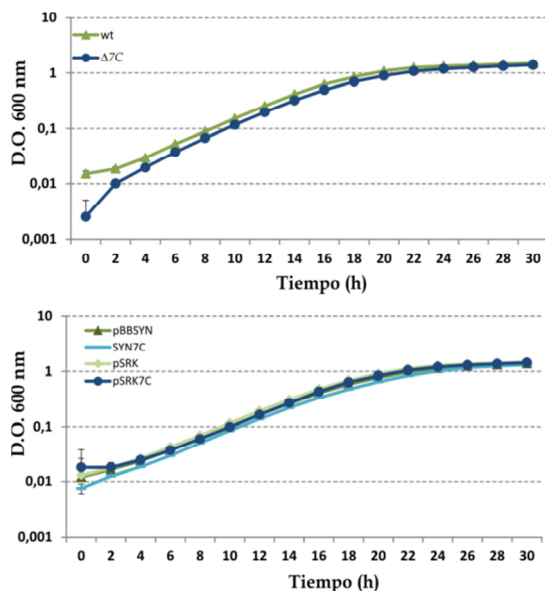
La expresión constitutiva de *Smr7C* desde pSRK se tradujo en un aumento del nivel del transcrito primario de casi 17 veces y del de su forma procesada de casi 4,4 veces, respecto a la expresión silvestre (Fig. III-30, carril 3). En este caso, no se detectaron bandas minoritarias de mayor tamaño.

#### **III-3.2.3.2. Análisis fenotípico de los mutantes *smr7C*: vida libre y simbiosis**

El mutante *S. meliloti* 1021 $\Delta$ *smr7C* y las cepas que sobreexpresan el sRNA, 1021pBBSYN*Smr7C* y 1021pSRK*Smr7C*, mostraron una morfología de colonia idéntica a la de la cepa silvestre 1021, al crecer cualquiera de ellas en placas de agar con medio TY o MM. La monitorización del crecimiento de la cepa salvaje 1021 y el mutante 1021 $\Delta$ *smr7C* en medio TY líquido, no reveló diferencias significativas entre ambas, que alcanzaron la fase estacionaria a la misma D.O. (Fig. III-31, panel superior).

Las cepas que sobreexpresan *Smr7C* se sometieron al mismo análisis, comparándose con la cepa silvestre 1021 transformada con los plásmidos pSRK o pBBSYN vacíos. Las curvas de crecimiento fueron similares en todos los casos (Fig. III-31, panel inferior). Concluimos, por tanto, que ni la delección ni la sobreexpresión del gen *smr7C* tienen un efecto aparente en el crecimiento de *S. meliloti* 1021 en medio TY.

Para evaluar el comportamiento simbiótico de los mutantes *smr7C* en *S. meliloti* 1021 se inocularon independientemente sets de 24 plantas crecidas en cultivo hidropónico en tubos de ensayo; la cepa 1021 y el mutante 1021 $\Delta$ *smr7C* para el estudio de la delección, y el control 1021pBBSYN y la derivada pBBBSYN*Smr7C* para el de la sobreexpresión. La monitorización del número de nódulos por planta y la

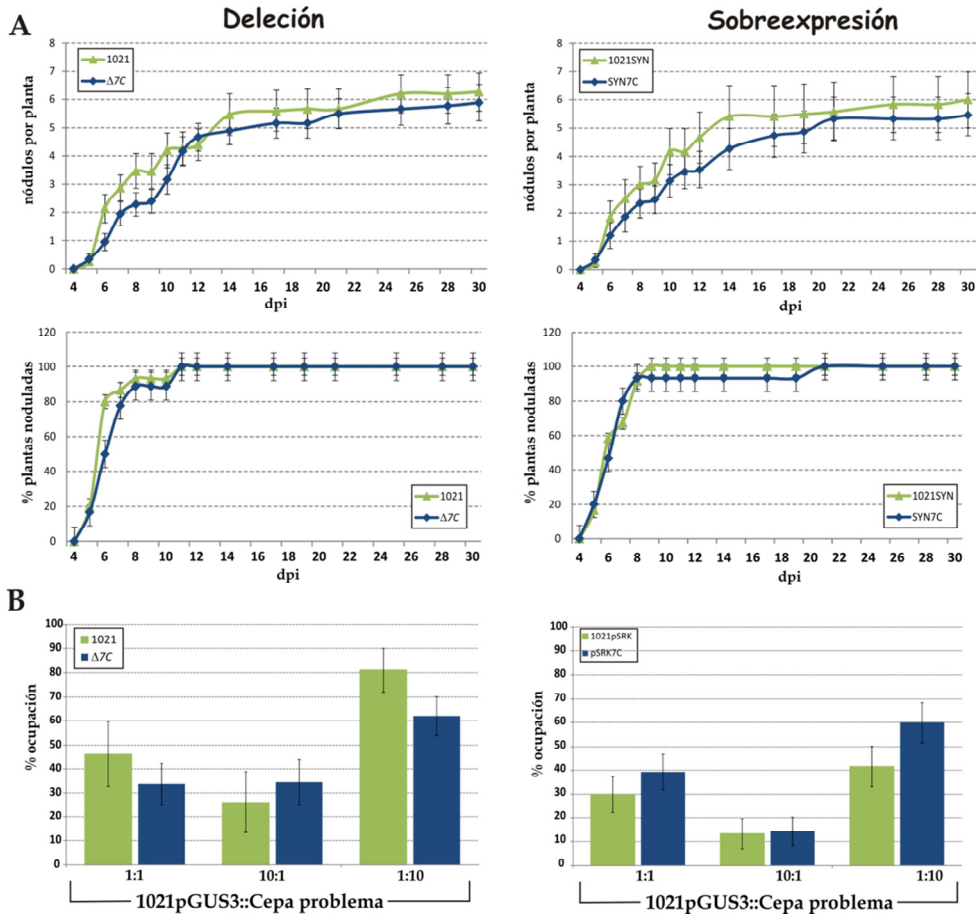


**Fig. III-31. Fenotipo de crecimiento de los mutantes *smr7C*.** Curvas de crecimiento correspondientes a los mutantes por delección (panel superior) y sobreexpresión (panel inferior).

cinética de nodulación no evidenció diferencias significativas en ningún caso cuando se compararon los mutantes con sus respectivos controles lo que sugiere que la alteración de la expresión de *Smr7C* no afecta la capacidad noduladora de *S. meliloti* (Fig. III-32A).

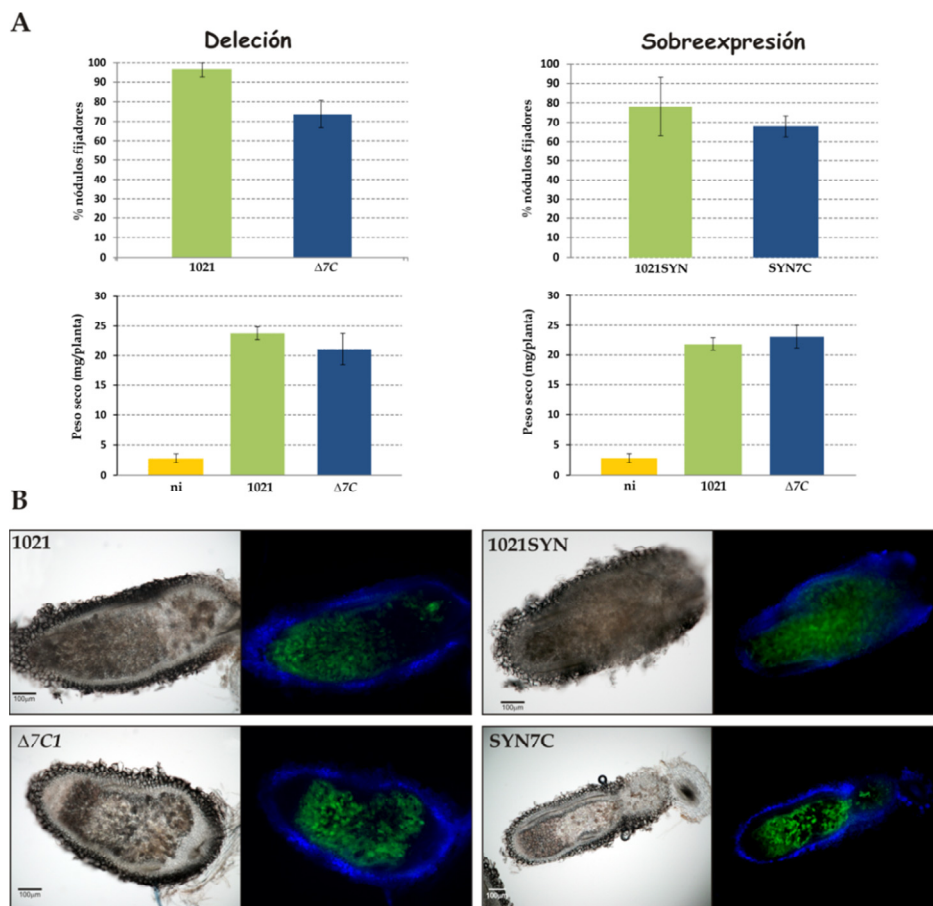
Los experimentos de co-inoculación (competitividad) de 1021pGUS3 con los mutantes *smr7C* (1:1, 10:1 y 1:10) no revelaron diferencias significativas en los porcentajes de ocupación de nódulos de éstos en relación a los de las cepas control en las mismas condiciones (Fig. III-32B). En estos ensayos sólo es destacable que el mutante 1021Δ*smr7C* presentó un porcentaje de ocupación ligeramente menor que la cepa silvestre, cuando ambas bacterias eran mayoritarias en las mezclas de co-inoculación con 1021pGUS3, que reversionó con la sobreexpresión del transcrito.

El conteo de nódulos rosados en los sets de plantas inoculadas independientemente con las cepas de referencia y sus mutantes reveló que existe una disminución del 23% de plantas inoculadas con 1021Δ*smr7C* respecto al número de nódulos fijadores inducidos por su cepa parental 1021 al final del experimento (30 dpi) (Fig. III-33A, sección superior izquierda). Esta disminución se reflejó también en un descenso en el peso seco de las mismas plantas, aunque los análisis estadísticos no consideraron éste significativo ( $p=0,288$ ;  $F=1,179$ ) (Fig. III-33A, sección inferior



**Fig. III-32. Cinéticas de nodulación y competitividad de los mutantes en el gen *smr7C*.** Los paneles situados a la izquierda corresponden a los mutantes por delección y los de la derecha a los mutantes por sobreexpresión. **A)** Infectividad y cinéticas de nodulación representadas por el número de nódulos por planta (paneles superiores) y porcentaje de plantas noduladas (paneles inferiores) de los mutantes *smr7C*. **B)** Ensayos de competitividad (porcentaje de ocupación) de nódulos de las cepas silvestre y mutantes en presencia de la reportera pGUS3. La proporción de la cepa problema sobre la control se indica en cada caso

izquierda). La sobreexpresión de *Smr7C* no tuvo influencia en este parámetro simbiótico (Fig. III-33A, sección derecha). El estudio mediante microscopía de campo claro de la histología de nódulos maduros de *M. sativa* tampoco reveló diferencias aparentes entre los inducidos por los mutantes y los de sus respectivas cepas parentales, mostrando la zonación característica de nódulos indeterminados ocupados por bacteroides autofluorescentes (Fig. III-33B). La elongación de la zona de senescencia o zona IV observada en los nódulos inducidos por mutantes que sobreexpresan *Smr15C1* y *Smr15C2* también se observó en el mutante 1021



**Fig. III-33. Fenotipos endosimbióticos de los mutantes *smr7C*.** **A)** Porcentaje de nódulos fijadores (paneles superiores) y medida del peso seco (paneles inferiores) de plantas de alfalfa ocupadas por los mutantes por delección (sección izquierda) y sobreexpresión (sección derecha) de los genes *smr7C* (ni, no inoculadas). **B)** Imágenes correspondientes a secciones de nódulos de *M.sativa* obtenidas mediante microscopía de campo claro (imágenes situadas a la izquierda de cada caso) y de fluorescencia utilizando filtros para el verde y el azul (520 y 470 nm, respectivamente). La cepa que ocupa estos nódulos se indica en cada caso. Barra de escala: 100  $\mu$ m.

pBBSYN7C, confirmando que este fenotipo no es específico de la sobreexpresión de los sRNAs, sino general de cepas portadoras del plásmido pBBSYN.

#### III-3.2.4. INTERACCIÓN DE *Smr7C* CON LA CHAPERONA *Hfq*

La posible asociación del sRNA *Smr7C* con *Hfq* también se estudió en primer lugar analizando su presencia en CoIP-RNA mediante hibridación Northern y mapeo de las lecturas de RNA-Seq en esta región genómica. La hibridación de RNA total obtenido de cultivos de *S. meliloti* 2011 en medio TY confirmó la mayor acumulación del transcrito en fase exponencial de crecimiento al igual que se observó en *S. meliloti* 1021 (Fig. III-33A). No obstante, en este caso la señal de hibridación de la forma

procesada de Smr7C fue siempre visiblemente más intensa que la correspondiente al transcrito primario. Este perfil de expresión fue también el observado en la cepa *S. meliloti* 2011hfq<sup>FLAG</sup>. Sin embargo, la sonda de Smr7C no detectó ninguna de las formas del transcrito ni en CoIP-RNA preparado de lisados de bacterias de esta última cepa crecida hasta fase logarítmica ni en las correspondientes a muestras control obtenidas de bacterias silvestres.

La visualización mediante IGB del mapeo de las lecturas de la secuenciación masiva de CoIP-RNA en esta región genómica confirmó los hallazgos de las hibridaciones Northern (Fig. III-33B). Smr7C apareció escasamente representado en CoIP-RNA de *S. meliloti* 2011hfq<sup>FLAG</sup> con apenas 50 lecturas en las regiones del transcrito con más cobertura (frente a las 20.000-30.000 lecturas obtenidas para Smr15C1 y Smr15C2), incluso por debajo de la cobertura que se obtuvo para el mismo en CoIP-RNA control (~150 lecturas).

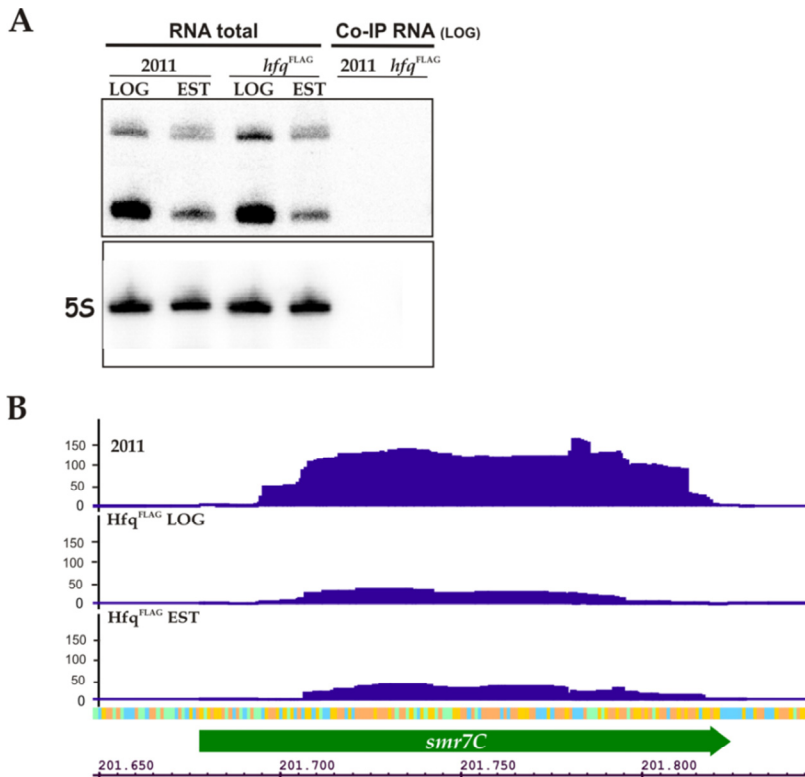


Fig. III-34. Asociación del sRNA Smr7C a la chaperona Hfq. **A)** Hibridación Northern sobre RNA total y CoIP-RNA de las cepas sometidas a CoIP con la sonda específica de Smr7C. **B)** Cobertura (número de lecturas) de la región *smr7C* en las tres librerías de cDNA sometidas a RNA-Seq visualizadas con IGB. La escala inferior corresponde a las coordenadas en el genoma de *S. meliloti* 1021

La ausencia de Hfq tampoco afectó negativamente la estabilidad de Smr7C. La vida media del transcrito primario determinada por cuantificación de las señales de

hibridación en muestras de RNA total de *S. meliloti* 1021, obtenidas 5, 15, 30 y 45 min después de parar la actividad transcripcional con rifampicina, fue de 35 min, mientras que la del procesado fue de 31 min. En el mutante *S. meliloti* 1021 $\Delta$ *hfq* estos valores fueron comparables con los determinados en la cepa silvestre; 41 y 28 min para Smr7C y Smr7C\_2, respectivamente.

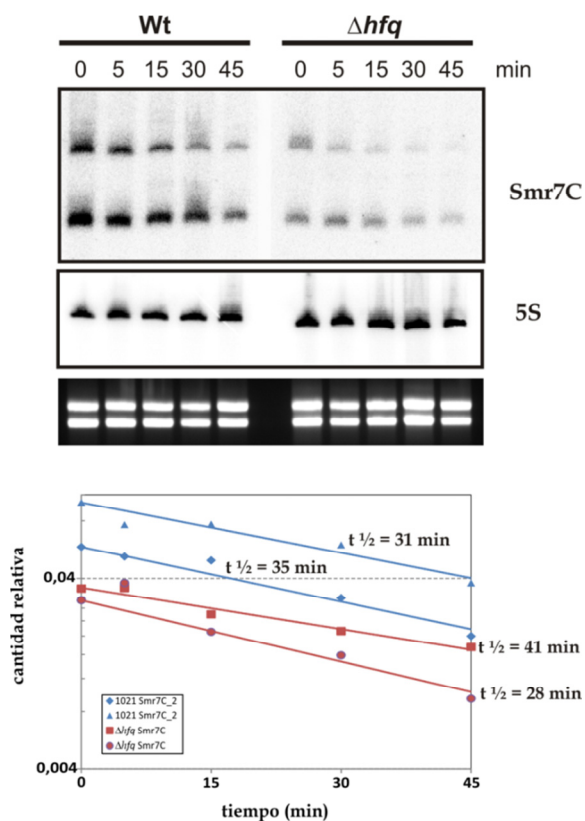


Fig. III-20. Estabilidad de los transcritos Smr7C/Smr7C\_2 en el fondo genético  $\Delta$ *hfq*. El panel superior representa la hibridación Northern sobre RNA total de la cepa salvaje 1021 y el mutante  $\Delta$ *hfq* y el inferior la determinación de las vidas medias ( $t_{1/2}$ ), en min, en los dos fondos genéticos, basada en la cuantificación de cada señal de hibridación en relación a la del RNA 5S. Los tiempos de extracción del RNA total después del tratamiento con Rifampicina se indica en min en la parte superior de cada panel.

En resumen, estos resultados sugieren que la actividad de Smr7C no es dependiente de su asociación con la chaperona Hfq.

### III-3.2.5. APROXIMACIÓN A LA IDENTIFICACIÓN DEL mRNA DIANA DE Smr7C.

#### III-3.2.5.1. Predicciones con IntaRNA.

Para la predicción de los posibles mRNA diana de Smr7C se utilizó IntaRNA con la secuencia del transcrito primario e idénticos parámetros de búsqueda a los utilizados para la generación de las listas de candidatos para Smr15C1 y Smr15C2. En este caso la búsqueda generó un total de 3.359 predicciones en el cromosoma, 1.290



en pSymA y 1.569 en pSymB. Estas listas se ordenaron de acuerdo a la energía libre del híbrido, dando preferencia a las de menor valor, más favorecidas termodinámicamente. Se extrajeron las primeras 50 predicciones correspondientes a cada replicón (con energías comprendidas entre -24,4 kcal/mol y -14,3 kcal/mol), que mayoritariamente involucraron a las horquillas estructurales SL2 (posiciones 35-54) y SL3 (75-89) en las interacciones con los mRNA diana, sin sesgos evidentes hacia ninguna de ellas. Esta lista inicial de 150 candidatos se filtró, dejando para su análisis posterior aquéllos con una interacción predicha con Smr7C en el entorno del sitio SD (i.e. posiciones -15/+15, respecto al codón de inicio de traducción anotado para cada mRNA), que fueron un total de 50; 17 localizados en el cromosoma, 18 en pSymA y 15 en pSymB. Las categorías funcionales de estos mRNAs fueron mucho más diversas que las representadas por los candidatos predichos para Smr15C1/C2, encontrándose un porcentaje particularmente alto de proteínas hipotéticas en este caso (un total de 23 entre los tres replicones). Las funciones predichas para las proteínas codificadas por el resto de mRNAs candidatos fueron: 10 enzimas metabólicas (6 codificadas en el cromosoma, 2 en pSymA y 2 en pSymB), 6 reguladores transcripcionales (uno en cromosoma, 2 en pSymA y 3 en pSymB), 4 transposasas (2 en cada plásmido simbiótico), un receptor de membrana implicado en quimiotaxis (Cpv) codificado en el cromosoma, una proteína del sistema de secreción HlyD codificada en pSymA y cinco transportadores codificados en pSymB.

Todas estas interacciones se revaluaron y se seleccionaron para su posterior ensayo *in vivo* aquéllas que involucraban series más largas de nucleótidos en los mRNA con función predecible, que en general fueron las más favorecidas termodinámicamente. Esta selección final generó una lista de **4 mRNAs candidatos** que se relacionan en la Tabla III-6, junto a los parámetros de interacción con Smr7C predichos por IntaRNA en cada caso.

**Tabla III-6. Interacciones más probables para Smr7C.**

	Energía (kcal/mol)	Posición sRNA	Posición mRNA <sup>1</sup>	Función mRNA
<i>eryD</i> (SMc01629)	-17,4	19-52	-10/+23	Represor operón eritritol
<i>wggR/expG</i> (Smb21317)	-17,4	75-88	-7/+16	Inductor síntesis EPSII
<i>SMa0206</i>	-15,7	75-88	-11/+5	Regulador de respuesta
<i>rhtA</i> (SMa2414)	-14,34	17-60	-19/+26	Receptor rizobactina

<sup>1</sup> Respecto al AUG anotado

### III-3.2.5.2. Evidencias experimentales: ensayo de doble plásmido.

Debido a que Smr7C no evidenció dependencia de la chaperona Hfq la única aproximación experimental que se llevó a cabo para la verificación de las

predicciones fue el ensayo de doble plásmido *in vivo*. Las características más relevantes de los plásmidos diana pR\_EryD::GFP, pR\_RhtA::GFP, pRO\_WggR::GFP y pRO\_SMa0206::GFP, que expresan las fusiones traduccionales entre las regiones 5' de cada mRNA y eGFP se recogen en la Tabla III-7. Estos plásmidos se movilizaron a las cepas *E. coli* DH5 $\alpha$  y a *S. meliloti*  $\Delta$ smr7C previamente transformadas de forma independiente con pSRKSmr7C, que expresa constitutivamente el sRNA y pSRK como control negativo de las interacciones. Los dobles transformantes (i.e. 8 cepas de *E. coli* DH5 $\alpha$  y 8 de *S. meliloti*  $\Delta$ smr7C) se crecieron en medio TY agar y se observaron en escáner de fluorescencia.

**Tabla III-7. Características relevantes de los plásmidos diana para Smr7C utilizados en el ensayo *in vivo*.**

mRNA <sup>1</sup>	Nombre del plásmido	Coordenadas genómicas del inserto <sup>2</sup>	Codón fusionado	Fluorescencia en placa <sup>3</sup>
<i>eryD</i> (SMc01629)	pR_EryD::GFP	2422200-2422192	16	++
<i>wggR/expG</i> (SMb21317)	pRO_WggR::GFP	983814-983901	14	+++
<i>SMa0204/SMa0206</i>	pRO_SMa0206::GFP	114468-114544	12	+++
<i>rhtA</i> (SMa2414)	pR_RhtA::GFP	1315730-1315745	17	++

<sup>1</sup> Nombre del mRNA (gen) y coordenadas según la anotación genómica de *S. meliloti* 1021.

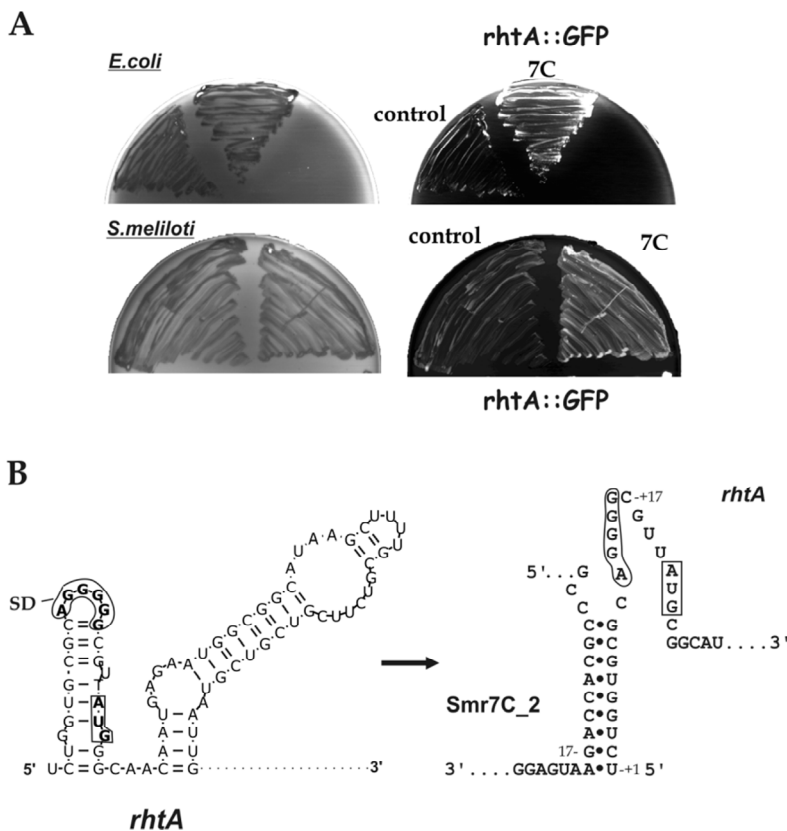
<sup>2</sup> Extremo 5' del fragmento del gen diana en relación al ATG anotado. Los sitios de inicio de la transcripción de cada mRNA fueron facilitados por A. Becker (no publicados).

<sup>3</sup> Intensidad de fluorescencia detectada: +++, muy intensa; ++, intensidad media; +, intensidad baja; -, fluorescencia no detectada

Estos ensayos no evidenciaron alteración del patrón de fluorescencia de las fusiones en presencia de Smr7C, con la excepción de pR\_RhtA::GFP (Fig. III-36A), cuya fluorescencia basal aumentó visiblemente cuando coexistía en la célula con pSRKSmr7C en un total de tres de cinco ensayos, pero nunca con el plásmido control. Este efecto fue también evidente en el fondo genético de *E. coli* DH5 $\alpha$ .

El gen *rhtA* codifica la proteína de membrana receptora del sideróforo rizobactina en *S. meliloti* 1021 y su transcripción es inducida en situaciones de estrés de hierro. IntaRNA predice una interacción de Smr7C con el mensajero de *rhtA* que involucra a los nucleótidos localizados entre las posiciones 17 y 60 del transcrito primario (i.e. parte del motivo SL1 y el motivo SL2), que son parcialmente complementarios a una secuencia comprendida entre las posiciones -19 y +26, respecto al codón AUG, en el mRNA (Fig. III-36B). RNAfold predice una horquilla estructural en la región 5'-UTR del mRNA de *rhtA* que incluye el sitio SD, el inicio de la traducción AUG y varios codones correspondientes a la secuencia codificante. Esta estructura podría impedir la traducción de RhtA, explicando así la baja fluorescencia basal de la fusión pR\_RhtA::GFP en placa (Fig. III-36A). El apareamiento de Smr7C en esa región

podría, por tanto, desorganizar esta horquilla inhibitoria para permitir la traducción del mensajero, con el consiguiente aumento de la fluorescencia de la fusión traduccional observada en nuestros ensayos. La transcripción de *Smr7C* es inducida también en situaciones de carencia de hierro, condición en la que se acumula preferentemente la forma procesada del transcrito, cuyos primeros 17 nt, complementarios a la región 5'-UTR del mensajero de *rhtA* están despareados y, por tanto, más accesibles para la interacción (Fig. III-36B). Estos resultados sugieren que en situaciones de carencia de hierro el procesamiento mayoritario de *Smr7C* podría favorecer su interacción con la región 5'-UTR de *rhtA* posibilitando la traducción de éste en un proceso independiente de Hfq.



**Fig. III-36. Interacción de Smr7C con RhtA.** **A)** Ensayo de doble plásmido in vivo en *E. coli* (sección superior) y *S. meliloti* (sección inferior) de la 5'UTR del mRNA *RhtA* con el plásmido que expresa Smr7C. Control: plásmido vacío. **B)** Mecanismo de complementariedad de bases predicho en donde Smr7C\_2 activa a *rhtA* de *S. meliloti*. El panel de la izquierda representa la 5'-UTR del mRNA *rhtA* en el estado "inactivo", con la horquilla que secuestra el SD (indicado) e inhibe la traducción. El panel derecho representa la activación de *rhtA* por Smr7C\_2 que supondría la disociación de la horquilla y la liberación del SD (remarcado) y el AUG (encuadrado). Las posiciones de cada RNA se indican en cada caso.

### III-4. DISCUSIÓN

Los *trans*-sRNAs son los ribo-reguladores bacterianos mejor caracterizados en lo referente a su biogénesis, función y mecanismos de acción (Storz *et al.*, 2011). Sin embargo no se conoce la función de ningún sRNA de este tipo entre los identificados en bacterias endosimbióticas vegetales. Una caracterización muy preliminar de los sRNAs Smr identificados mediante genómica comparada y presentada en el Capítulo

1 anticipaba que estos transcritos podrían adscribirse a la categoría de los *trans*-sRNAs. En este Capítulo hemos profundizado en la caracterización funcional de los sRNAs Smr15C1, Smr15C2, y Smr7C. Nuestros resultados revelan que estos tres sRNAs son prescindibles para la interacción simbiótica con *M. sativa* y se acumulan diferencialmente en condiciones biológicas diversas para la regulación post-transcripcional de proteínas de transporte mediante mecanismos dependientes (Smr15C1/C2) e independientes (Smr7C) de la chaperona de RNA Hfq.

### ***Características estructurales y distribución de las familias de sRNAs $\alpha$ 15 y $\alpha$ 7.***

La caracterización estructural de cualquier sRNA requiere del conocimiento previo de su secuencia nucleotídica primaria exacta. En este trabajo hemos empleado la tecnología RACE para el mapeo experimental de los extremos 5' y 3' de los sRNAs Smr15C1, Smr15C2 y Smr7C. Nuestras determinaciones están en perfecta correlación con los extremos identificados para estos transcritos mediante RNA-Seq de la fracción de RNAs de pequeño tamaño de *S. meliloti* 2011 (Schlüter *et al.*, 2010).

Los *trans*-sRNAs suelen presentar una organización modular muy conservada, consistente en varias horquillas estructurales, de las cuales la terminal corresponde al terminador de transcripción independiente de Rho, y series de nucleótidos desapareados generalmente implicados en el establecimiento de las interacciones con secuencias complementarias en el mRNA diana (Wagner *et al.*, 2002). La simulación del plegamiento de los transcritos Smr15C1 y Smr15C2 predice una estructura secundaria muy similar para ambos constituida por tres horquillas (SL1-3), idéntica a la de sus homólogos en *A. tumefaciens* (AbcR1/R2) que ha sido mapeada experimentalmente (Wilms *et al.*, 2011). En el bucle de la primera horquilla de ambos sRNAs se reconocen motivos aSD expuestos a una posible interacción en el entorno del RBS de sus posibles mRNA diana. Otros motivos típicos de *trans*-sRNA de función conocida reconocibles en esta estructura son una corta secuencia desapareada rica en A/U y una cola poli-U en el extremo 3' que han sido propuestos como motivos de unión a la chaperona Hfq (Storz *et al.*, 2011). Por el contrario, la secuencia de Smr7C pliega en una estructura integrada por cuatro horquillas (SL1-4), en la que no son identificables motivos aSD expuestos en regiones desapareadas ni potenciales sitios de unión a Hfq. En este caso toda la secuencia final de uridinas se predice apareada e integrada en el tallo que conforma el terminador, por lo que el grupo hidroxilo de la uridina terminal estaría menos accesible para la interacción con la chaperona (Schumacher *et al.*, 2002; Sauer and Weichenrieder, 2011). El procesamiento de este transcrito en la posición 42 genera una amplia región 5' desapareada muy accesible a potenciales interacciones con mRNAs.

Estos modelos estructurales ( $\alpha 15$  y  $\alpha 7$ ) aparecen únicamente representados en bacterias filogenéticamente relacionadas del orden Rhizobiales del grupo de las  $\alpha$ -proteobacterias.  $\alpha 7C$  representa un caso típico de familia de sRNAs codificados en genes cromosómicos de copia única y localizados en regiones genómicas altamente sinténicas en las distintas especies. La filogenia de sus miembros sugiere que la evolución de los sRNAs  $\alpha 7$  ha ocurrido exclusivamente por herencia vertical. Por el contrario, la distribución de la familia  $\alpha 15$  es mucho más compleja. Este modelo identificó en cada genoma individual un número variable de genes  $\alpha 15$  localizados en plásmidos accesorios además de las copias cromosómicas, generalmente dispuestas en tándem en la misma RIG. El agrupamiento estructural de todos los representantes de esta familia así como el grado diverso de sintenia entre sus regiones genómicas sugieren que tanto las duplicaciones en distintas etapas de la evolución como la transferencia genética horizontal, sobre todo de los *loci* extracromosómicos, han contribuido a la generación de estas reiteraciones. La localización de genes que codifican *trans*-sRNAs en regiones genómicas adquiridas por transferencia horizontal no es excepcional. Así por ejemplo, la isla de patogenicidad SPI-1 de *S. enterica* expresa un *trans*-sRNA muy abundante y dependiente de Hfq, InvR, que regula la síntesis de la proteína de membrana OmpD codificada en el cromosoma (Pfeiffer *et al.*, 2007). Nuestros experimentos de hibridación Northern, sin embargo, no han detectado expresión del tercer parálogo de Smr15C1/C2 en *S. meliloti*, cuyo *loci* fue identificado en el plásmido simbiótico pSymA.

### **Los sRNAs Smr15C1/C2 y Smr7C están diferencialmente regulados en *S. meliloti***

Una de las características que identifica a los *trans*-sRNAs es su expresión en respuesta a estímulos medioambientales concretos. Los datos presentados en el capítulo 1 ya anticipaban la acumulación diferencial de los transcritos Smr15C1/C2 y Smr7C en situaciones de crecimiento en vida libre y en endosimbiosis. En este capítulo hemos ampliado el estudio de la expresión de estos sRNAs a otras condiciones biológicas relevantes para el establecimiento competitivo de *S. meliloti* en suelo y para la interacción simbiótica con *M. sativa*. Smr15C1 y Smr15C2 muestran una expresión totalmente divergente en *S. meliloti* 1021. En medio TY, Smr15C1 presenta el patrón de acumulación mayoritario de los sRNAs bacterianos, con máxima expresión en fase estacionaria tardía de crecimiento, mientras que Smr15C2 lo hace preferentemente en la fase exponencial. Siempre se ha argumentado que la inducción de los sRNAs con la entrada en fase estacionaria refleja su regulación transcripcional por regulones diversos de respuesta a estrés, que son gradualmente activados a medida que va cesando el crecimiento bacteriano por agotamiento de

nutrientes (Papenfort *et al.*, 2006). En la línea de esta hipótesis nuestros resultados revelan una inducción de Smr15C1 en prácticamente todas las situaciones de estrés imitadas en los cultivos, con la excepción del estrés de membrana inducido por etanol y la limitación de hierro. Por el contrario, los sRNAs que se expresan durante el crecimiento exponencial en medio rico podrían responder a señales más específicas o participar, como el caso de GcvB, en la represión generalizada del transporte de nutrientes alternativos (i.e. aminoácidos) cuando las fuentes de energía prioritarias (i.e. azúcares) son abundantes (Sharma *et al.*, 2008). Smr15C2 podría ser un ejemplo de este tipo de sRNAs, pues su inducción en relación a la acumulación del transcrito en fase exponencial sólo fue observada en situación de carencia de hierro imitada con el quelante 2,2'-dipiridil. En tejido nodular también fue evidente una expresión diferente de ambos transcritos; Smr15C1 no fue detectado en ninguna de las zonas típicas de los nódulos indeterminados maduros, mientras que la expresión de Smr15C2 fue particularmente intensa en las bacterias no diferenciadas que ocupan la zona de invasión (i.e. zona II). Smr15C1 y Smr15C2 no parecen actuar de forma jerárquica, pues la ausencia de cualquiera de los dos no altera la cinética de acumulación del otro en medio completo TY (Fig. III-14C). Sus patrones de expresión sugieren que sus promotores están diferencialmente regulados, pudiendo actuar de forma aditiva en el control del mismo proceso, como los sRNAs parálogos Qrr que regulan el *quorum sensing* en *V. harveyi* (Tu and Bassler, 2007), o independiente, para la regulación de sets diferentes o parcialmente solapantes de mRNAs.

Al contrario que Smr15C1 y Smr15C2, sus homólogos en *A. tumefaciens* (AbcR1 y AbcR2) presentan cinéticas de expresión idénticas, con acumulación máxima en fase estacionaria tardía de crecimiento (Wilms *et al.*, 2011). El alineamiento múltiple de las secuencias promotoras de todos los sRNAs  $\alpha 15$  agrupó Smr15C1 con sus homólogos en bacterias del género *Sinorhizobium* y a Smr15C2 con el resto de los miembros de la familia. El análisis de las secuencias conservadas en cada grupo de promotores identificó posibles motivos de unión a sendos factores de transcripción del tipo LacI, potencialmente implicados en la respuesta a la disponibilidad de determinadas fuentes de carbono en el medio externo. Estas predicciones podrían justificar los patrones de expresión de las parejas de parálogos  $\alpha 15$  en *A. tumefaciens* y *S. meliloti* a la vez que anticipan la participación de Smr15C1 y Smr15C2 en el ajuste del metabolismo de *S. meliloti* en respuesta a situaciones de estrés diversas.

Por su parte, Smr7C presentó una cinética de expresión en medio rico algo diferente a Smr15C1/C2, con una acumulación progresiva del transcrito primario y su forma procesada durante toda la fase de crecimiento exponencial, para hacerse ambos prácticamente indetectables en fase estacionaria tardía. Este sRNA también es

inducido en respuesta a  $H_2O_2$ , carencia de hierro, acidez y microaerobiosis. Al igual que Smr15C2, en nódulos maduros de alfalfa sólo se detecta expresión de Smr7C en bacterias no diferenciadas, en un patrón de acumulación probablemente asociado a la respuesta al estrés oxidativo en la zona de invasión del nódulo. Llama la atención que en las situaciones de máxima inducción de Smr7C, sea la forma procesada y no el transcrito primario, la que se acumula preferentemente (Fig. III-27). Como se discute en epígrafes posteriores, estos hallazgos sugieren que el procesamiento de Smr7C podría generar la forma funcionalmente activa del sRNA para la interacción con el mRNA diana.

### ***Smr15C1/C2 y Smr7C son prescindibles para el crecimiento en medio rico de S. meliloti y la interacción simbiótica con M. sativa***

La alteración de la expresión endógena de los sRNAs Smr15C1/C2 y Smr7C por delección o sobreexpresión constitutiva de sus genes respectivos, no alteró en ningún caso la tasa de crecimiento de *S. meliloti* 1021 en medio completo TY, ni tampoco comprometió su capacidad para inducir infecciones en alfalfa, su competitividad por la nodulación, la organogénesis nodular o la eficiencia de la fijación simbiótica de nitrógeno. Estos resultados no son, en ningún caso, sorprendentes y pueden atribuirse a dos factores: i) nuestra caracterización fenotípica de los mutantes en los sRNAs Smr no ha considerado condiciones biológicas más específicamente relacionadas con los perfiles de expresión determinados para cada transcrito; es razonable que sólo un pequeño porcentaje de los cientos de *trans*-sRNAs expresados por *S. meliloti* 1021 controlen procesos simbióticos críticos cuya desregulación tenga manifestaciones fenotípicas evidentes que, no obstante, podrían identificarse mediante genética reversa, ii) los *trans*-sRNAs son reguladores finos de la expresión génica, muy versátiles y en muchos casos muy reiterados en genomas individuales, que optimizan las respuestas de las bacterias ante los súbitos cambios en el medio externo, pero no son esenciales para la supervivencia de éstas en situaciones de estrés. Por ello en la mayoría de los casos los efectos de su actividad son sólo evidentes a nivel molecular. En *E. coli* y *S. meliloti* ningún *trans*-sRNAs ha sido identificado mediante escrutinios genéticos clásicos basados en la mutagénesis al azar con transposones, probablemente porque en ningún caso las inserciones en estos genes provocaron los fenotipos buscados (Meade *et al.*, 1982; de Lorenzo *et al.*, 1990; de Lorenzo and Timmis, 1994; Wassarman *et al.*, 1999).

### ***Hfq une y estabiliza Smr15C1/C2 pero no Smr7C***

La actividad de la mayoría de los *trans*-sRNAs conocidos requiere la asistencia de la chaperona de RNA Hfq, cuya acción se traduce, entre otros efectos, en una

estabilización del sRNA (Valentin-Hansen *et al.*, 2004; Brennan and Link, 2007). La dependencia de Hfq es, por tanto, un aspecto a estudiar necesariamente en la caracterización funcional y del mecanismo de acción de los ribo-reguladores expresados por bacterias que, como *S. meliloti*, codifican homólogos reconocibles de esta proteína. Confirmando las predicciones *in silico*, las hibridaciones Northern y el análisis de los datos de la RNA-Seq revelaron que Smr15C1 y Smr15C2 se asocian con gran afinidad por Hfq (factores de enriquecimiento 44 y 21, respectivamente CoIP-RNA; Tabla I-7). Por el contrario, ni el transcrito primario ni la forma procesada de Smr7C aparecieron representados específicamente en CoIP-RNA.

Smr15C1/C2, Smr7C y Smr7C\_2 son todos transcritos muy estables en *S. meliloti* 1021. Sus tiempos de vida media respectivos, calculados después del bloqueo de la actividad de transcripción con rifampicina fueron superiores a 20 min en todos los casos, muy altos en comparación a los 2-4 min determinados en experimentos similares para sRNAs expresados por *Salmonella* y catalogados inestables, como GlmY, GlmZ o GcvB (Sharma *et al.*, 2008; Urban and Vögel, 2008). En bacterias, la poli(A) polimerasa I (PAP I), contribuye a la desestabilización de los transcritos por poliadenilación de sus extremos 3', facilitando así la degradación exonucleolítica 3'-5' de RNAs con regiones 3' muy estructuradas como las de los *trans*-sRNAs. De esta forma PAP I podría contribuir también a la expresión diferencial típica de estos ribo-reguladores (Dreyfus *et al.*, 2002; Joanny *et al.*, 2007). De hecho hay varios casos bien documentados de *trans*-sRNAs en *E. coli* y *Salmonella* en los que la poliadenilación por PAP I es un factor esencial en su regulación post-transcripcional y mecanismo de acción (i.e. GlmY, SraL) (Urban and Vögel, 2008; Viegas *et al.*, 2007). En *S. meliloti* PAP I está codificada por el gen cromosómico *SMc02700* (Galibert *et al.*, 2001), sin embargo, los experimentos RACE no revelaron signos de poliadenilación de ninguno de los transcritos Smr, lo que podría contribuir a su alta estabilidad. Por el contrario, la ausencia de Hfq comprometió la estabilidad de Smr15C1/C2 de forma muy evidente (vida media de 2 y 11 min, respectivamente en el mutante *hfq*) pero no afectó la de ninguna de las formas en las que existe Smr7C. En conjunto, estos resultados sugieren que en *S. meliloti* Hfq es la chaperona que asiste la actividad de Smr15C1/C2 pero no la de Smr7C.

Los *trans*-sRNAs cuya actividad es totalmente independiente de Hfq son una clase minoritaria entre los de función conocida en bacterias que expresan esta RNA chaperona, de forma que sólo hay descritos ejemplos en *B. subtilis* y *S. aureus* (Silvaggi *et al.*, 2005; Boisset *et al.*, 2007; Heidrich *et al.*, 2007). El de Smr7C no es, sin embargo, un caso excepcional en *S. meliloti*. Conviene recordar aquí que casi el 60% de los *trans*-sRNAs representados en nuestras preparaciones de CoIP-RNA no habían



sido anotados por Schlüter *et al.* (2010) tras la secuenciación masiva de la subpoblación de sRNAs expresados por *S. meliloti* 2011 en ocho condiciones biológicas diferentes. Este dato revela que en *S. meliloti* co-existen amplias poblaciones de *trans*-sRNAs dependientes e independientes de Hfq y predice la existencia de otras RNA chaperonas que pueden contribuir a la ribo-regulación. Una de ellas podría ser la codificada por el gen *SMc01113*, que comparte características estructurales con el dominio MID de las proteínas Argonauta (AGO) que forman parte del complejo RISC (*RNA-induced-silencing-complex*) eucariótico al que se asocian los miRNAs (Pandey *et al.*, 2011). La delección de *SMc01113* tiene efectos pleiotrópicos similares a la de *hfq* en *S. meliloti*, afectando fenotipos bacterianos de vida libre y simbióticos, así como la expresión de alguno de los sRNAs conocidos, por lo que podría tener un impacto en la ribo-regulación similar al de Hfq.

***Identificación de los mRNA diana de los sRNAs Smr15C1, Smr15C2 y Smr7C: valoración de las predicciones in silico y del ensayo genético in vivo***

La identidad de los mRNAs directamente regulados por los *trans*-sRNAs es, sin duda, el dato más revelador de la función de éstos. Este tipo de ribo-reguladores interactúan con sus mRNA diana a través de apareamientos con series cortas y discontinuas de nucleótidos complementarios, mayoritariamente localizados en la región 5'-UTR de los mRNAs. En la mayoría de los casos, aunque la complementariedad pueda ser más extensa, el apareamiento de 6-7 nt (secuencia *seed* del sRNA), generalmente desapareados en cualquiera de las regiones del sRNA, es suficiente para una regulación efectiva del mRNA, habilitando incluso a un único sRNA para su interacción con múltiples mensajeros. Estas circunstancias hacen de la predicción/identificación del(los) mRNA diana una labor muy compleja que constituye un auténtico desafío y el verdadero cuello de botella para el progreso en la caracterización funcional de los *trans*-sRNAs. De hecho los *trans*-sRNAs con dianas validadas experimentalmente son absoluta minoría en relación al número de ellos conocidos en cada especie bacteriana (Storz *et al.*, 2011).

Una primera aproximación a este problema son las predicciones bioinformáticas, potencialmente aplicables a cualquier bacteria con genoma secuenciado, basadas en la termodinámica de la interacción sRNA-mRNA y en la accesibilidad para el apareamiento en ambas moléculas de las secuencias supuestamente implicadas en el mismo. Como era esperado este tipo de predicciones generaron listas de miles de mRNAs (regiones 5') como probables dianas de Smr15C1/C2 y Smr7C.

Estos programas no tienen en cuenta por desconocidos otros factores supuestamente relevantes para una ribo-regulación productiva como los sitios de

asociación de las RNA chaperonas en el dúplex sRNA-mRNA (e.g. Hfq) u otros elementos estructurales específicos, que podrían reducir sustancialmente las listas de potenciales dianas para cada sRNA (Tjaden *et al.*, 2006; Richter *et al.*, 2010). Sin embargo, la relación funcional entre los mRNAs candidatos se considera un primer índice de confianza en las predicciones (Tjaden *et al.*, 2006). En este sentido entre los múltiples mRNAs diana predichos por estas aproximaciones para sRNAs expresados por *E. coli* y *Salmonella* como GcvB, RybB y RyhB predominan los que codifican transportadores ABC, proteínas externas de membrana (OMPs) y proteínas del metabolismo del hierro, respectivamente, para muchos de los cuales se tienen sólidas evidencias experimentales (Sharma *et al.*, 2008; Sharma *et al.*, 2011; Papenfort *et al.*, 2006; Massé and Gottesman, 2002). Esta relación funcional fue evidente en el catálogo de candidatos para Smr15C1/C2, con amplia mayoría de mRNAs que especifican los componentes estructurales de los transportadores ABC, fundamentalmente las proteínas periplásmicas implicadas en el reconocimiento y unión de los sustratos específicos en cada caso y en ellas centramos nuestra atención posterior. Además, las interacciones Smr15C1/C2-mRNA predichas *in silico* involucran en la mayoría de los casos al motivo aSD del dominio 5' de ambos sRNAs en el apareamiento con secuencias complementarias que incluyen todo el RBS de los mRNAs, solapan con éste o se extienden a las secuencias nucleotídicas inmediatamente adyacentes, en la línea de lo observado en la mayoría de las interacciones sRNA-mRNA descritas en la literatura (Storz *et al.*, 2011). Es más, en muchos casos estas regiones 5' de los *trans*-sRNAs se han catalogado como motivos autónomos, funcionales para el apareamiento con los mRNA diana con independencia del contexto estructural en el que se integran, lo que añade un nuevo índice de confianza a muchas de las predicciones para Smr15C1/C2 (Guillier and Gottesman, 2008; Papenfort *et al.*, 2010). Aunque las secuencias de los bucles de la horquilla 5' de estos dos sRNAs conservan el motivo aSD referido, también evidencian cierta divergencia, lo que explicaría que las listas de candidatos para cada sRNA fueran parcialmente coincidentes, anticipando que a pesar de su homología, Smr15C1 y Smr15C2 podrían regular diferentes sets de mRNAs en *S. meliloti*.

Por el contrario, la lista de candidatos para Smr7C no evidenció ninguno de los sesgos referidos, por lo que la selección de mRNAs para su estudio posterior se basó en la asunción generalizada de que las interacciones más extensas entre nucleótidos son probablemente las implicadas en las interacciones sRNA-mRNA más específicas (Storz *et al.*, 2011).

La validación inequívoca de las interacciones sRNA-mRNA diana siempre requiere la acumulación de varias evidencias experimentales (Urban and Vögel,

2007). En nuestro caso, como prueba experimental directa de la especificidad de las interacciones utilizamos un ensayo genético *in vivo* basado en la observación de cambios visibles en la fluorescencia de fusiones traduccionales a GFP de las regiones 5' de una selección de mRNAs diana predichos *in silico* cuando eran co-expresadas con los sRNAs Smr15C1, Smr15C2 y Smr7C. Estos experimentos sólo validaron dos interacciones entre todas las testadas. El bajo índice de resultados positivos obtenidos en nuestros ensayos puede ser debido a varios factores. En primer lugar, la actividad de la GFP incorporada en estas construcciones genéticas posiblemente no sea la óptima para la visualización de efectos sutiles de los sRNAs sobre la traducción de alguna de las fusiones, de forma que la utilización de variantes de GFP con fluorescencia optimizada podría mejorar la reproducibilidad de los ensayos (Pedelacq *et al.*, 2006). Segundo, la gran mayoría de mRNAs testados codifican proteínas que son exportadas al periplasma o integradas en la membrana, por lo que quizá en muchos casos las secuencias aminoacídicas que especifican estas localizaciones no estaban incluidas en las fusiones y mantengan las proteínas recombinantes confinadas al citoplasma. Recientemente ha sido publicado que la localización en membrana del péptido naciente codificado por el mRNA de *ptsG* es absolutamente necesaria para su ribo-regulación (represión) por el *trans*-sRNA SgrS (Kawamoto *et al.*, 2005). Por tanto, será muy importante testar en el futuro cómo la inclusión de los péptidos señal precisos puede afectar la regulación de algunas de nuestras fusiones por los sRNAs Smr. En tercer lugar, varios estudios recientes han revelado una considerable heterogeneidad de la transcripción de genes bacterianos a nivel poblacional, lo que puede ser cierto también para procesos post-transcripcionales como los analizados en este trabajo (Bagge *et al.*, 2004). La utilización de GFP como proteína reportera que no precisa de sustratos cromogénicos adicionales posibilitará en el futuro el estudio de la regulación génica a nivel de célula única, lo que será de gran utilidad también en la investigación de la ribo-regulación en bacterias. Finalmente, este ensayo genético de doble plásmido está concebido con carácter cualitativo aunque admite cuantificación a varios niveles (i.e. determinación de la fluorescencia mediante citometría de flujo o valoración mediante Western blot de la tasa de traducción de la proteína de fusión en cada ensayo). Es previsible, por tanto, que la introducción de mejoras en el ensayo y su cuantificación puedan poner de manifiesto en el futuro interacciones sRNAs Smr-mRNAs no evidenciadas en nuestros estudios.

### ***Las evidencias experimentales sugieren que Smr15C1 y Smr15C2 regulan múltiples transportadores ABC***

El ensayo genético *in vivo* para la validación de los mRNA diana predichos para Smr15C1/C2 reveló que Smr15C2 induce la expresión de la proteína codificada por el gen *SMa1650*, segundo de un operón que codifica la permeasa de membrana de un transportador ABC cuya transcripción es inducida por casaminoácidos (Mauchline *et al.*, 2006). Éstos constituyen un grupo muy heterogéneo de aminoácidos y pequeños péptidos, generados por hidrólisis ácida de la caseína, que son muy utilizados en la formulación de los medios de cultivo ricos en nutrientes como el TY de *S. meliloti*. Con independencia de otras implicaciones biológicas, la regulación positiva de la traducción de *SMa1650* por Smr15C2 durante la fase exponencial de crecimiento contribuiría, por tanto, a la utilización de estas fuentes de energía por las bacterias.

Como ya se ha apuntado en la discusión de los resultados presentados en el Capítulo 2, entre los mRNAs/proteínas dependientes de Hfq (i.e. diferencialmente expresados en un mutante *hfq*) y los rescatados por co-inmunoprecipitación con la proteína se han identificado las dianas de varios de los sRNAs bien caracterizados en *Salmonella* (e.g. mRNAs *dppA*, *fadL*, *ompD* o *oppA*) (Sittka *et al.*, 2008). Una nueva inspección de los mRNAs/proteínas catalogados como diferencialmente expresados por los mutantes *hfq* de *S. meliloti* y de los mRNAs representados en CoIP-RNA obtenido en fase logarítmica y estacionaria de crecimiento y tres condiciones biológicas más (A. Becker y J.-P. Schlüter, comunicación personal) reveló la presencia en ambas listas de dianas predichas *in silico* para Smr15C1 (i.e. *aglF*, *aglE*, *oppA*) o para los dos transcritos (i.e. *SMa0495*, *SMb20442*, *SMb20144* y el ya referido *SMa1650*) (Fig. III-22). Aunque en general la especificidad de sustrato de estos transportadores no está caracterizada en *S. meliloti* algunos datos experimentales sobre la transcripción de estos genes y las predicciones *in silico* sugieren que *SMa0495*, *SMb20144* y *OppA* estarían implicados en la utilización de compuestos nitrogenados como aminoácidos y péptidos (Mauchline *et al.*, 2006; Nogales *et al.*, 2009), *SMb20144* en la de azúcares diversos como fucosa o manosa (Mauchline *et al.*, 2006) y *AgIF* en la de  $\alpha$ -glucósidos. En general, las secuencias de estos mRNAs más representadas en CoIP-RNA fueron las regiones 5' de cada transcrito donde se predicen sus apareamientos con los sRNAs Smr15C1/C2 (Fig. III-22), lo que constituiría una evidencia experimental más de la existencia de estas interacciones *in vivo*. No obstante, el efecto de Smr15C1/C2 sobre la traducción de cada proteína y sus bases moleculares deben ser investigadas en el futuro mediante aproximaciones experimentales complementarias.

La determinación de los cambios globales en la expresión de proteínas por la alteración de la expresión endógena de los *trans*-sRNAs ha sido también una aproximación experimental productiva para la identificación de los mRNA diana de alguno de los ribo-reguladores mejor caracterizados en *E. coli* y *Salmonella* (Udekwu *et al.*, 2005; Rasmussen *et al.*, 2005; Altuvia *et al.*, 1997; Massé *et al.*, 2005; Papenfort *et al.*, 2006). Guiados por las predicciones bioinformáticas, en este trabajo hemos comparado también los cambios en los perfiles de las proteínas periplásmicas provocados por la sobreexpresión constitutiva de Smr15C1 o Smr15C2 en un mutante de *S. meliloti* generado por delección de ambos *locus*. El análisis de los geles 2D-PAGE reveló la represión específica de la proteína LivK por Smr15C2. El mRNA/proteína codificado por el gen *livK* siempre fue catalogado como inducido en ausencia de Hfq tanto en nuestros estudios como en otros similares documentados en la bibliografía (Gao *et al.*, 2010; Barra-Bily *et al.*, 2010b), lo que estaría de acuerdo con la degradación de Smr15C2 en ausencia de la chaperona. El mensajero de *livK* también apareció representado en CoIP-RNA obtenido de bacterias en fase exponencial de crecimiento pero no en nuestra lista priorizada de candidatos generada por las predicciones *in silico*. Sin embargo, una búsqueda posterior reveló un apareamiento entre Smr15C2 y el mRNA de *livK* que involucra 8 nt del motivo SL1 del sRNA y la totalidad del RBS en el mRNA, coincidiendo con la región de éste más enriquecida en el CoIP-RNA (Fig. III-23). Esta interacción no es predicha para Smr15C1 con los mismos parámetros de búsqueda. En *A. tumefaciens* un abordaje similar reveló también la represión de LivK por el homólogo de Smr15C2 pero no por el de Smr15C1, a pesar de que en este caso las interacciones de cada transcrito con el mRNA involucran a series más extensas de nucleótidos (Wilms *et al.*, 2011).

En *A. tumefaciens*, LivK transporta el ácido  $\gamma$ -aminobutírico (GABA) que es exudado por las plantas, como respuesta a la infección por el patógeno, para estimular la degradación de la señal de *quorum sensing* N-(3-oxo-octanoil) homoserín lactona producida por las bacterias, atenuando así su virulencia (Chevrot *et al.*, 2006). Aunque en *S. meliloti* el transportador LivK no ha sido caracterizado experimentalmente, los estudios de homología *in silico* predicen su afinidad por un amplio espectro de aminoácidos de cadena ramificada (e.g. leucina). En *R. leguminosarum* los aminoácidos de cadena ramificada suministrados por la planta son transportados activamente al interior de los bacteroides maduros, o que inician su diferenciación, contribuyendo al control de su desarrollo y supervivencia intracelular (Prell *et al.*, 2009). En *S. meliloti* no está documentado un mecanismo similar, aunque la represión de Smr15C2 en la zona de fijación de nitrógeno de los nódulos maduros posibilitaría también el transporte de estas moléculas a través de LivK. En definitiva,

estas observaciones sugieren que aunque sRNAs homólogos pueden regular los mismos mRNAs en bacterias diferentes, la significación biológica de esta regulación podría diferir sustancialmente en cada sistema.

El modelo de actividad propuesto para los *trans*-sRNAs Smr15C1 y Smr15C2 en *S. meliloti* se presenta en Fig. III-34, que recopila las predicciones *in silico* y los hallazgos experimentales descritos y discutidos en los epígrafes anteriores. En conjunto, nuestros resultados sugieren que en *S. meliloti* estos dos transcritos homólogos se acumulan diferencialmente en respuesta a fluctuaciones en la concentración extracelular de fuentes de carbono diferentes para la regulación masiva, mediante mecanismos dependientes de la RNA chaperona Hfq, de la traducción de conjuntos independientes o parcialmente coincidentes de proteínas de transporte de nutrientes diversas, optimizando así a las reprogramaciones metabólicas impuestas por la oligotrofia, la limitación de nutrientes y otros estreses abióticos de suelo y rizosfera.

#### ***Smr7C regula el transportador del sideróforo rizobactina en S. meliloti 1021***

El ensayo genético de doble plásmido *in vivo* reveló que la expresión de Smr7C estimula la traducción de la fusión de la región 5' del gen *rhtA* a GFP. Este gen, localizado en el plásmido pSymA, codifica el receptor de membrana externa del sideróforo rizobactina cuya biosíntesis está especificada por el operón *rhbABCDE*. La transcripción de este operón y la del gen *rhtA* están reguladas positivamente por el producto de un octavo gen en el *cluster* denominado *rhrA*, que exhibe características típicas de un activador transcripcional del tipo AraC (Lynch *et al.*, 2001). El modelo de actividad propuesto para este ribo-regulador y las hipótesis de trabajo futuras se resumen en el panel derecho de la Fig. III-34.

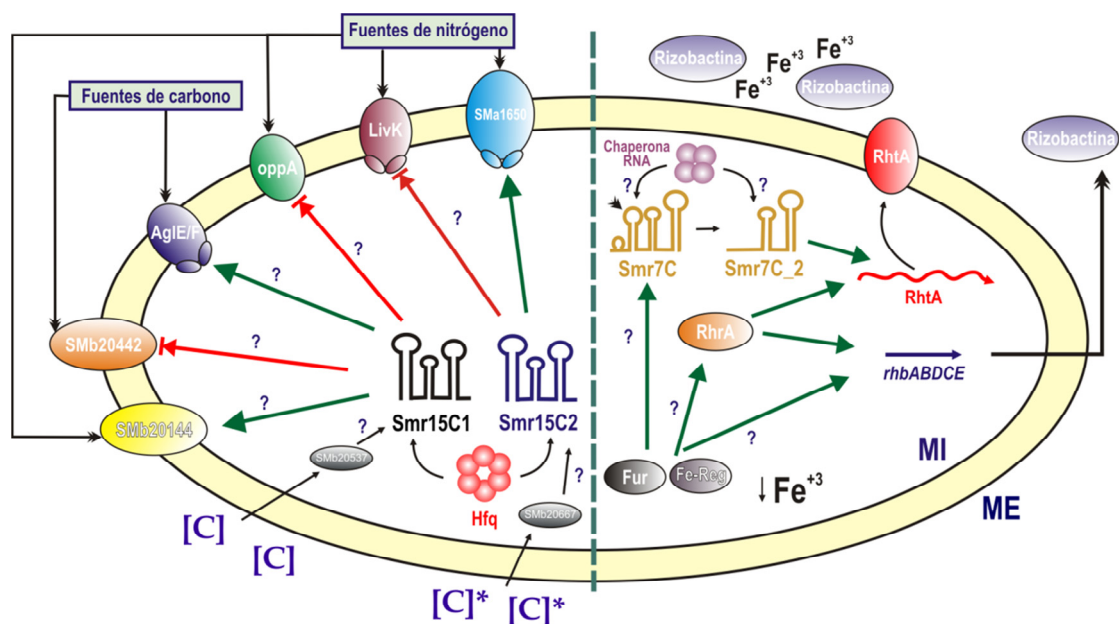
En *S. meliloti* los genes de la biosíntesis y transporte de rizobactina se expresan en bacterias en vida libre en respuesta a las limitaciones de hierro extracelular y posiblemente también en los primeros estadios de la simbiosis con alfalfa, pero no en los bacteroides maduros, siendo prescindibles para el establecimiento de una simbiosis eficiente (Lynch *et al.*, 2001). El hierro que demanda el proceso de fijación simbiótica de nitrógeno parece que se incorpora mayoritariamente al bacteroide a través de un transportador de hierro no acoplado a sideróforos codificado por el operón cromosómico *sitABCD* (Chao *et al.*, 2004). Nuestros resultados revelan efectivamente una correlación entre el perfil de expresión de Smr7C y el descrito para el mRNA de *rhtA*. En enterobacterias y otras bacterias Gram-negativas los genes que garantizan la homeostasis de hierro intracelular están mayoritariamente reprimidos por el regulador Fur en situaciones de alta disponibilidad del metal. En *S. meliloti*, sin

embargo, la función de Fur no se conoce con tanto detalle y todas las evidencias experimentales apuntan a que esta proteína no es un regulador global del metabolismo del hierro (Wexler *et al.*, 2003; Chao *et al.*, 2004). De hecho, en ausencia de Fur sólo los genes *sitABCD*, entre todos los conocidos en *S. meliloti* relacionados con la captación de hierro, aparecen inducidos mientras que todos los de la rizobactina son reprimidos (Chao *et al.*, 2004). El análisis de la región promotora de *Smr7C* no evidenció cajas Fur típicas como las reconocibles en el promotor del operón *sitABCD* lo que sugiere que la transcripción de este sRNA y del activador *rhtA* no está regulada directamente por Fur.

La simulación del plegamiento de la región 5'-UTR del mRNA de *rhtA* revela una posible horquilla inhibitoria que secuestra el sitio SD y el codón de inicio de la traducción, lo que justificaría la baja fluorescencia basal de su fusión a GFP observada en nuestros ensayos (Fig. III-33). El apareamiento de *Smr7C* con el mRNA de *rhtA* involucra a una secuencia de 9 nt localizados entre las posiciones +1 y +9, respecto al inicio de transcripción determinado para este mRNA (A. Becker y J.-P. Schlüter, comunicación personal), que son totalmente complementarios a la serie de nucleótidos situada entre las posiciones 52 y 60 de la molécula del sRNA, que en el transcrito primario se predicen apareados en el tallo de la segunda horquilla en la que se estructura este transcrito. El procesamiento mayoritario de *Smr7C* en las condiciones en las que se observó su máxima expresión (i.e. limitación de hierro y pH ácido) liberaría estos nucleótidos haciéndolos accesibles a la interacción con el mRNA. El apareamiento *Smr7C\_2-rhtA* liberaría el RBS y el codón AUG del mRNA haciéndolo accesible a los ribosomas y posibilitando la traducción del mensaje, en un mecanismo que precedimos independiente de Hfq.

La homeostasis de hierro está ribo-regulada en bacterias como *E. coli*, *Salmonella* o *P. aeruginosa*, en las que los homólogos del sRNA RyhB, cuya expresión es dependiente de Fur, reprimen la traducción de proteínas que acumulan o utilizan hierro (e.g. superóxido dismutasa SodB) cuando la disponibilidad de éste es limitada (Massé and Gottesman, 2002; Wilderman *et al.*, 2004). La actividad de RyhB es asistida por Hfq en estas especies bacterianas. En *B. subtilis*, sin embargo, la represión de las proteínas acumuladoras de hierro es mediada por otro sRNA también regulado por Fur, denominado FsrA, cuya actividad no depende de Hfq sino de al menos tres proteínas básicas (FbpA-C), que actúan como chaperonas de RNA en el proceso (Gaballa *et al.*, 2008). La posibilidad de la existencia de este tipo de proteínas en *S. meliloti* debe ser, por tanto, explorada.

En definitiva, Smr7C representaría un nuevo modelo de *trans* ribo-regulación de la homeostasis del hierro en bacterias, independiente del regulador Fur y de la RNA chaperona Hfq.



**Fig. III-34. Modelo de regulación propuesta para los sRNAs Smr15C1/C2 y Smr7C en *S. meliloti*.** La parte derecha representa la regulación propuesta para Smr15C1/C2 y la parte izquierda para Smr7C. Las flechas verdes indican inducción de la expresión génica y las rojas represión. Los interrogantes indican la ausencia de verificaciones experimentales confirmatorias, aunque todos los candidatos propuestos para Smr15C1/C2 presentan alguna evidencia experimental obtenida del análisis de perfiles transcriptómicos/proteómicos, RNA-Seq de Hfq. Las interacciones no marcadas con interrogantes son las verificadas mediante ensayo *in vivo* mediante doble plásmido. [C] [C]\*, fuentes de carbono; Fe-Reg, posible regulador del metabolismo del hierro.





## **ANEXO CAPÍTULO 3**



## OLIGONUCLEÓTIDOS UTILIZADOS

Tabla III-8. Oligonucleótidos utilizados en este trabajo.

Nombre	Aplicación	Secuencia
Cir7C	RACE 3'	5'-GCCTTGTCCCGCTTTCAAAA-3'
Cir7C-i	RACE 3'	5'-ATGAGGCCTTTGTCCTCATC-3'
Cir15C1	RACE 3'	5'-CCTCTGGAGGTTTGAAACCT-3'
Cir15C1-i	RACE 3'	5'-AACAGCTGCTGCAGCGGCTG-3'
Cir15C2	RACE 3'	5'-GTGCCACCGCAGCAGCTGT-3'
Cir15C2-i	RACE 3'	5'-TGGGAGGAGAAGCCACCAAAA-3'
57C	Mut. delección	5'-GCATGCTCCTGCGAATCGAGCCTGAA-3'
37C	Mut. delección	5'-TCTAGAGACCGGCTGCACATCGTAAT-3'
515C1-16	Mut. delección	5'-GCATGCGGATTGACGAGGTAGGTCTT-3'
315C1-16	Mut. delección	5'-TCTAGAGGCTTGATCCTGACCGAACA-3'
57C-i	Mut. delección	5'-GAGGTACCCGACCGAATCGCGCATT-3'
37C-i	Mut. delección	5'-GAGGTACCTTTGCCCGACGGTCGAAA-3'
515C1-i	Mut. delección	5'-GAGGTACCTTTGCCCGGCTTATGG-3'
315C1-i	Mut. delección	5'-GAGGTACCATATGGGTAGGGCCGTAG-3'
515C2-i	Mut. delección	5'-GAGGTACCACAGCCCTCAGTTGCGAA-3'
315C2-i	Mut. delección	5'-GAGGTACCTTGACCCGCAAGGATT-3'
sec7C	Mut. delección	5'-AAGGTCCATTGCGCTTTTCC-3'
sec7C-i	Mut. delección	5'-AATGCCTGCAATGAAGCCGA-3'
sec15C1	Mut. delección	5'-GAGGTACCACAGCCCTCAGTTGCGAA-3'
sec15C1-i	Mut. delección	5'-TGAGATGAAACCGTCGACTG-3'
sec15C2	Mut. delección	5'-ACTGGAACCTCTGACGATCC-3'
5Ery-Kpn	Mut. delección	5'-GGTACCTATAAAAAATAATTCTTGAC-3'
3Ery-Kpn	Mut. delección	5'-GGTACCGAGCGTTCACCGACAAAACA-3'
Fsyn	Mut. sobreexpresión	5'-CTATAAAAAATAATTCTTGACATA TTTGCTAAAATTGGTATAAATAG-3'
Rsyn	Mut. sobreexpresión	5'-GATCCTATTATACCAAATTTTAGCAAATA TGTCAGAATATTTTATAGGTAC-3'
SmrC7Fw	Mut. sobreexpresión	5'-GGATCCGAGCGGCACTCATACAATG-3'
SmrC7Rv	Mut. sobreexpresión	5'-GAGTCCGTCGGGCAAAAGGCAACAA-3'
SmrC15Fw	Mut. sobreexpresión	5'-GGATCCAGCTGGTGCATCTAGCCG-3'
SmrC15Rv	Mut. sobreexpresión	5'-GAGTCCGCGGGCAAAATGCCGAC-3'
SmrC16Fw	Mut. sobreexpresión	5'-GGATCCAGCTGATGCATCTTTGGT-3'
SmrC16Rv	Mut. sobreexpresión	5'-GAGTCTGGGGCTCCGCCAGGCAA-3'
secSRK	Mut. sobreexpresión	5'-TTCCATTCGCCATTCAGGCT-3'
SRKFw	Mut. sobreexpresión	5'-ACTAAAGGGATCCAAAGCTGGAGC-3'
SRKRv	Mut. sobreexpresión	5'-GCTCACAATTGGATCCAACATACGAG-3'
SRKRv2	Mut. sobreexpresión	5'-GGTCATATGCTGGATCCTGTGTGA-3'
Syn-i	Ensayo <i>in vivo</i>	5'-CGAGGATCCTATTATACCAAATTTTAGCA-3'
3_GFP_E	Ensayo <i>in vivo</i>	5'-GAATTCCTACTTGTACAGATTCGTC-3'
5SynH	Ensayo <i>in vivo</i>	5'-AAGCTTATAAAAAATAATTCTTGACA-3'
FlacZ	Ensayo <i>in vivo</i>	5'-AGATCTGGAGGTCTCCAACATGGACTACAAAAG ATGACGACGATAAAAACCATGATTACGGATTC-3'
GFP_i2	Ensayo <i>in vivo</i>	5'-CCGGATCCGCTAGCAAGGGCGAGGAGCTGTT-3'
FlacZ_59	Ensayo <i>in vivo</i>	5'-GCTAGCTTTTTTGGATCCCCATTCGCCAATCAG-3'
RhoIT_AS	Ensayo <i>in vivo</i>	5'-AGCTTACAGGGATAAAAAAAGCAGCACT GTAATCAGTGTCTTTGTGCTTTTTTTCAGAGCT-3'
RhoIT_S	Ensayo <i>in vivo</i>	5'-CTGAAAAAACGACAAAAGCAGCAC TGATTACAGTGTCTTTTTTATCCCTGTA-3'
eryD_F	Ensayo <i>in vivo</i>	5'-GCTAGCGCCGTCACCATCAATGATT-3'
exsE_F	Ensayo <i>in vivo</i>	5'-GGATCCTGTCCCAGCTTCCGGGTGA-3'
exsE_R	Ensayo <i>in vivo</i>	5'-GGATCCTGTGGTGCAGGGGGCGTT-3'
oppA_F(1)	Ensayo <i>in vivo</i>	5'-GGATCCTGGTGTCTCCGTGCAAGCAG-3'
oppA_F(2)	Ensayo <i>in vivo</i>	5'-GCTAGCCACTGTGTGATTAGTCCGG-3'
oppA_R	Ensayo <i>in vivo</i>	5'-GGATCCGTTCTTTGGGCTGATATC-3'

rhaRS_R	Ensayo <i>in vivo</i>	5'-GCTAGCAAAGCTGATACCGGCCAACA-3'
rhaS_F	Ensayo <i>in vivo</i>	5'-GGATCCTTCACGAGCCCGTTACAGA-3'
rhtA_F	Ensayo <i>in vivo</i>	5'-GCTAGCCATGGCCAAGCCGATAGCGC-3'
SMa0495_F	Ensayo <i>in vivo</i>	5'-GCTAGCCATTGCAACCGCCGACCCCA-3'
SMa0495_R	Ensayo <i>in vivo</i>	5'-GGATCCTAGAAGGCATCGAATTTCCA-3'
SMa1337_F	Ensayo <i>in vivo</i>	5'-GCTAGCGGCAAGCGCAATGCTTAAAA-3'
SMa1337_R	Ensayo <i>in vivo</i>	5'-GGATCCTCGCGACCCGCGCCCTG-3'
SMa1651/50_F	Ensayo <i>in vivo</i>	5'-GGATCCAGACCTATCCGCTCGCCGAA-3'
SMa1651_R	Ensayo <i>in vivo</i>	5'-GCTAGCGAACGAGCCGAACAATTCG-3'
SMb20127_R	Ensayo <i>in vivo</i>	5'-GGATCCCGCCTGATTGCCAAATATC-3'
SMb21432_F	Ensayo <i>in vivo</i>	5'-GCTAGCCCATTTCAGCGTTTCGGCGT-3'
SMb21432_R	Ensayo <i>in vivo</i>	5'-GGATCCTCCGGCCGACATTTGAGGAG-3'
SMc00245_F(1)	Ensayo <i>in vivo</i>	5'-GGATCCCACGAGGGCTTCGCATCGCC-3'
SMc00245_F(2)	Ensayo <i>in vivo</i>	5'-GCTAGCCACCATCGCGAGCGTGGCGA-3'
SMc00245_R	Ensayo <i>in vivo</i>	5'-GGATCCCTGATACITGTCCGAAACCG-3'
SMc02417_F	Ensayo <i>in vivo</i>	5'-GCTAGCATCGTTTATGGATTCCATCC-3'
SMc02417_R	Ensayo <i>in vivo</i>	5'-GGATCCCTCAAGAGCACGCAATTCG-3'
SMc03121_F(1)	Ensayo <i>in vivo</i>	5'-GCTAGCGGCCGTCACCATCAATGATT-3'
SMc03121_F(2)	Ensayo <i>in vivo</i>	5'-GGATCCTGTCCCGAGCTCCGGGTGA-3'
SMc03121_R	Ensayo <i>in vivo</i>	5'-GGATCCTCTGGTGCAGGGGGCGTT-3'
ttpCD_F	Ensayo <i>in vivo</i>	5'-GGATCCTGGTGTTCCTGTCAAGCAG-3'
ttpCD_R	Ensayo <i>in vivo</i>	5'-GCTAGCCACTGCTGTGATTAGTCCGG-3'
wggR_F	Ensayo <i>in vivo</i>	5'-GGATCCGTTTCGTTTGGGCCTGATATC-3'

## PLÁSMIDOS UTILIZADOS.

Tabla III-9. Construcciones utilizadas en este trabajo.

Nombre	Descripción
pGEMgsmr7C	pGEM-T con la región genómica <i>smr7C</i>
pGEMgsmr15C1/C2	pGEM-T con la región <i>smr15C1/C2</i>
pGEM $\Delta$ smr7C	pGEM-T con la región genómica y sin el ORF de <i>smr7C</i>
pGEM $\Delta$ smr15C1	pGEM-T con la región genómica y sin el ORF de <i>smr15C1</i>
pGEM $\Delta$ smr15C2	pGEM-T con la región genómica y sin el ORF de <i>smr15C1</i>
pGEM $\Delta$ smr15C1/C2	pGEM-T con la región genómica y sin los ORF <i>smr15C1/C2</i>
pGEMssDUTI	pGEM-T con el <i>cassette</i> de eritromicina
pGEM $\Delta$ smr7C-Ery	pGEM-T con la región genómica y sin el ORF de <i>smr7C</i> y el marcador Ery <sup>r</sup>
pGEM $\Delta$ smr15C1-Ery	pGEM-T con la región genómica y sin el ORF de <i>smr15C1</i> y el marcador Ery <sup>r</sup>
pGEM $\Delta$ smr15C2-Ery	pGEM-T con la región genómica y sin el ORF de <i>smr15C1</i> y el marcador Ery <sup>r</sup>
pGEM $\Delta$ smr15C1/C2-Ery	pGEM-T con la región genómica y sin los ORF <i>smr15C1/C2</i> y el marcador Ery <sup>r</sup>
pK18 $\Delta$ smr7C-Ery	pk18 <i>mobsacB</i> con la región genómica $\Delta$ smr7C-Ery
pK18 $\Delta$ smr15C1-Ery	pk18 <i>mobsacB</i> con la región genómica $\Delta$ smr15C1-Ery
pK18 $\Delta$ smr15C2-Ery	pk18 <i>mobsacB</i> con la región genómica $\Delta$ smr15C2-Ery
pK18 $\Delta$ smr15C1/C2-Ery	pk18 <i>mobsacB</i> con la región genómica $\Delta$ smr15C1/C2-Ery
pSRK*	pSRKKm sin diana BamHI
pSRK_C	pSRKKm sin el operador de lacIQ
pR_eGFP	Plásmido reportero control para la expresión de 5'-UTR::GFP
pRO_eGFP	Plásmido reportero control para la expresión de 5'-UTR::GFP en operones
pGEM $\Psi$ syn-eGFP	pGEM-T que expresa la GFP desde el $\Psi$ syn
pJB_eGFP	pJB3Tc19que expresa la GFP desde el $\Psi$ syn
pGEM_GFP-AUTR	pGEM-T que contiene la unidad $\Psi$ syn-eGFP para clonar 5'-UTR
pGEM_GFP-AUTR-T	pGEM-T que contiene la unidad $\Psi$ syn-eGFP para clonar 5'-UTR con el terminador T1
pGEM_GFP_O-T	pGEM-T que contiene la unidad $\Psi$ syn-eGFP para clonar 5'-UTR de operones

# **DISCUSIÓN GENERAL**



## DISCUSIÓN GENERAL

En la tercera edición del libro *The RNA World* (2006), un libro de referencia en el conocimiento sobre RNAs, se describió a los sRNAs bacterianos como una nueva clase de reguladores que había permanecido hasta la fecha inexplorada. Al igual que se había visto en eucariotas, estos sRNAs actúan modificando la expresión génica de sus mRNAs o modulando la actividad de proteínas, al igual que lo hacen los factores de transcripción clásicos proteicos.

Una de las cuestiones más interesantes en el estudio de los sRNAs bacterianos ha sido su origen en el mundo RNA. Aunque el apareamiento de bases entre RNAs ha sido una de las formas más tempranas de regulación, la pregunta de cuándo surgieron los sRNAs en el mundo bacteriano sigue siendo un misterio. Sin embargo, la rápida acumulación de información referente a secuencias genómicas junto con el interés de dilucidar el papel de sRNAs desconocidos en multitud de bacterias, está proporcionando indicios de cómo estas moléculas han evolucionado en la historia.

En el caso de los sRNAs dependientes de Hfq, que son el grupo mejor estudiado, la evolución de los RNAs reguladores parece haber sido rápida. Debido a la falta de similitudes entre secuencias y estructuras secundarias de la mayoría de estos transcritos no ha sido posible extraer una imagen clara de su evolución, incluso entre especies relacionadas (i.e. bacterias Gram-negativas y/o  $\gamma$ -proteobacterias). Sin embargo existen ciertas consideraciones generales o contribuciones que pueden tenerse en cuenta. En primer lugar, dado que la mayoría de este tipo de sRNAs actúan mediante apareamiento con sus mRNA diana, la evolución del sRNA y sus dianas debe ser conjunta. La mayoría de ellos comparten una conservación del dominio que interacciona con el mRNA (Sharma *et al.*, 2007). Un claro exponente es el caso de la multiplicidad de dianas por parte de algunos sRNAs, en los que, dada la similitud de las secuencias implicadas en la interacción entre ambas moléculas, se puede predecir que la evolución haya implicado un origen ancestral común. En segundo lugar, estos sRNAs pueden presentar una regulación conservada, siendo el estudio de los reguladores que controlan su expresión una herramienta muy útil a la hora de estudiar su evolución. Por ejemplo, un caso muy estudiado es la regulación de la disponibilidad de hierro por parte de Fur (Carpenter *et al.*, 2009). Este factor de transcripción y sus motivos de unión están muy conservados en un amplio abanico de grupos bacterianos que comprenden tanto a Gram-positivas como a Gram-negativas. Los sRNAs que contienen motivos de unión a Fur (e.g. RhyB de *E. coli* o PrrE de *Pseudomonas*) y regulan la respuesta de forma similar, sin embargo no comparten ninguna identidad en secuencia o estructura entre ellos (Wilderman *et al.*,



2004), por lo que posiblemente hayan divergido rápidamente y su evolución haya sido independiente. En tercer lugar, la reiteración en los genomas por parte de los sRNAs puede mostrar contextos en los que estas moléculas pueden extenderse en los grupos bacterianos. Al igual que los miRNAs, los sRNAs pueden formar familias de transcritos relacionadas entre sí. Los RNAs Qrr de *V. cholerae* se presentan en regiones cromosómicas separadas y poseen grados de variación diversos, mientras que los sRNAs OmrA y OmrB de *P. aeruginosa* se localizan en una posición adyacente y son muy similares en secuencia (Gottesman *et al.*, 2007). La capacidad de los sRNAs de duplicación y generación de copias ha sido relevante en la formación de nuevos sRNAs y representa un mecanismo de distribución de estas moléculas en grupos bacterianos. Además, la presencia de elementos móviles asociados a muchos sRNAs refleja que pueden estar sujetos a una transferencia horizontal. Es el caso por ejemplo de los sRNAs que se encuentran en islas de patogenicidad, presentes en *Salmonella* o *S. aureus* (Pichon and Felden, 2005).

Los ejemplos descritos anteriormente proporcionan evidencias de una evolución reciente de los sRNAs en bacterias y sugieren que los límites de su evolución son bastante lábiles pudiendo cambiar su secuencia y estructura relativamente rápido. Debido a esta velocidad de evolución no ha sido posible desarrollar evidencias sólidas de las fuentes de origen de estos sRNAs, dando pie a la libre especulación. No hay razones para pensar un camino evolutivo único para los sRNAs fuertemente divergentes por lo que todas las posibilidades pueden ser válidas. En la revisión Gottesman and Storz, 2010, los autores proponen una serie de posibles rutas evolutivas que pueden apoyar el origen de los sRNAs. Una de las hipótesis que se barajan se explica con la teoría del “ruido transcripcional”, que se sustenta en la existencia de una transcripción a bajo nivel en la célula de ciertas regiones que se soporta por la detección de actividades de transcripción antisentido de promotores sin detección de transcritos estables (Kawano *et al.*, 2005). Es relativamente fácil de imaginar que algunos de estos transcritos hayan supuesto una ventaja en la regulación de ciertos procesos y la evolución los haya seleccionado para convertirse en RNAs expresados a un nivel fisiológico. Otra fuente de sRNAs reguladores podrían ser los tRNAs o los ancestros de los tRNAs. Existen muchas evidencias de las relaciones entre sRNAs y tRNAs (e.g. tmRNA) o el sRNA de *E. coli* CyaR. Estudios recientes han demostrado una posible asociación entre Hfq y los tRNA, posiblemente porque la chaperona pueda modular su procesamiento (Zhang *et al.*, 2003; Lee and Feig, 2008). La habilidad de los tRNAs de asociarse con Hfq podría apoyar la idea de su ancestro común. Otra explicación al origen de los sRNAs implicaría a los mRNAs. El grupo de los *cis*-sRNAs y concretamente los *riboswitches* representan un estado en el que la identidad del sRNA está totalmente unida al mensajero. La capacidad de los

*riboswitches* de modular la expresión de los mRNA hace pensar que en algún momento alguno de ellos pudiera dar el salto para integrarse en una unidad transcripcional independiente.

En los últimos 5 años las técnicas de identificación de RNAs han dado ha conocer extensos catálogos de transcritos que no codifican proteínas y que suponen un nivel de la transcripción celular más abundante incluso que los propios mRNAs. Dada su versatilidad funcional y su implicación en mecanismos post-transcripcionales de regulación génica, los sRNAs han permitido dilucidar procesos clave en la adaptación y respuesta de las bacterias en sus nichos ecológicos. El suelo y concretamente la rizosfera son ambientes muy heterogéneos, en donde las bacterias que viven en ellos están sometidas a gran número de estreses externos que requieren una respuesta rápida y eficaz para su supervivencia. En este sentido, las bacterias del suelo suponen un modelo de estudio óptimo para la regulación mediada por sRNAs y el análisis de los endosimbiontes supone además una prospección de las interacciones procariota-eucariota.

Esta tesis doctoral comprende la identificación de sRNAs en el genoma de *S. meliloti*, cuyo RNoma únicamente contaba con un representante anotado en su genoma (tmRNA). Se ha abordado esta identificación utilizando desde aproximaciones clásicas basadas en genómica comparada y hasta la utilización de técnicas de secuenciación masiva. Estos resultados han catalogado 460 sRNAs que comprenden las categorías de *trans* y *cis*-sRNAs, la mayoría de ellos de nueva anotación. La comprensión del impacto de la chaperona Hfq en la biología de esta bacteria, también desarrollada en esta tesis, supone la determinación de las rutas susceptibles de sufrir regulación por RNAs no codificantes en el endosimbionte y, junto con los sRNAs identificados mediante coIP aporta una materia prima susceptible de ser utilizada en futuras líneas de investigación que persigan la caracterización de los procesos que intervienen en la interacción simbiótica de la bacteria con su huésped natural, la alfalfa. Con la primera aproximación a la función de sRNAs en *S. meliloti*, nuestro trabajo establece una serie de criterios y una estrategia basada en herramientas bioinformáticas y evidencias experimentales para la caracterización funcional de estos transcritos. Nuestros resultados presentan a Smr7C, Smr15C1 y Smr15C2 como ejemplos típicos de sRNAs que actúan en *trans* modificando la traducción de mRNAs que no están implicados de forma directa en procesos simbióticos pero que posiblemente intervengan en ajustes metabólicos de *S. meliloti* en el ambiente rizosférico o en estadios iniciales de la simbiosis.



# CONCLUSIONS



**1.** The combination of computational algorithms relying on probabilistic models to assess structure conservation and thermodynamic stability of RNA molecules was a reliable approach for predicting sRNA genes in the *Sinorhizobium meliloti* genome with high rates of sensitivity and specificity.

**2.** This *in silico* comparative genomics screen coupled to experimental validation of predictions by Northern hybridization and TAP-based RACE mapping led to the identification of ten sRNAs which had not been annotated previously in the *S. meliloti* genome. Three of these sRNAs are structural homologs of the housekeeping *E. coli* RNase P, SRP and 6S RNAs. The remaining seven transcripts are novel bacterial sRNAs that were termed Smr, which are conserved in phylogenetically related species of the order of the Rhizobiales and expressed differentially from independent transcriptional units with recognizable promoter and termination signatures. Therefore, they were catalogued as putative *trans*-acting sRNAs.

**3.** Deep-sequencing of the sub-population of transcripts associated to the *S. meliloti* RNA chaperone Hfq in free-living bacteria grown in rich medium revealed that this protein mostly binds mRNAs. Nonetheless, this strategy resulted in the identification of 226 *trans*-sRNAs and 234 *cis*-encoded antisense sRNAs, 69% of which were unknown in *S. meliloti*. However, it seems that this approach did not saturate the search for *trans*-RNAs expressed by this bacterium, suggesting that in *S. meliloti* co-exist large populations of Hfq-dependent and independent *trans*-acting riboregulators.

**4.** The RNA chaperone Hfq is a global post-transcriptional regulator of gene expression in *S. meliloti* influencing transport and metabolic pathways, rhizospheric competence, intracellular survival of differentiated bacteroids and symbiotic nitrogen fixation.

**5.** The *trans*-encoded Smr sRNAs Smr15C1, Smr15C2 and Smr7C are dispensable for optimal free-living growth of *S. meliloti* in rich medium as well as for the establishment of an efficient symbiosis with its cognate legume host alfalfa under laboratory conditions.

**6.** The seemingly homologous sRNAs Smr15C1 and Smr15C2 are differentially regulated in *S. meliloti* 1021 to act independently or additively, rather than hierarchically for the post-transcriptional regulation of different or partially overlapping arrays of ABC transporter genes. Smr15C1/C2-mediated riboregulation is predicted to be mainly exerted by interaction of unpaired nucleotides within the 5' hairpin loop of each sRNA with complementary nucleotides within or in the vicinity of the RBS of the target mRNAs in an Hfq-dependent manner.

**7.** The Smr7C sRNA exists in two stable forms in *S. meliloti* 1021 that accumulate under iron depletion in free-living and undifferentiated endosymbiotic bacteria to activate translation of the rhizobactin membrane transporter RhtA, most probably by melting an inhibitory hairpin at the 5'-UTR of the target mRNA through an Hfq-independent base-pairing interaction that is predicted to be favored by the post-transcriptional processing of the primary Smr7C transcript.

# **BIBLIOGRAFÍA**





- Aiba, H.** (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol.* **10**, (2), 134-9.
- Altman, S.** (1989). Ribonuclease P: an enzyme with a catalytic RNA subunit. *Adv Enzymol Relat Areas Mol Biol.* **62**, 1-36.
- Altman, S., Kirsebom, L. and Talbot, S.** (1993). Recent studies of ribonuclease P. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology.* **7**, (1), 7-14.
- Ansong, C., Yoon, H., Porwollik, S., Mottaz-Brewer, H., Petritis, B. O., Jaitly, N., Adkins, J. N., McClelland, M., Heffron, F. and Smith, R. D.** (2009). Global systems-level analysis of Hfq and SmpB deletion mutants in *Salmonella*: implications for virulence and global protein translation. *PLoS One.* **4**, (3), e4809.
- Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H. and Altuvia, S.** (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Current biology : CB.* **11**, (12), 941-50.
- Arnvig, K. B., Comas, I., Thomson, N. R., Houghton, J., Boshoff, H. I., Croucher, N. J., Rose, G., Perkins, T. T., Parkhill, J., Dougan, G. and Young, D. B.** (2011). Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog.* **7**, (11), e1002342.
- Avner, P. and Heard, E.** (2001). X-chromosome inactivation: counting, choice and initiation. *Nature reviews. Genetics.* **2**, (1), 59-67.
- Azam, T. A., Hiraga, S. and Ishihama, A.** (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes to cells : devoted to molecular & cellular mechanisms.* **5**, (8), 613-26.
- Babitzke, P. and Romeo, T.** (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current opinion in microbiology.* **10**, (2), 156-63.
- Bailey, T. L. and Elkan, C.** (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* **2**, 28-36.
- Barends, S., Barend, K. and van Wezel, G. P.** (2011). The tmRNA-tagging mechanism and the control of gene expression: a review. *Wiley Interdiscip Rev RNA.* **2**, (2), 233-46.
- Barnett, M. J., Fisher, R. F., Jones, T., Komp, C., Abola, A. P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J., Gurjal, M., Hong, A., Huizar, L., Hyman, R. W., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Palm, C., Peck, M. C., Surzycki, R., Wells, D. H., Yeh, K. C., Davis, R. W., Federspiel, N. A. and Long, S. R.** (2001). Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proceedings of the National Academy of Sciences of the United States of America.* **98**, (17), 9883-8.

- Barnett, M. J., Toman, C. J., Fisher, R. F. and Long, S. R.** (2004). A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proceedings of the National Academy of Sciences of the United States of America*. **101**, (47), 16636-41.
- Barra-Bily, L., Pandey, S. P., Trautwetter, A., Blanco, C. and Walker, G. C.** (2010). The *Sinorhizobium meliloti* RNA chaperone Hfq mediates symbiosis of *S. meliloti* and alfalfa. *J Bacteriol*. **192**, (6), 1710-8.
- Barra-Bily, L., Fontenelle, C., Jan, G., Flechard, M., Trautwetter, A., Pandey, S. P., Walker, G. C. and Blanco, C.** (2010). Proteomic alterations explain phenotypic changes in *Sinorhizobium meliloti* lacking the RNA chaperone Hfq. *J Bacteriol*. **192**, (6), 1719-29.
- Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L. and Breaker, R. R.** (2005). 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA*. **11**, (5), 774-84.
- Batut, J., Andersson, S. G. and O'Callaghan, D.** (2004). The evolution of chronic infection strategies in the alpha-proteobacteria. *Nature reviews. Microbiology*. **2**, (12), 933-45.
- Bejerano-Sagie, M. and Xavier, K. B.** (2007). The role of small RNAs in *quorum sensing*. *Current opinion in microbiology*. **10**, (2), 189-98.
- Belasco, J. G.** (2010). All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat Rev Mol Cell Biol*. **11**, (7), 467-78.
- Berghoff, B. A., Glaeser, J., Nuss, A. M., Zobawa, M., Lottspeich, F. and Klug, G.** (2011). Anoxygenic photosynthesis and photooxidative stress: a particular challenge for *Roseobacter*. *Environmental microbiology*. **13**, (3), 775-91.
- Beringer, J. E.** (1974). R factor transfer in *Rhizobium leguminosarum*. *Journal of General Microbiology*. **84**, (1), 188-98.
- Birnboim, H. C. and Doly, J.** (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids research*. **7**, (6), 1513-23.
- Bittner, A. N., Foltz, A. and Oke, V.** (2007). Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. *Journal of bacteriology*. **189**, (5), 1884-9.
- Blum, H., Beier, H. and Gross, H. J.** (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*. **8**, (2), 93-99.
- Bohn, C., Rigoulay, C., Chabelskaya, S., Sharma, C. M., Marchais, A., Skorski, P., Borezee-Durant, E., Barbet, R., Jacquet, E., Jacq, A., Gautheret, D., Felden, B., Vogel, J. and Bouloc, P.** (2010). Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. *Nucleic acids research*. **38**, (19), 6620-36.

- Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A. C., Benito, Y., Jacquier, A., Gaspin, C., Vandenesch, F. and Romby, P.** (2007). *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes & development*. **21**, (11), 1353-66.
- Boughammoura, A., Matzanke, B. F., Bottger, L., Reverchon, S., Lesuisse, E., Expert, D. and Franza, T.** (2008). Differential role of ferritins in iron metabolism and virulence of the plant-pathogenic bacterium *Erwinia chrysanthemi* 3937. *Journal of bacteriology*. **190**, (5), 1518-30.
- Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H. and Vogel, J.** (2008). Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Molecular cell*. **32**, (6), 827-37.
- Boysen, A., Moller-Jensen, J., Kallipolitis, B., Valentin-Hansen, P. and Overgaard, M.** (2010). Translational regulation of gene expression by an anaerobically induced small non-coding RNA in *Escherichia coli*. *J Biol Chem*. **285**, (14), 10690-702.
- Brantl, S.** (2007). Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Current opinion in microbiology*. **10**, (2), 102-9.
- Brownlee, G. G.** (1971). Sequence of 6S RNA of *E. coli*. *Nat New Biol*. **229**, (5), 147-9.
- Burrowes, E., Abbas, A., O'Neill, A., Adams, C. and O'Gara, F.** (2005). Characterisation of the regulatory RNA RsmB from *Pseudomonas aeruginosa* PAO1. *Research in microbiology*. **156**, (1), 7-16.
- Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G. and Luisi, B. F.** (2005). Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*. **437**, (7062), 1187-91.
- Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Puhler, A., Purnelle, B., Ramsperger, U., Renard, C., Thebault, P., Vandenbol, M., Weidner, S. and Galibert, F.** (2001). Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proceedings of the National Academy of Sciences of the United States of America*. **98**, (17), 9877-82.
- Carter, R. J., Dubchak, I. and Holbrook, S. R.** (2001). A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic acids research*. **29**, (19), 3928-38.
- Cavanagh, A. T., Klocko, A. D., Liu, X. and Wassarman, K. M.** (2008). Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70. *Mol Microbiol*. **67**, (6), 1242-56.
- Chao, Y. and Vogel, J.** (2010). The role of Hfq in bacterial pathogens. *Curr Opin Microbiol*. **13**, (1), 24-33.

- Chen, J. L., Blasco, M. A. and Greider, C. W.** (2000). Secondary structure of vertebrate telomerase RNA. *Cell*. **100**, (5), 503-14.
- Chen, S., Lesnik, E. A., Hall, T. A., Sampath, R., Griffey, R. H., Ecker, D. J. and Blyn, L. B.** (2002). A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems*. **65**, (2-3), 157-77.
- Christiansen, J. K., Larsen, M. H., Ingmer, H., Sogaard-Andersen, L. and Kallipolitis, B. H.** (2004). The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bacteriol*. **186**, (11), 3355-62.
- Christiansen, J. K., Nielsen, J. S., Ebersbach, T., Valentin-Hansen, P., Sogaard-Andersen, L. and Kallipolitis, B. H.** (2006). Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA*. **12**, (7), 1383-96.
- Church, G. M. and Gilbert, W.** (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. **81**, (7), 1991-5.
- Coornaert, A., Lu, A., Mandin, P., Springer, M., Gottesman, S. and Guillier, M.** (2010). MicA sRNA links the PhoP regulon to cell envelope stress. *Mol Microbiol*. **76**, (2), 467-79.
- Crick, F.** (1970). Central dogma of molecular biology. *Nature*. **227**, (5258), 561-3.
- Davis, B. M., Quinones, M., Pratt, J., Ding, Y. and Waldor, M. K.** (2005). Characterization of the small untranslated RNA RyhB and its regulon in *Vibrio cholerae*. *Journal of bacteriology*. **187**, (12), 4005-14.
- de Haseth, P. L. and Uhlenbeck, O. C.** (1980). Interaction of *Escherichia coli* host factor protein with oligoriboadenylates. *Biochemistry*. **19**, (26), 6138-46.
- Ding, Y., Davis, B. M. and Waldor, M. K.** (2004). Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol Microbiol*. **53**, (1), 345-54.
- Ditta, G., Stanfield, S., Corbin, D. and Helinski, D. R.** (1980). Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proceedings of the National Academy of Sciences of the United States of America*. **77**, (12), 7347-51.
- Dixon, R. and Kahn, D.** (2004). Genetic regulation of biological nitrogen fixation. *Nature reviews. Microbiology*. **2**, (8), 621-31.
- Do, C. B., Woods, D. A. and Batzoglou, S.** (2006). CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*. **22**, (14), e90-8.
- Dondrup, M., Goesmann, A., Bartels, D., Kalinowski, J., Krause, L., Linke, B., Rupp, O., Sczyrba, A., Puhler, A. and Meyer, F.** (2003). EMMA: a platform for consistent storage and efficient analysis of microarray data. *Journal of biotechnology*. **106**, (2-3), 135-46.
- Durand, S. and Storz, G.** (2010). Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol Microbiol*. **75**, (5), 1215-31.

- Eggenhofer, E., Haslbeck, M. and Scharf, B.** (2004). MotE serves as a new chaperone specific for the periplasmic motility protein, MotC, in *Sinorhizobium meliloti*. *Mol Microbiol.* **52**, (3), 701-12.
- Erdmann, V. A., Szymanski, M., Hochberg, A., Groot, N. and Barciszewski, J.** (2000). Non-coding, mRNA-like RNAs database Y2K. *Nucleic acids research.* **28**, (1), 197-200.
- Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F. and Gesteland, R. F.** (1997). Probing the structure of the *Escherichia coli* 10Sa RNA (tmRNA). *RNA.* **3**, (1), 89-103.
- Fender, A., Elf, J., Hampel, K., Zimmermann, B. and Wagner, E. G.** (2010). RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* **24**, (23), 2621-6.
- Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F. J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B. and Puhler, A.** (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N<sub>2</sub>-fixing endosymbiont *Sinorhizobium meliloti*. *Proceedings of the National Academy of Sciences of the United States of America.* **98**, (17), 9889-94.
- Fischer, H. M.** (1994). Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev.* **58**, (3), 352-86.
- Foussard, M., Garnerone, A. M., Ni, F., Soupene, E., Boistard, P. and Batut, J.** (1997). Negative autoregulation of the *Rhizobium meliloti* fixK gene is indirect and requires a newly identified regulator, FixT. *Mol Microbiol.* **25**, (1), 27-37.
- Frank, D. N. and Pace, N. R.** (1998). Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem.* **67**, 153-80.
- Franze de Fernandez, M. T., Eoyang, L. and August, J. T.** (1968). Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature.* **219**, (5154), 588-90.
- Fry, J., Wood, M. and Poole, P. S.** (2001). Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Molecular plant-microbe interactions : MPMI.* **14**, (8), 1016-25.
- Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dreano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F. J., Weidner, S., Wells, D. H., Wong, K., Yeh, K. C. and Batut, J.** (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science.* **293**, (5530), 668-72.

Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñiz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., García-Sotelo, J. S., López-Fuentes, A., Porrón-Sotelo, L., Alquicira-Hernández, S., Medina-Rivera, A., Martínez-Flores, I., Alquicira-Hernández, K., Martínez-Adame, R., Bonavides-Martínez, C., Miranda-Ríos, J., Huerta, A. M., Mendoza-Vargas, A., Collado-Torres, L., Taboada, B., Vega-Alvarado, L., Olvera, M., Olvera, L., Grande, R., Morett, E. and Collado-Vides, J. (2011). RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* **39**, (Database issue), D98-105.

Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñiz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., García-Sotelo, J. S., López-Fuentes, A., Porrón-Sotelo, L., Alquicira-Hernández, S., Medina-Rivera, A., Martínez-Flores, I., Alquicira-Hernández, K., Martínez-Adame, R., Bonavides-Martínez, C., Miranda-Ríos, J., Huerta, A. M., Mendoza-Vargas, A., Collado-Torres, L., Taboada, B., Vega-Alvarado, L., Olvera, M., Olvera, L., Grande, R., Morett, E. and Collado-Vides, J. (2011). RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* **39**, (Database issue), D98-105.

Gao, M., Barnett, M. J., Long, S. R. and Teplitski, M. (2010). Role of the *SinoRhizobium meliloti* global regulator Hfq in gene regulation and symbiosis. *Mol Plant Microbe Interact.* **23**, (4), 355-65.

Garcia-Rodriguez, F. M. and Toro, N. (2000). *SinoRhizobium meliloti* (nodulation formation efficiency) genes exhibit temporal and spatial expression patterns similar to those of genes involved in symbiotic nitrogen fixation. *Molecular plant-microbe interactions : MPMI.* **13**, (6), 583-91.

Gardiner, K. and Pace, N. R. (1980). RNase P of *Bacillus subtilis* has a RNA component. *J Biol Chem.* **255**, (16), 7507-9.

Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., Wilkinson, A. C., Finn, R. D., Griffiths-Jones, S., Eddy, S. R. and Bateman, A. (2009). Rfam: updates to the RNA families database. *Nucleic acids research.* **37**, (Database issue), D136-40.

Geissmann, T., Chevalier, C., Cros, M. J., Boisset, S., Fechter, P., Noirot, C., Schrenzel, J., Francois, P., Vandenesch, F., Gaspin, C. and Romby, P. (2009). A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic acids research.* **37**, (21), 7239-57.

Geissmann, T., Chevalier, C., Cros, M. J., Boisset, S., Fechter, P., Noirot, C., Schrenzel, J., Francois, P., Vandenesch, F., Gaspin, C. and Romby, P. (2009). A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic acids research.* **37**, (21), 7239-57.

- Geng, J., Song, Y., Yang, L., Feng, Y., Qiu, Y., Li, G., Guo, J., Bi, Y., Qu, Y., Wang, W., Wang, X., Guo, Z., Yang, R. and Han, Y. (2009). Involvement of the post-transcriptional regulator Hfq in *Yersinia pestis* virulence. *PLoS One*. **4**, (7), e6213.
- Giacomini, A., Ollero, F. J., Squartini, A. and Nuti, M. P. (1994). Construction of multipurpose gene cartridges based on a novel synthetic promoter for high-level gene expression in gram-negative bacteria. *Gene*. **144**, (1), 17-24.
- Gibson, K. E., Kobayashi, H. and Walker, G. C. (2008). Molecular determinants of a symbiotic chronic infection. *Annu Rev Genet*. **42**, 413-41.
- Gillet, R. and Felden, B. (2001). Emerging views on tmRNA-mediated protein tagging and ribosome rescue. *Mol Microbiol*. **42**, (4), 879-85.
- Glazebrook, J., Ichige, A. and Walker, G. C. (1993). A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. *Genes & development*. **7**, (8), 1485-97.
- Goethals, K., Van Montagu, M. and Holsters, M. (1992). Conserved motifs in a divergent nod box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proceedings of the National Academy of Sciences of the United States of America*. **89**, (5), 1646-50.
- Gong, W., Hao, B., Mansy, S. S., Gonzalez, G., Gilles-Gonzalez, M. A. and Chan, M. K. (1998). Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. *Proceedings of the National Academy of Sciences of the United States of America*. **95**, (26), 15177-82.
- Gong, Z., Zhu, J., Yu, G. and Zou, H. (2007). Disruption of *nifA* gene influences multiple cellular processes in *Sinorhizobium meliloti*. *J Genet Genomics*. **34**, (9), 783-9.
- Gorke, B. and Stulke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature reviews. Microbiology*. **6**, (8), 613-24.
- Gottesman, S. (2004). The small RNA regulators of *Escherichia coli*: roles and mechanisms\*. *Annual review of microbiology*. **58**, 303-28.
- Gottesman, S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet*. **21**, (7), 399-404.
- Gottesman, S. and Storz, G. (2010). Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb Perspect Biol*.
- Griffin, B. E. (1971). Separation of 32P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of '4.5 S' and 5 S from *E. coli*. *FEBS letters*. **15**, (3), 165-168.
- Grosshans, H. and Slack, F. J. (2002). Micro-RNAs: small is plentiful. *The Journal of cell biology*. **156**, (1), 17-21.
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. and Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic acids research*. **36**, (Web Server issue), W70-4.



- Gu, J., Chen, Y. and Reddy, R.** (1998). Small RNA database. *Nucleic acids research*. **26**, (1), 160-2.
- Guillier, M. and Gottesman, S.** (2008). The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic acids research*. **36**, (21), 6781-94.
- Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E. and Gross, C. A.** (2007). Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in *Escherichia coli*. *J Bacteriol*. **189**, (5), 1963-73.
- Gutierrez, P., Li, Y., Osborne, M. J., Pomerantseva, E., Liu, Q. and Gehring, K.** (2005). Solution structure of the carbon storage regulator protein CsrA from *Escherichia coli*. *Journal of bacteriology*. **187**, (10), 3496-501.
- Harris, M. E. and Christian, E. L.** (2003). Recent insights into the structure and function of the ribonucleoprotein enzyme ribonuclease P. *Curr Opin Struct Biol*. **13**, (3), 325-33.
- Heidrich, N., Moll, I. and Brantl, S.** (2007). In vitro analysis of the interaction between the small RNA SR1 and its primary target ahrC mRNA. *Nucleic acids research*. **35**, (13), 4331-46.
- Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. and Luhrmann, R.** (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *Embo J*. **14**, (9), 2076-88.
- Heuer, H., Krsek, M., Baker, P., Smalla, K. and Wellington, E. M.** (1997). Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Applied and environmental microbiology*. **63**, (8), 3233-41.
- Heurlier, K., Dénervaud, V. and Haas, D.** (2006). Impact of quorum sensing on fitness of *Pseudomonas aeruginosa*. *Int J Med Microbiol*. **296**, (2-3), 93-102.
- Heurlier, K., Williams, F., Heeb, S., Dormond, C., Pessi, G., Singer, D., Camara, M., Williams, P. and Haas, D.** (2004). Positive control of swarming, rhamnolipid synthesis, and lipase production by the posttranscriptional RsmA/RsmZ system in *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology*. **186**, (10), 2936-45.
- Hindley, J.** (1967). Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *Journal of molecular biology*. **30**, (1), 125-36.
- Hori, K. and Yanazaki, Y.** (1974). Nucleotide sequence specific interaction of host factor I with bacteriophage Q beta RNA. *FEBS letters*. **43**, (1), 20-2.
- Huang, B., Jones, S. A., Brandenburg, B. and Zhuang, X.** (2008). Whole-cell 3D STORM reveals interactions between cellular structures with nanometer-scale resolution. *Nature methods*. **5**, (12), 1047-52.

- Ikeda, Y., Yagi, M., Morita, T. and Aiba, H.** (2011). Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in *Escherichia coli*. *Mol Microbiol.* **79**, (2), 419-32.
- Ish-Horowicz, D. and Burke, J. F.** (1981). Rapid and efficient cosmid cloning. *Nucleic acids research.* **9**, (13), 2989-98.
- Jacques, J. F., Jang, S., Prevost, K., Desnoyers, G., Desmarais, M., Imlay, J. and Masse, E.** (2006). RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in *Escherichia coli*. *Mol Microbiol.* **62**, (4), 1181-90.
- Jasiecki, J. and Wegrzyn, G.** (2005). Localization of *Escherichia coli* poly(A) polymerase I in cellular membrane. *Biochemical and biophysical research communications.* **329**, (2), 598-602.
- Joanny, G., Le Derout, J., Brechemier-Baey, D., Labas, V., Vinh, J., Regnier, P. and Hajnsdorf, E.** (2007). Polyadenylation of a functional mRNA controls gene expression in *Escherichia coli*. *Nucleic acids research.* **35**, (8), 2494-502.
- Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E. and Walker, G. C.** (2007). How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol.* **5**, (8), 619-33.
- Jordan, D. C.** (1984). Family III. Rhizobiaceae. *Baltimore: Willians and Wilkins.*
- Kajitani, M., Kato, A., Wada, A., Inokuchi, Y. and Ishihama, A.** (1994). Regulation of the *Escherichia coli* *hfq* gene encoding the host factor for phage Q beta. *J Bacteriol.* **176**, (2), 531-4.
- Kalamorz, F., Reichenbach, B., Marz, W., Rak, B. and Gorke, B.** (2007). Feedback control of glucosamine-6-phosphate synthase GlnS expression depends on the small RNA GlnZ and involves the novel protein YhbJ in *Escherichia coli*. *Mol Microbiol.* **65**, (6), 1518-33.
- Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J. and Nagai, K.** (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell.* **96**, (3), 375-87.
- Kaminski, P. A. and Elmerich, C.** (1998). The control of *Azorhizobium caulinodans* *nifA* expression by oxygen, ammonia and by the HF-I-like protein, NrfA. *Mol Microbiol.* **28**, (3), 603-13.
- Kaminski, P. A., Desnoues, N. and Elmerich, C.** (1994). The expression of *nifA* in *Azorhizobium caulinodans* requires a gene product homologous to *Escherichia coli* HF-I, an RNA-binding protein involved in the replication of phage Q beta RNA. *Proceedings of the National Academy of Sciences of the United States of America.* **91**, (11), 4663-7.

- Kawamoto, H., Morita, T., Shimizu, A., Inada, T. and Aiba, H.** (2005). Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in *Escherichia coli*. *Genes & development*. **19**, (3), 328-38.
- Kawano, M., Reynolds, A. A., Miranda-Rios, J. and Storz, G.** (2005). Detection of 5'- and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic acids research*. **33**, (3), 1040-50.
- Kay, E., Humair, B., Denervaud, V., Riedel, K., Spahr, S., Eberl, L., Valverde, C. and Haas, D.** (2006). Two GacA-dependent small RNAs modulate the quorum-sensing response in *Pseudomonas aeruginosa*. *Journal of bacteriology*. **188**, (16), 6026-33.
- Keenan, R. J., Freymann, D. M., Stroud, R. M. and Walter, P.** (2001). The signal recognition particle. *Annual review of biochemistry*. **70**, 755-75.
- Keiler, K. C., Waller, P. R. and Sauer, R. T.** (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*. **271**, (5251), 990-3.
- Khan, S. R., Gaines, J., Roop, R. M. and Farrand, S. K.** (2008). Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid *quorum sensing*. *Appl Environ Microbiol*. **74**, (16), 5053-62.
- Khemici, V. and Carpousis, A. J.** (2004). The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol*. **51**, (3), 777-90.
- Khemici, V., Poljak, L., Luisi, B. F. and Carpousis, A. J.** (2008). The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol Microbiol*. **70**, (4), 799-813.
- Kingsford, C. L., Ayanbule, K. and Salzberg, S. L.** (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome biology*. **8**, (2), R22.
- Kiss, T.** (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *Embo J*. **20**, (14), 3617-22.
- Kitphati, W., Ngok-Ngam, P., Suwanmaneerat, S., Sukchawalit, R. and Mongkolsuk, S.** (2007). *Agrobacterium tumefaciens* fur has important physiological roles in iron and manganese homeostasis, the oxidative stress response, and full virulence. *Applied and environmental microbiology*. **73**, (15), 4760-8.
- Koleva, R. I., Austin, C. A., Kowaleski, J. M., Neems, D. S., Wang, L., Vary, C. P. and Schlux, P. J.** (2006). Interactions of ribosomal protein S1 with DsrA and rpoS mRNA. *Biochemical and biophysical research communications*. **348**, (2), 662-8.
- Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M., 2nd and Peterson, K. M.** (1994). pBBR1MCS: a broad-host-range cloning vector. *Biotechniques*. **16**, (5), 800-2.

- Krol, E. and Becker, A.** (2004). Global transcriptional analysis of the phosphate starvation response in *SinoRhizobium meliloti* strains 1021 and 2011. *Molecular genetics and genomics : MGG.* **272**, (1), 1-17.
- Kulesus, R. R., Diaz-Perez, K., Slechta, E. S., Eto, D. S. and Mulvey, M. A.** (2008). Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. *Infect Immun.* **76**, (7), 3019-26.
- Kulkarni, P. R., Cui, X., Williams, J. W., Stevens, A. M. and Kulkarni, R. V.** (2006). Prediction of CsrA-regulating small RNAs in bacteria and their experimental verification in *Vibrio fischeri*. *Nucleic acids research.* **34**, (11), 3361-9.
- Laemmli, U. K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* **227**, (5259), 680-5.
- Landt, S. G., Abeliuk, E., McGrath, P. T., Lesley, J. A., McAdams, H. H. and Shapiro, L.** (2008). Small non-coding RNAs in *Caulobacter crescentus*. *Mol Microbiol.* **68**, (3), 600-14.
- Lenz, D. H., Miller, M. B., Zhu, J., Kulkarni, R. V. and Bassler, B. L.** (2005). CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol Microbiol.* **58**, (4), 1186-202.
- Lenz, D. H., Mok, K. C., Lilley, B. N., Kulkarni, R. V., Wingreen, N. S. and Bassler, B. L.** (2004). The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell.* **118**, (1), 69-82.
- Leonard, L. T.** (1943). A Simple Assembly for Use in the Testing of Cultures of Rhizobia. *Journal of bacteriology.* **45**, (6), 523-7.
- Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T. and Geurts, R.** (2005). Formation of organelle-like N<sub>2</sub>-fixing symbiosomes in legume root nodules is controlled by DMI2. *Proceedings of the National Academy of Sciences of the United States of America.* **102**, (29), 10375-80.
- Lindell, M., Romby, P. and Wagner, E. G.** (2002). Lead(II) as a probe for investigating RNA structure in vivo. *RNA.* **8**, (4), 534-41.
- Liu, J. M., Livny, J., Lawrence, M. S., Kimball, M. D., Waldor, M. K. and Camilli, A.** (2009). Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic acids research.* **37**, (6), e46.
- Liu, J., Wang, J. T., Hu, J. and Tian, B.** (2005). A method for aligning RNA secondary structures and its application to RNA motif detection. *BMC Bioinformatics.* **6**, 89.
- Liu, M. Y., Gui, G., Wei, B., Preston, J. F., 3rd, Oakford, L., Yuksel, U., Giedroc, D. P. and Romeo, T.** (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem.* **272**, (28), 17502-10.

- Livny, J. and Waldor, M. K.** (2007). Identification of small RNAs in diverse bacterial species. *Current opinion in microbiology*. **10**, (2), 96-101.
- Livny, J., Brencic, A., Lory, S. and Waldor, M. K.** (2006). Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic acids research*. **34**, (12), 3484-93.
- Livny, J., Fogel, M. A., Davis, B. M. and Waldor, M. K.** (2005). sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic acids research*. **33**, (13), 4096-105.
- Lybecker, M. C., Abel, C. A., Feig, A. L. and Samuels, D. S.** (2010). Identification and function of the RNA chaperone Hfq in the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol*. **78**, (3), 622-35.
- Lynch, D., O'Brien, J., Welch, T., Clarke, P., Cuiv, P. O., Crosa, J. H. and O'Connell, M.** (2001). Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. *Journal of bacteriology*. **183**, (8), 2576-85.
- MacLellan, S. R., MacLean, A. M. and Finan, T. M.** (2006). Promoter prediction in the rhizobia. *Microbiology*. **152**, (Pt 6), 1751-63.
- Majdalani, N., Vanderpool, C. K. and Gottesman, S.** (2005). Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol*. **40**, (2), 93-113.
- Mandin, P. and Gottesman, S.** (2010). Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *Embo J*. **29**, (18), 3094-107.
- Marlow, V. L., Haag, A. F., Kobayashi, H., Fletcher, V., Scocchi, M., Walker, G. C. and Ferguson, G. P.** (2009). Essential role for the BacA protein in the uptake of a truncated eukaryotic peptide in *Sinorhizobium meliloti*. *Journal of bacteriology*. **191**, (5), 1519-27.
- Masse, E. and Gottesman, S.** (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*. **99**, (7), 4620-5.
- Masse, E., Vanderpool, C. K. and Gottesman, S.** (2005). Effect of RyhB small RNA on global iron use in *Escherichia coli*. *Journal of bacteriology*. **187**, (20), 6962-71.
- Mauchline, T. H., Fowler, J. E., East, A. K., Sartor, A. L., Zaheer, R., Hosie, A. H., Poole, P. S. and Finan, T. M.** (2006). Mapping the *SinoRhizobium meliloti*1021 solute-binding protein-dependent transportome. *Proceedings of the National Academy of Sciences of the United States of America*. **103**, (47), 17933-8.
- McNealy, T. L., Forsbach-Birk, V., Shi, C. and Marre, R.** (2005). The Hfq homolog in *Legionella pneumophila* demonstrates regulation by LetA and RpoS and interacts with the global regulator CsrA. *J Bacteriol*. **187**, (4), 1527-32.

- Meade, H. M., Long, S. R., Ruvkun, G. B., Brown, S. E. and Ausubel, F. M.** (1982). Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *J Bacteriol.* **149**, (1), 114-22.
- Metzker, M. L.** (2010). Sequencing technologies - the next generation. *Nature reviews. Genetics.* **11**, (1), 31-46.
- Miczak, A., Srivastava, R. A. and Apirion, D.** (1991). Location of the RNA-processing enzymes RNase III, RNase E and RNase P in the *Escherichia coli* cell. *Mol Microbiol.* **5**, (7), 1801-10.
- Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D. and Feig, A. L.** (2004). *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol.* **11**, (12), 1206-14.
- Moller, T., Franch, T., Udesen, C., Gerdes, K. and Valentin-Hansen, P.** (2002). Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes & development.* **16**, (13), 1696-706.
- Morita, T., Maki, K. and Aiba, H.** (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & development.* **19**, (18), 2176-86.
- Mulcahy, H., O'Callaghan, J., O'Grady, E. P., Adams, C. and O'Gara, F.** (2006). The posttranscriptional regulator RsmA plays a role in the interaction between *Pseudomonas aeruginosa* and human airway epithelial cells by positively regulating the type III secretion system. *Infection and immunity.* **74**, (5), 3012-5.
- Nawrocki, E. P., Kolbe, D. L. and Eddy, S. R.** (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics.* **25**, (10), 1335-7.
- Nguyen, V. T., Kiss, T., Michels, A. A. and Bensaude, O.** (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature.* **414**, (6861), 322-5.
- Nielsen, J. S., Lei, L. K., Ebersbach, T., Olsen, A. S., Klitgaard, J. K., Valentin-Hansen, P. and Kallipolitis, B. H.** (2010). Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Res.* **38**, (3), 907-19.
- Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., Arkin, A. P., Mironov, A. A. and Dubchak, I.** (2010). RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* **38**, (Web Server issue), W299-307.
- Novick, R. P. and Geisinger, E.** (2008). *Quorum sensing* in staphylococci. *Annu Rev Genet.* **42**, 541-64.
- Ogawa, J. and Long, S. R.** (1995). The *Rhizobium meliloti* *groEL* locus is required for regulation of early *nod* genes by the transcription activator NodD. *Genes & development.* **9**, (6), 714-29.

- Oglesby, A. G., Murphy, E. R., Iyer, V. R. and Payne, S. M.** (2005). Fur regulates acid resistance in *Shigella flexneri* via RyhB and ydeP. *Mol Microbiol.* **58**, (5), 1354-67.
- Olivares, J., Casadesus, J. and Bedmar, E. J.** (1980). Method for Testing Degree of Infectivity of *Rhizobium meliloti* Strains. *Applied and environmental microbiology.* **39**, (5), 967-70.
- Opdyke, J. A., Kang, J. G. and Storz, G.** (2004). GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol.* **186**, (20), 6698-705.
- Otaka, H., Ishikawa, H., Morita, T. and Aiba, H.** (2011). PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proceedings of the National Academy of Sciences of the United States of America.* **108**, (32), 13059-64.
- Pandey, S. P., Minesinger, B. K., Kumar, J. and Walker, G. C.** (2011). A highly conserved protein of unknown function in *SinoRhizobium meliloti* affects sRNA regulation similar to Hfq. *Nucleic acids research.* **39**, (11), 4691-708.
- Papenfort, K. and Vogel, J.** (2009). Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Research in microbiology.* **160**, (4), 278-87.
- Papenfort, K. and Vogel, J.** (2010). Regulatory RNA in bacterial pathogens. *Cell host & microbe.* **8**, (1), 116-27.
- Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M. and Vogel, J.** (2010). Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci U S A.*
- Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J. C. and Vogel, J.** (2006). SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol.* **62**, (6), 1674-88.
- Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. and Barford, D.** (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Molecular cell.* **33**, (2), 204-14.
- Passalacqua, K. D., Varadarajan, A., Ondov, B. D., Okou, D. T., Zwick, M. E. and Bergman, N. H.** (2009). Structure and complexity of a bacterial transcriptome. *J Bacteriol.* **191**, (10), 3203-11.
- Pavesi, G., Mauri, G., Stefani, M. and Pesole, G.** (2004). RNAProfile: an algorithm for finding conserved secondary structure motifs in unaligned RNA sequences. *Nucleic acids research.* **32**, (10), 3258-69.
- Peer, A. and Margalit, H.** (2011). Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. *Journal of bacteriology.* **193**, (7), 1690-701.

- Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. and Vogel, J.** (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nature structural & molecular biology*. **16**, (8), 840-6.
- Pfeiffer, V., Sittka, A., Tomer, R., Tedin, K., Brinkmann, V. and Vogel, J.** (2007). A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. *Mol Microbiol*. **66**, (5), 1174-91.
- Pobigaylo, N., Wetter, D., Szymczak, S., Schiller, U., Kurtz, S., Meyer, F., Nattkemper, T. W. and Becker, A.** (2006). Construction of a large signature-tagged mini-Tn5 transposon library and its application to mutagenesis of *Sinorhizobium meliloti*. *Appl Environ Microbiol*. **72**, (6), 4329-37.
- Postic, G., Frapy, E., Dupuis, M., Dubail, I., Livny, J., Charbit, A. and Meibom, K. L.** (2010). Identification of small RNAs in *Francisella tularensis*. *BMC Genomics*. 625.
- Poulsen, L. K., Refn, A., Molin, S. and Andersson, P.** (1991). The *gef* gene from *Escherichia coli* is regulated at the level of translation. *Mol Microbiol*. **5**, (7), 1639-48.
- Prell, J. and Poole, P.** (2006). Metabolic changes of rhizobia in legume nodules. *Trends in microbiology*. **14**, (4), 161-8.
- Prell, J., Bourdes, A., Karunakaran, R., Lopez-Gomez, M. and Poole, P.** (2009). Pathway of gamma-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. *Journal of bacteriology*. **191**, (7), 2177-86.
- Prevost, K., Salvail, H., Desnoyers, G., Jacques, J. F., Phaneuf, E. and Masse, E.** (2007). The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol Microbiol*. **64**, (5), 1260-73.
- Pueppke, S. G. and Broughton, W. J.** (1999). *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular plant-microbe interactions : MPMI*. **12**, (4), 293-318.
- Ramos, C. G., Sousa, S. A., Grilo, A. M., Feliciano, J. R. and Leitao, J. H.** (2011). The second RNA chaperone, Hfq2, is also required for survival under stress and full virulence of *Burkholderia cenocepacia* J2315. *Journal of bacteriology*. **193**, (7), 1515-26.
- Regalia, M., Rosenblad, M. A. and Samuelsson, T.** (2002). Prediction of signal recognition particle RNA genes. *Nucleic acids research*. **30**, (15), 3368-77.
- Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T. and Mondragon, A.** (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*. **468**, (7325), 784-9.
- Repoila, F. and Gottesman, S.** (2003). Temperature sensing by the *dsrA* promoter. *Journal of bacteriology*. **185**, (22), 6609-14.



**Repoila, F., Majdalani, N. and Gottesman, S.** (2003). Small non-coding RNAs, coordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol Microbiol.* **48**, (4), 855-61.

**Rigaud, J. and Puppo, A.** (1975). Indole-3 Acetic Acid Catabolism by Soybean Bacteroids. *J. Gen. Microbiol.* **88**, 223-228.

**Rivas, E.** (2005). Evolutionary models for insertions and deletions in a probabilistic modeling framework. *BMC Bioinformatics.* **6**, 63.

**Rivas, E., Klein, R. J., Jones, T. A. and Eddy, S. R.** (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol.* **11**, (17), 1369-73.

**Robertsen, B. K., Aman, P., Darvill, A. G., McNeil, M. and Albersheim, P.** (1981). Host-Symbiont Interactions : V. the structure of acidic extracellular polysaccharides secreted by *Rhizobium leguminosarum* and *Rhizobium trifolii*. *Plant physiology.* **67**, (3), 389-400.

**Robertson, G. T. and Roop, R. M., Jr.** (1999). The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol Microbiol.* **34**, (4), 690-700.

**Robledo, M., Jimenez-Zurdo, J. I., Velazquez, E., Trujillo, M. E., Zurdo-Pineiro, J. L., Ramirez-Bahena, M. H., Ramos, B., Diaz-Minguez, J. M., Dazzo, F., Martinez-Molina, E. and Mateos, P. F.** (2008). *Rhizobium* cellulase CelC2 is essential for primary symbiotic infection of legume host roots. *Proceedings of the National Academy of Sciences of the United States of America.* **105**, (19), 7064-9.

**Roop, R. M., 2nd, Robertson, G. T., Ferguson, G. P., Milford, L. E., Winkler, M. E. and Walker, G. C.** (2002). Seeking a niche: putative contributions of the *hfq* and *bacA* gene products to the successful adaptation of the brucellae to their intracellular home. *Vet Microbiol.* **90**, (1-4), 349-63.

**Ruvkun, G.** (2001). Molecular biology. Glimpses of a tiny RNA world. *Science.* **294**, (5543), 797-9.

**Samarsky, D. A., Ferbeyre, G., Bertrand, E., Singer, R. H., Cedergren, R. and Fournier, M. J.** (1999). A small nucleolar RNA:ribozyme hybrid cleaves a nucleolar RNA target in vivo with near-perfect efficiency. *Proceedings of the National Academy of Sciences of the United States of America.* **96**, (12), 6609-14.

**Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1989). Molecular cloning. A laboratory Manual. *Cold Spring Harbor Laboratory Press.* **Cold Spring Harbor, NY.**

**Sauer, E. and Weichenrieder, O.** (2011). Structural basis for RNA 3'-end recognition by Hfq. *Proceedings of the National Academy of Sciences of the United States of America.* **108**, (32), 13065-70.

**Sauter, C., Basquin, J. and Suck, D.** (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res.* **31**, (14), 4091-8.

- Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G. and Puhler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene*. **145**, (1), 69-73.
- Schluter, J. P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Janicke, S., Becker, J. D., Giegerich, R. and Becker, A. (2010). A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics*. **11**, 245.
- Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P. and Brennan, R. G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J*. **21**, (13), 3546-56.
- Schwarz, G., Mendel, R. R. and Ribbe, M. W. (2009). Molybdenum cofactors, enzymes and pathways. *Nature*. **460**, (7257), 839-47.
- Serganov, A. (2009). The long and the short of riboswitches. *Curr Opin Struct Biol*. **19**, (3), 251-9.
- Serganov, A., Huang, L. and Patel, D. J. (2009). Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature*. **458**, (7235), 233-7.
- Sharma, A. K. and Payne, S. M. (2006). Induction of expression of *hfq* by DksA is essential for *Shigella flexneri* virulence. *Mol Microbiol*. **62**, (2), 469-79.
- Sharma, C. M. and Vogel, J. (2009). Experimental approaches for the discovery and characterization of regulatory small RNA. *Curr Opin Microbiol*. **12**, (5), 536-46.
- Sharma, C. M., Darfeuille, F., Plantinga, T. H. and Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes & development*. **21**, (21), 2804-17.
- Silvaggi, J. M., Perkins, J. B. and Losick, R. (2005). Small untranslated RNA antitoxin in *Bacillus subtilis*. *Journal of bacteriology*. **187**, (19), 6641-50.
- Silvaggi, J. M., Perkins, J. B. and Losick, R. (2006). Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. *Journal of bacteriology*. **188**, (2), 532-41.
- Simpson, L., Thiemann, O. H., Savill, N. J., Alfonzo, J. D. and Maslov, D. A. (2000). Evolution of RNA editing in trypanosome mitochondria. *Proceedings of the National Academy of Sciences of the United States of America*. **97**, (13), 6986-93.
- Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., Hinton, J. C. and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet*. **4**, (8), e1000163.
- Sittka, A., Pfeiffer, V., Tedin, K. and Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol*. **63**, (1), 193-217.

- Sittka, A., Sharma, C. M., Rolle, K. and Vogel, J.** (2009). Deep sequencing of *Salmonella* RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol.* **6**, (3), 266-75.
- Siu, F. Y., Spangord, R. J. and Doudna, J. A.** (2007). SRP RNA provides the physiologically essential GTPase activation function in cotranslational protein targeting. *RNA.* **13**, (2), 240-50.
- Sleutels, F., Zwart, R. and Barlow, D. P.** (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature.* **415**, (6873), 810-3.
- Sobrero, P. and Valverde, C.** (2011). Evidences of autoregulation of hfq expression in *Sinorhizobium meliloti* strain 2011. *Archives of Microbiology.* 1-11.
- Sonnleitner, E., Abdou, L. and Haas, D.** (2009). Small RNA as global regulator of carbon catabolite repression in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America.* **106**, (51), 21866-71.
- Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jager, K. E. and Blasi, U.** (2003). Reduced virulence of a hfq mutant of *Pseudomonas aeruginosa* O1. *Microb Pathog.* **35**, (5), 217-28.
- Sonnleitner, E., Schuster, M., Sorger-Domenigg, T., Greenberg, E. P. and Blasi, U.** (2006). Hfq-dependent alterations of the transcriptome profile and effects on *quorum sensing* in *Pseudomonas aeruginosa*. *Mol Microbiol.* **59**, (5), 1542-58.
- Sonnleitner, E., Sorger-Domenigg, T., Madej, M. J., Findeiss, S., Hackermuller, J., Huttenhofer, A., Stadler, P. F., Blasi, U. and Moll, I.** (2008). Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. *Microbiology.* **154**, (Pt 10), 3175-87.
- Soto, M. J., Dominguez-Ferreras, A., Perez-Mendoza, D., Sanjuan, J. and Olivares, J.** (2009). Mutualism versus pathogenesis: the give-and-take in plant-bacteria interactions. *Cell Microbiol.* **11**, (3), 381-8.
- Soto, M. J., Sanjuan, J. and Olivares, J.** (2006). Rhizobia and plant-pathogenic bacteria: common infection weapons. *Microbiology.* **152**, (Pt 11), 3167-74.
- Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W. and Hess, W. R.** (2008). The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet.* **4**, (8), e1000173.
- Storz, G.** (2002). An expanding universe of noncoding RNAs. *Science.* **296**, (5571), 1260-3.
- Storz, G., Vogel, J. and Wassarman, K. M.** (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell.* **43**, (6), 880-91.
- Stuart, K., Kable, M. L., Allen, T. E. and Lawson, S.** (1998). Investigating the mechanism and machinery of RNA editing. *Methods.* **15**, (1), 3-14.

- Studier, F. W.** (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of molecular biology*. **219**, (1), 37-44.
- Sukhodolets, M. V. and Garges, S.** (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry*. **42**, (26), 8022-34.
- Sun, X., Zhulin, I. and Wartell, R. M.** (2002). Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.* **30**, (17), 3662-71.
- Tabei, Y., Tsuda, K., Kin, T. and Asai, K.** (2006). SCARNA: fast and accurate structural alignment of RNA sequences by matching fixed-length stem fragments. *Bioinformatics*. **22**, (14), 1723-9.
- Tafer, H. and Hofacker, I. L.** (2008). RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics*. **24**, (22), 2657-63.
- Taghbalout, A. and Rothfield, L.** (2007). RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America*. **104**, (5), 1667-72.
- Takada, A., Umitsuki, G., Nagai, K. and Wachi, M.** (2007). RNase E is required for induction of the glutamate-dependent acid resistance system in *Escherichia coli*. *Bioscience, biotechnology, and biochemistry*. **71**, (1), 158-64.
- Timmermans, J. and Van Melderen, L.** (2010). Post-transcriptional global regulation by CsrA in bacteria. *Cellular and molecular life sciences : CMLS*. **67**, (17), 2897-908.
- Timmers, A. C., Soupène, E., Auriac, M. C., de Billy, F., Vasse, J., Boistard, P. and Truchet, G.** (2000). Saprophytic intracellular rhizobia in alfalfa nodules. *Mol Plant Microbe Interact.* **13**, (11), 1204-13.
- Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S. and Storz, G.** (2006). Target prediction for small, noncoding RNAs in bacteria. *Nucleic acids research*. **34**, (9), 2791-802.
- Tjaden, B., Saxena, R. M., Stolyar, S., Haynor, D. R., Kolker, E. and Rosenow, C.** (2002). Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic acids research*. **30**, (17), 3732-8.
- Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., Barthelemy, M., Vergassola, M., Nahori, M. A., Soubigou, G., Regnault, B., Coppee, J. Y., Lecuit, M., Johansson, J. and Cossart, P.** (2009). The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*. **459**, (7249), 950-6.
- Toledo-Arana, A., Repoila, F. and Cossart, P.** (2007). Small noncoding RNAs controlling pathogenesis. *Current opinion in microbiology*. **10**, (2), 182-8.
- Tomizawa, J.** (1984). Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript. *Cell*. **38**, (3), 861-70.

- Trainer, M. A. and Charles, T. C.** (2006). The role of PHB metabolism in the symbiosis of rhizobia with legumes. *Applied microbiology and biotechnology*. **71**, (4), 377-86.
- Tramonti, A., De Canio, M. and De Biase, D.** (2008). GadX/GadW-dependent regulation of the *Escherichia coli* acid fitness island: transcriptional control at the gadY-gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol Microbiol*. **70**, (4), 965-82.
- Tsui, H. C., Leung, H. C. and Winkler, M. E.** (1994). Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Mol Microbiol*. **13**, (1), 35-49.
- Tu, K. C. and Bassler, B. L.** (2007). Multiple small RNAs act additively to integrate sensory information and control quorum sensing in *Vibrio harveyi*. *Genes & development*. **21**, (2), 221-33.
- Ulve, V. M., Sevin, E. W., Cheron, A. and Barloy-Hubler, F.** (2007). Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in *SinoRhizobium meliloti* strain 1021. *BMC Genomics*. **8**, 467.
- Updegrave, T. B., Correia, J. J., Galletto, R., Bujalowski, W. and Wartell, R. M.** (2010). *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim Biophys Acta*. **1799**, (8), 588-96.
- Urban, J. H. and Vogel, J.** (2007). Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic acids research*. **35**, (3), 1018-37.
- Urban, J. H. and Vogel, J.** (2008). Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol*. **6**, (3), e64.
- Urbanowski, M. L., Stauffer, L. T. and Stauffer, G. V.** (2000). The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol*. **37**, (4), 856-68.
- Valencia-Burton, M. and Broude, N. E.** (2007). Visualization of RNA using fluorescence complementation triggered by aptamer-protein interactions (RFAP) in live bacterial cells. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]*. **Chapter 17**, Unit 17 11.
- Valencia-Burton, M., McCullough, R. M., Cantor, C. R. and Broude, N. E.** (2007). RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature methods*. **4**, (5), 421-7.
- Valentin-Hansen, P., Eriksen, M. and Udesen, C.** (2004). The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol*. **51**, (6), 1525-33.
- Valoczi, A., Varallyay, E., Kauppinen, S., Burgyan, J. and Havelda, Z.** (2006). Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. *The Plant journal : for cell and molecular biology*. **47**, (1), 140-51.

- Valverde, C. and Haas, D.** (2008). Small RNAs controlled by two-component systems. *Advances in experimental medicine and biology*. **631**, 54-79.
- Valverde, C., Livny, J., Schluter, J. P., Reinkensmeier, J., Becker, A. and Parisi, G.** (2008). Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. *BMC Genomics*. **9**, 416.
- Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaitre, B., Alunni, B., Bourge, M., Kucho, K., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E. and Mergaert, P.** (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*. **327**, (5969), 1122-6.
- Vasse, J., de Billy, F., Camut, S. and Truchet, G.** (1990). Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol*. **172**, (8), 4295-306.
- Vecerek, B., Beich-Frandsen, M., Resch, A. and Bläsi, U.** (2010). Translational activation of *rpoS* mRNA by the non-coding RNA DsrA and Hfq does not require ribosome binding. *Nucleic Acids Res*. **38**, (4), 1284-93.
- Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R. and Blasi, U.** (2008). The C-terminal domain of *Escherichia coli* Hfq is required for regulation. *Nucleic Acids Res*. **36**, (1), 133-43.
- Ventre, I., Goodman, A. L., Vallet-Gely, I., Vasseur, P., Soscia, C., Molin, S., Bleves, S., Lazdunski, A., Lory, S. and Filloux, A.** (2006). Multiple sensors control reciprocal expression of *Pseudomonas aeruginosa* regulatory RNA and virulence genes. *Proceedings of the National Academy of Sciences of the United States of America*. **103**, (1), 171-6.
- Vioque, A., Arnez, J. and Altman, S.** (1988). Protein-RNA interactions in the RNase P holoenzyme from *Escherichia coli*. *Journal of molecular biology*. **202**, (4), 835-48.
- Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. and Gelfand, M. S.** (2002). Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res*. **30**, (14), 3141-51.
- Vockenhuber, M. P., Sharma, C. M., Statt, M. G., Schmidt, D., Xu, Z., Dietrich, S., Liesegang, H., Mathews, D. H. and Suess, B.** (2011). Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *RNA biology*. **8**, (3),
- Vogel, J.** (2009). A rough guide to the non-coding RNA world of *Salmonella*. *Mol Microbiol*. **71**, (1), 1-11.
- Vogel, J. and Luisi, B. F.** (2011). Hfq and its constellation of RNA. *Nature reviews. Microbiology*. **9**, (8), 578-89.
- Vogel, J. and Papenfort, K.** (2006). Small non-coding RNAs and the bacterial outer membrane. *Current opinion in microbiology*. **9**, (6), 605-11.

**Vogel, J. and Wagner, E. G.** (2007). Target identification of small noncoding RNAs in bacteria. *Current opinion in microbiology*. **10**, (3), 262-70.

**Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jager, J. G., Huttenhofer, A. and Wagner, E. G.** (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic acids research*. **31**, (22), 6435-43.

**Vytvytska, O., Moll, I., Kaberdin, V. R., von Gabain, A. and Bläsi, U.** (2000). Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes Dev*. **14**, (9), 1109-18.

**Wagner, E. G. and Simons, R. W.** (1994). Antisense RNA control in bacteria, phages, and plasmids. *Annual review of microbiology*. **48**, 713-42.

**Wagner, E. G., Altuvia, S. and Romby, P.** (2002). Antisense RNAs in bacteria and their genetic elements. *Advances in genetics*. **46**, 361-98.

**Wais, R. J., Wells, D. H. and Long, S. R.** (2002). Analysis of differences between *SinoRhizobium meliloti*1021 and 2011 strains using the host calcium spiking response. *Molecular plant-microbe interactions : MPMI*. **15**, (12), 1245-52.

**Wang, C., Ding, C., Meraz, R. F. and Holbrook, S. R.** (2006). PSoL: a positive sample only learning algorithm for finding non-coding RNA genes. *Bioinformatics*. **22**, (21), 2590-6.

**Washietl, S., Hofacker, I. L. and Stadler, P. F.** (2005). Fast and reliable prediction of noncoding RNAs. *Proceedings of the National Academy of Sciences of the United States of America*. **102**, (7), 2454-9.

**Wassarman, K. M. and Storz, G.** (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell*. **101**, (6), 613-23.

**Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. and Gottesman, S.** (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev*. **15**, (13), 1637-51.

**Wassarman, K. M., Zhang, A. and Storz, G.** (1999). Small RNAs in *Escherichia coli*. *Trends in microbiology*. **7**, (1), 37-45.

**Will, S., Reiche, K., Hofacker, I. L., Stadler, P. F. and Backofen, R.** (2007). Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol*. **3**, (4), e65.

**Willems, A., Fernandez-Lopez, M., Munoz-Adelantado, E., Goris, J., De Vos, P., Martinez-Romero, E., Toro, N. and Gillis, M.** (2003). Description of new Ensifer strains from nodules and proposal to transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an opinion. *International journal of systematic and evolutionary microbiology*. **53**, (Pt 4), 1207-17.

- Willkomm, D. K., Minnerup, J., Hüttenhofer, A. and Hartmann, R. K.** (2005). Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res.* **33**, (6), 1949-60.
- Wilms, I., Voss, B., Hess, W. R., Leichert, L. I. and Narberhaus, F.** (2011). Small RNA-mediated control of the *Agrobacterium tumefaciens* GABA binding protein. *Mol Microbiol.*
- Wilusz, C. J. and Wilusz, J.** (2005). Eukaryotic Lsm proteins: lessons from bacteria. *Nature structural & molecular biology.* **12**, (12), 1031-6.
- Worhunsky, D. J., Godek, K., Litsch, S. and Schlax, P. J.** (2003). Interactions of the non-coding RNA DsrA and RpoS mRNA with the 30 S ribosomal subunit. *J Biol Chem.* **278**, (18), 15815-24.
- Yang, S., Pelletier, D. A., Lu, T. Y. and Brown, S. D.** (2010). The *Zymomonas mobilis* regulator *hfq* contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. *BMC Microbiol.* **10**, 135.
- Yang, Z., Zhu, Q., Luo, K. and Zhou, Q.** (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature.* **414**, (6861), 317-22.
- Yoder-Himes, D. R., Chain, P. S., Zhu, Y., Wurtzel, O., Rubin, E. M., Tiedje, J. M. and Sorek, R.** (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America.* **106**, (10), 3976-81.
- Young, J. M.** (2003). The genus name *Ensifer* (Casida 1982) takes priority over *Sinorhizobium* (Chen et al. 1988), and *Sinorhizobium morelense* (Wang et al. 2002) is a later synonym of *Ensifer adhaerens* (Casida 1982). Is the combination "*Sinorhizobium adhaerens*" (Casida 1982, Willems et al. 2003) legitimate? Request for an Opinion. *International journal of systematic and evolutionary microbiology.* **53**, (Pt 6), 2107-10.
- Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G. and Gottesman, S.** (2003). Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol.* **50**, (4), 1111-24.
- Zhang, Y. and Hong, G.** (2009). Post-transcriptional regulation of NifA expression by Hfq and RNase E complex in *Rhizobium leguminosarum* bv. *viciae*. *Acta Biochim Biophys Sin (Shanghai).* **41**, (9), 719-30.



