A survey of sRNA families in α-proteobacteria

Coral del Val, 1, 2,* Roscío Romero-Zaliz, 2 Omar Torres-Quesada, 2 Alexandra Peregrina, 3 Nicolás Torre 1 and José I. Jiménez-Zurdo 1

1Grupo de Ecología Genética de la Rizosfera; Estación Experimental del Zaidín; Consejo Superior de Investigaciones Científicas; CSIC; Granada, Spain
2Department of Computer Science and Artificial Intelligence; E.T.S.I. Informatics; Universidad de Granada; Granada, Spain; 3CITIC-UGR, Centro de Investigación en Tecnologías de la Información y de las Comunicaciones de la Universidad de Granada; Granada, Spain

Keywords: Sinorhizobium meliloti, symbiotic nitrogen fixation, RNome, riboregulation, Hfq, Brucella, rhizobia, subB, speF

Abbreviations: sRNA, bacterial small non-coding RNA; RNome, ribonome; RACE, rapid amplification of cDNA ends; Smr, S. meliloti sRNA; TAP, tobacco acid pyrophosphatase; TSS, transcription start site; CM, covariance model; RBS, ribosome binding site; sSD, anti-Shine-Dalgarno; S-UTR, S untranslated region; GABA, γ-amino butyric acid

Submitted: 09/28/11
Revised: 11/03/11
Accepted: 11/03/11
http://dx.doi.org/10.4161/rna.18643
*Correspondence to Coral del Val; Email: delval@decsai.ugr.es

We have performed a computational comparative analysis of six small non-coding RNA (sRNA) families in α-proteobacteria. Members of these families were first identified in the intergenic regions of the nitrogen-fixing endosymbiont S. meliloti by a combined bioinformatics screen followed by experimental verification. Consensus secondary structures inferred from covariance models for each sRNA family evidenced in some cases conserved motifs putatively relevant to the function of trans-encoded base-pairing sRNAs, i.e., Hfq-binding signatures and exposed anti-Shine-Dalgarno sequences. Two particular family models, namely r15 and r35, shared own sub-structural modules with the Rfam model subB (RF00519) and the uncharacterized sRNA family r25b, respectively. A third sRNA family, termed r45, has homology to the cis-acting regulatory element speF (RF00518). However, new experimental data further confirmed that the S. meliloti r45 represents an Hfq-binding sRNA processed from or expressed independently of speF, thus refining the Rfam speF model annotation. All the six families have members in phylogenetically related plant-interacting bacteria and animal pathogens of the order of the Rhizobiales, some occurring with high levels of paralogy in individual genomes. In silico and experimental evidences predict differential regulation of paralogous sRNAs in S. meliloti. The distribution patterns of these sRNA families suggest major contributions of vertical inheritance and extensive ancestral duplication events to the evolution of sRNAs in plant-interacting bacteria. Post-genomic research has rendered bacterial small non-coding RNAs (sRNAs) as major players in the post-transcriptional regulation of gene expression underlying a wide range of important cellular processes, e.g., general responses to abiotic stimuli, cell division, quorum sensing or virulence. However, very little is known about the role of riboregulation in the control of symbiotic and pathogenic plant-microbe interactions. The α-subdivision of the proteobacteria includes Gram-negative microorganisms with different life styles, frequently involving long-term mutualistic or pathogenic interactions with higher eukaryotes. Sinorhizobium meliloti is an environmentally and agronomically relevant α-proteobacterium belonging to the order of the Rhizobiales. It is well recognized as a genetically tractable model microbe for the investigation and exploitation of the nitrogen-fixing endosymbiosis with legume plants. The outcome of these interactions is the formation in the cognate legume (i.e., Medicago species for S. meliloti) of the so-called root nodules which finally host invading bacteria in their differentiated nitrogen-fixing competent form of bacteria. The S. meliloti genome has a multipartite architecture consisting of a single chromosome (3.65 Mbp) and two large plasmids termed pSymA (1.35 Mbp) and pSymB (1.68 Mbp). Megaplasmid pSymB harbors the clusters of genes specifying symbiotic functions, among others, but is dispensable for bacterial free-living growth whereas pSymB exhibits chromosome-like features e.g., it accommodates the essential RNA-Agn encoding gene. This composite arrangement is common to the
genomes of many bacterial species of the order of the Rhizobiales in which second chromosomes have been proposed to evolve from an early-acquired ancestral plasmid. Similarly to S. meliloti, many β-proteobacteria interacting with plants usually host a variable number of accessory extrachromosomal replicons besides the ancestral set of primary chromosome and megaplasmid. These non-essential plasmids most likely have a mosaic origin and contribute to the adaptive flexibility demanded by the transition of bacteria from a free-living to an intracellular state. At the regulatory level, these adaptations from a free-living to an intracellular state.

Results

Generation of Smr sRNA family models. The starting point of this study was the set of eight non-coding transcripts identified previously in our laboratory on the basis of structural conservation and experimental verification. These sRNAs were initially termed Smr7C, Smr9C, Smr14C, Smr15C, Smr16C, Smr22C, Smr35B and Smr45C for S. meliloti RNA, where the suffix indicates their respective positions in the output table of candidates along with the genomic location of each sRNA to defined positions in the S. meliloti genome. Their 3'-ends were assumed to map to the last residue of the consecutive stretches of Us of Rho-independent terminators predicted for most of the transcripts, except for Smr22C and Smr45C which 3'-ends have been inferred from published experimental data. Recent RNA-Seq based characterization of the small RNA fraction (50–350 nt) of the closely related strain S. meliloti 2011 mapped the full-length Smr transcripts in the S. meliloti 1021 genome essentially to the same positions reported earlier. The nucleotide sequences of the full-length Smr transcripts were first used to query the Rfam database v. 10.0 (www.sanger.ac.uk/Software/Rfam). This search revealed full homology of Smr22C to the well-characterized 6S RNA family and therefore, this sRNA was not further considered in this study. Of the remaining seven RNAs, Smr15C/16C and Smr45C exhibited partial structural homology to the suhB (RF00519) and speF (RF00518) RNA families respectively, whereas the remaining query transcripts did not match any Rfam entry. These seven sRNA sequences, likely representing previously unknown bacterial sRNA families, were next BLASTed with default parameters against all available bacterial genomes (1,615 sequences as of April 20, 2011; www.ncbi.nlm.nih.gov). The genomic regions exhibiting significant degree of homology to the query sequences (78–80% similarity) were collected to generate initial alignments for each RNA that were manually curated to construct an Infernal Model (covariance model; CM) for each sRNA. As expected from their primary nucleotide sequence similarity, this analysis merged the tandemly-encoded Smr15C and Smr16C transcripts into the same RNA family and they were renamed accordingly as Smr15C1 and Smr15C2, respectively. The six RNA families, resulting from this study, have homologies limited to species of the order of the Rhizobiales within the β-subgroup of proteobacteria. Consistent with the naming scheme of the query sRNAs, their family models have been referred to as sRNAs. The predicted Smr family pri

| Name         | Alternative names* | 5'-end** | 3'-end** | Length (nt) |
|--------------|--------------------|----------|----------|-------------|
| Smr7C        | Smr15C1/Smr15C2    | 201,079  | 201,428  | 150         |
| Smr9C        | Smr45C            | 1,398,425| 1,398,277| 149         |
| Smr14C       | Smr15C1           | 1,667,613| 1,667,491| 123         |
| Smr15C       | Smr15C1           | 1,698,731| 1,698,617| 115         |
| Smr16C       | Smr15C1           | 1,698,937| 1,698,817| 121         |
| Smr35B       | Smr35B            | 577,730  | 577,868  | 139         |
| Smr45C       | Smr45C            | 3,105,445| 3,105,298| 148         |

*Alternative reported names for the Smr transcripts. Coordinates according to the S. meliloti 1021 genome database at http://sint.toulouse.inra.fr/bacteria/amr15/sgp1/sgp1rshome.cgi or www.rhizogate.de; **RACE-based mapping; ENREF_32; **Deep sequencing
Figure 1. For figure legend, see page 122.
five main hairpin loops generally interrupted by internal stem-loops and/or single stranded sequence stretches. These structures are supported by a variable degree of nucleotide covariance that was particularly high in the three stem-loops of the τ7, τ14 and τ15 family members and the 5’ domains of τ9 and τ35 families. In most cases, the 3’ domain consists of a GC-rich hairpin followed by tails of uridine residues, thus matching the main structural feature of the Rho-independent terminators of transcription.

The exception was τ145 which last hairpin is supported by a string conserved of the primary nucleotide sequences but does not resemble a bona fide Rho-independent terminator.

A remarkable and complex structural situation was found in the τ15 and τ35 families. Members of the τ15 family showed partial homology to the Rlam model (RF00519) known as sub-Hfq. In all cases this structural homology to the full-length sub-Hfq transcripts was restricted to the second hairpin and the Rho-independent terminator. Sub-Hfq-like genes have been computationally predicted to occur in multiple copies in a wide range of γ-proteobacterial genomes and some meta-genomes.14,15 Similarly, τ35 sRNAs have three well-defined hairpin loops. The second and third structural motifs are maintained by extensive primary nucleotide sequence conservation and define a sequence stretch with wider occurrence in the genomes of the Rhizobiaceae (40 sequences) outside the full-length τ35 sRNAs (not shown).

Therefore, sub-Hfq and this newly identified τ35 sub-structural domain (τ35b) likely represent widely distributed variants of the τ15 and τ35 sRNA families with a highly variable or even missing 5’ stem loops characteristic of the later transcripts. The τ35 sRNA families mostly include putative trans-encoded transcripts, which are expected to influence translation of target mRNAs through short base-pairing interactions that usually occlude the 3’ domain of the poly(A)-rich region of an oligo(U)-rich tail of an sRNA.15 This scaffold is expected to stabilize the sRNA molecule and could also bind mRNA sequences if they are released and exposed to the target with the aid of proteins. The RNA chaperone Hfq has been shown to fulfill this function in most of the sRNA-mRNA target interactions documented to date. Internal single-stranded A/U-rich regions as well as a free 3’-hydroxyl end of an oligo(U) stretch (e.g., of Rho-independent terminators) have been proposed as preferential sRNA interaction sites for Hfq.16 Both Hfq-binding signatures coexist in the τ9 and τ15 sRNAs, whereas exposed 3’-end poly(U) tails of different lengths are also evident in τ9 and τ15 sRNAs. However, the minimal uridines of the Rho-independent terminators predicted for τ7, τ14 and τ35 family members are mostly base-paired to upstream sequences and hence could not be easily available for Hfq binding. In good correlation with these observations, the S. meliloti Smr9C (τ9), Smr15C1, Smr15C2 (both τ15) and Smr45C (τ45) sRNAs have been detected in the sub-population of transcripts co-immunoprecipitated with a chromosomally-encoded epitope-tagged Hfq protein in lysates of free-living bacteria.17

Sme45C and speF are likely expressed as independent RNA elements in S. meliloti. The τ45 RNA family partially matched the Rlam model speF (RF00518), a family of cis-acting RNA elements likely involved in the regulation of polyamine biosynthesis that have been identified in several γ-proteobacterial species.17 Consistent with its proposed role, speF sRNAs are mostly leader sequences of orthologs of ornithine decarboxylase-encoding genes.13 The S. meliloti speF structural homolog has been predicted to map between positions 3,105,448 and 3,105,137 in the chromosome of the reference strain 1021, upstream the SMΔ02983 gene which encodes a putative ornithine/arginine decarboxylase (Fig. 2). Therefore, the 148 nt-long sRNA Smr45C, deduced from experimental mapping,11 would entirely match the 5’ region of speF (Fig. 2A).

To solve this apparent inconsistency in the annotation of S. meliloti speF, the transcriptional output of this genomic region was further investigated. A closer inspection of the SMΔ02983/SMΔ02984 IGR identified two short sRNA stretches that met the criterium (TTGACG...CTCATAT of τ7-dependent promoters in S. meliloti and other γ-proteobacteria).18 One of these transcription signatures (P1) had been previously identified as the putative promoter of Smr45C and is located immediately upstream the TSS determined from sRNA Smr45C. The second one (P2) overlaps the 3’ region of the Smr45C coding sequence (Fig. 2A).

Transcription initiation from the P2 promoter is predicted to occur at the T residue at 3,105,289 nt position in the S. meliloti genome (Fig. 2A). Confirming previously reported data, a probe complementary to the 3’ region of Smr45C detected a unique RNA species of the expected size accumulating differentially in free-living microorganisms but not expressed in endosymbiotic bacteria (Fig. 2B). In contrast, a 25-mer oligonucleotide probe targeting a sequence 16 nt downstream the Smr45C 3’-end hybridized to a major RNA molecule visible at the top of the gel with an expression profile very similar to that of Sme45C (Fig. 2B). The RNA species detected by this oligonucleotide probe most likely corresponds to the SMΔ02983 mRNA with a speF leader starting downstream the position previously predicted in silico. This RNA molecule could be originated either by processing of a larger undetectable and hence unstable RNA species transcribed from P1 or, most likely, by transcription...
from the newly identified promoter P2, independently of Smr45C in the biological conditions tested. In agreement with this observation, a S. meliloti map of TSS generated by RNA-Seq of total RNA revealed transcripts with 5'-ends at 3,105,292 and 3,105,166 nt positions in this region of the S. meliloti chromosome (A. Becker and J.P. Schlüter, personal communication). Altogether, these new experimental evidences further support classification of Smr45C as a Hfq-binding sRNA, likely unrelated to the speF RNA element.

Distribution of the 16r sRNA families in the Rhizobiales. The occurrence of the 16r sRNA families in sequenced bacterial species of the Rhizobiales was further assessed using the Infernal models (CMs) generated in this work. The results of this comparative analysis are summarized in Figure 3. With the only exception of Smr35B (nr35), which is encoded in the chromosome-like replicon pSymB, all our
query sRNA genes are chromosomally located in *S. meliloti* 1021. Overall, structure-based clustering of the homolog identified with each of the CMs essentially correlates with the phylogeny of the order (en.wikipedia.org/wiki/Small_non-coding_RNAs_in_the_endosymbiotic_genome_of_the_order (en.wikipedia.org/wiki/Small_non-coding_RNAs_in_the_endosymbiotic_genome_of_the_order (en.wikipedia.org/wiki/Small_non-coding_RNAs_in_the_endosymbiotic_genome_of_the_order (en.wikipedia.org/wiki/Small_non-coding_RNAs_in_the_endosymbiotic_genome_of_the_order). Thus, the current distribution pattern of the RNA appears variable across the Rhizobiales. CMs generated in this work along with the name of the query sRNA genes located in the ancestral chromosome of the Rhizobiales. In contrast, sRNA genes of the r15 and r14 families exist in highly variable copy numbers in the individual genomes; many of them located on extrachromosomal replicons i.e., large accessory plasmids in Rhizobiaceae/Phyllobacteriaceae representatives and the second chromosome in Brucella species. r15 members occur in two chromosomal copies in 19 genomes of bacteria belonging to the Rhizobiaceae and Brucellaceae families. These two genes are clustered in the same IGR in genomes from Rhizobiaceae whereas in Brucella species map to distant positions on chromosome 1. The second chromosomal r15 loci were missed by our search in the genomes of *B. melitensis* bv. abortus 2308.
R. meliloti bb 1 16M and Ochrobactrum anthropi ATCC49188. With the exceptions of A. tumefaciens CS8 and R. leguminosarum bv. trifoli 2304, at least a third \( \sigma r15 \) gene is located in extrachromosomal replicons of the host genomes. The \( \sigma r14 \) RNA family showed an even more complex distribution pattern in the Rhizobiaceae. Two tandem copies of the S. meliloti 1021 genome (formerly Smr14C3) and Smr14C3 homologous genes were also identified in Sinorhizobium and Mesorhizobium species whereas in O. anthropi ATCC49188, Agrobacterium and Brucella species the second chromosomal gene predicted by the \( \sigma r14 \) CM does not occur in such a syntenic context. A variable number of additional \( \sigma r14 \) copies (up to six more in the genome of R. leguminosarum bv. trifoli WSM1325) were identified in the main chromosomal and accessory plasmids of most of the bacterial species belonging to the Rhizobiaceae and Phyllobacteriaceae families. The \( \sigma r15 \) and \( \sigma r14 \) family members are mostly encoded in IGRs with a few exceptions of genes predicted within or amongst the annotated ORFs. However, these ORFs are frequently small, putatively coding for hypothetical proteins and/or absent from syntenic positions in bacterial genomes, thus representing probable mis-annotations as protein coding regions (en.wikipedia.org/wiki/%CE%B1r14_RNA#Promoter_Analysis; en.wikipedia.org/wiki/%CE%B1r15_RNA#Genomic_Context). In general, tandemly-arranged \( \sigma r15 \) and \( \sigma r14 \) genes occur in complete or partial microsynteny with the flanking genes in genomes of Rhizobiaceae and Phyllobacteriaceae as do their homologs on the main chromosome of O. anthropi ATCC49188 and Brucella species. However, microsynteny is much more fragmented or even absent for many of the remaining chromosomal and plasmidic copies of the \( \sigma r14 \) and \( \sigma r15 \) loci. Altogether, these observations suggest that \( \sigma r14 \) and \( \sigma r15 \) constitute families of paralogous sRNA gene copies in the Rhizobiaceae probably emanated from duplication events of their respective ancestral chromosomal genes over evolutionary time scales. Nonetheless, horizontal transfer events could certainly contribute to the current distribution patterns of some \( \sigma r14 \) and \( \sigma r15 \) gene copies, particularly of those occurring without signs of microsynteny in the accessory plasmids of plant-interacting bacteria. Noteworthy, some of the \( \sigma r15 \) loci were flanked by insertion sequences or transposase-encoding genes, among other genetic elements involved in mobility events and genomic rearrangements (en.wikipedia.org/wiki/%CE%B1r15_RNA#Promoter_Analysis; en.wikipedia.org/wiki/%CE%B1r14_RNA#Genomic_Context).

Finally, the \( \sigma r15 \) family exhibits a more restricted and dispersed representation, not only at the species but also at the strain levels. Only seven candidates were identified by the \( \sigma r15 \) Infernal models in addition to the S. meliloti Smr15B sRNA. Three of these predicted Smr15B homologs are encoded on the chromosomes of A. tumefaciens CS8, O. anthropi ATCC49188, and R. leguminosarum bv. vicieae 3841, whereas the remaining four \( \sigma r15 \) genes are extrachromosomal and were identified on the R. etli CFPN42 plasmid pRl102 and R. leguminosarum bv. vicieae 3841 plasmid pRL11 and R. leguminosarum bv. trifoli 1325 plasmids pRL132502 and pRL132504. Again, the majority of the \( \sigma r15 \) genes appeared to be independent transcription units with recognizable promoters with the exceptions of the chromosomal and plasmidic loci of R. leguminosarum bv. vicieae 3841 and R. etli CFPN42, respectively, which putatively overlap to annotated ORFs of unpredicted function. S. meliloti 1021 and O. anthropi ATCC49188 \( \sigma r35 \) genes occur in complete microsynteny with the flanking genes whereas the genomic regions of the other six \( \sigma r35 \) representatives revealed partial or no conservation at all (en.wikipedia.org/wiki/%CE%B1r35_RNA#Genomic_Context). \( \sigma r14 \) and \( \sigma r15 \) representatives are differently regulated in S. meliloti. The \( \sigma r14 \) and \( \sigma r15 \) CMs also identified several related genes in the S. meliloti 1021 genome. A third copy of the Smr15C locus was found in the megaplasmid pSymA (Smr15A) and up to 5 additional copies of the query Smr14C2-encoding gene were also identified; two of them chromosomally located (Smr14C1 and Smr14C3), two in pSymA (Smr14A1 and Smr14A2) and the remaining one in pSymB (Smr14B) (Fig. 4). Similarly to the situation of Smr15C1/Smr15C2, genes arranged in tandem in the same S. meliloti 1021 IGR encode Smr14C2 and Smr14C3. All the newly predicted Smr14- and Smr15-like sRNAs in the S. meliloti genome are encoded in IGRs, with the exception of Smr14B, which is encoded antisense to the Smr15B RNA gene (Fig. 4).

Oligonucleotides specific to all the Smr14 and Smr15 loci used were to probe S. meliloti RNA obtained from log and stationary phase cultures in TY broth (Fig. 4). These experiments confirmed the growth-dependent expression of Smr14C2, Smr15C1 and Smr15C2 transcripts with preferential accumulation of Smr15C1 upon entry of bacteria into stationary phase (Fig. 4). Despite their sequence and structural similarity Smr15C1 and Smr15C2 displayed opposite expression profiles. Strikingly, this set of Northern hybridizations did not reveal signs of expression of any of the other five Smr14 genes whereas the Smr15A transcript was barely detected on gels (Fig. 4). Multiple nucleotide sequence alignments of the promoter regions of all the genes encoding \( \sigma r14 \) and \( \sigma r15 \) members in species of the Rhizobiaceae identified diverse conserved motifs that could contribute to the differential expression of these genes in specific biological conditions (en.wikipedia.org/wiki/%CE%B1r15_RNA#Promoter_Analysis; en.wikipedia.org/wiki/%CE%B1r14_RNA#Promoter_Analysis). Supporting this prediction, RNA-Seq of the S. meliloti sRNAs expressed in a number of stress conditions has rendered variable number of reads for the S. meliloti \( \sigma r14 \)- and \( \sigma r15 \)-like transcripts, possibly correlating with a diversity of accumulation profiles.7

Discussion

The repertoire of non-coding RNAs expressed by the legume endosymbiont S. meliloti is one of the best characterized among the case of \( \sigma \)-proteobacterial counterparts.4,9 However, current information about the function of these transcripts in bacteria is certainly scarce. The first set of sRNAs identified in the reference strain S. meliloti 1021 included eight transcripts with genomic boundaries experimentally determined by independent approaches.5,6
Here, we have performed a comprehensive computational comparative analysis of these eight sRNA sequences to identify conserved structural motifs putatively relevant to their function as well as to assess their conservation patterns in bacterial genomes. CMs derived from alignments of the Smr sRNA homologs first identified Smr22C as the \( S.\) meliloti ortholog of the ubiquitous 6S sRNA. This RNA constitutes an example of a well-characterized trans-acting protein-binding sRNA. 19 The remaining seven transcripts represent structural and functional novel prokaryotic sRNAs and were collected into six different Infernal models. These CMs were used to accurately identify new members of each family in available sequenced bacterial genomes. This search revealed conservation of the Smr sRNAs in bacterial species belonging to the order of the Rhizobiales within the \( \alpha \)-subgroup of proteobacteria and, hence these RNA families were accordingly termed \( \alpha r \). Such a distribution pattern, limited to phylogenetically related bacterial species, is a general feature of the Hfq-dependent base-pairing riboregulators. Indeed, the consensus secondary structures deduced from each family model evidenced Hfq-binding and exposed aSD signatures in \( \alpha r 15 \) and \( \alpha r 14 \) transcripts as recognizable functional motifs involved in the sRNA-target mRNA interaction. In this regard, it is also noteworthy that previously reported pull-down experiments as well as stability assays on a \( S.\) meliloti \( hfq \) mutant background independently confirmed the Smr-Hfq interactions predicted by our CMs.17,20 Two particular CMs representing the \( \alpha r 15 \) and \( \alpha r 45 \) families rendered partial hits to the Rfam models corresponding to the suhB and speF non-coding RNA elements, respectively. The secondary structure of the \( \alpha r 15 \) sRNAs is predicted to consist of three hairpin motifs, in good agreement with the mapping of the

Figure 4. Northern analysis of the Smr14 and Smr15 sRNAs in \( S.\) meliloti. Maps of the genomic regions (not drawn to scale) of all the genes predicted by the \( \alpha r 14 \) and \( \alpha r 15 \) CMs in \( S.\) meliloti 1021 are shown to the left of the panels. Numbers denote coordinates of the genes in the genome. Name of the oligonucleotide probes used to hybridize each membrane are indicated to the right and their corresponding nucleotide sequences are listed in Table 2. RNA samples were obtained from logarithmic (log) and stationary phase (st) \( S.\) meliloti 1021 cultures in TY broth. 5S RNA was also probed as RNA loading control.

Table 2. RNA Biology Volume 9 Issue 2
A. tumefaciens Smr15C1 homolog (AbcR1) by enzymatic probing. Furthermore, the 5′SD-containing 5′ hairpin loop of *A. tumefaciens* AbcR1 has been shown to be the functional domain of this transcript for targeting the 5'-UTR of the mRNA encoding the GABA-binding protein. Conferring these experimental results preliminary predictions of Smr15C1/C2 sRNA interactions in *S. meliloti* using diverse bioinformatics tools anticipate a major involvement of the 5′ hairpin in target recognition (O. Torres-Quispeza and J.J. Jiménez-Zurdo, unpublished results). This 5′ stem loop is a variable or missing domain in shB-like transcripts. Our comparative analysis revealed a similar situation for the τ35 sRNA family and its variant τ35B. The disrupted occurrence of the τ35 loci in the Rhizobiales points also to the primary hairpins of these molecules as a functional domain, which probably has co-evolved with its target protein or mRNA in these genomes. Some 5′ located sRNA domains have been shown to be critical elements for specific pairing-hairpin mRNA target recognition that can act autonomously when fused to unrelated sRNA molecules. Therefore, the structural modules shared by τ35/τ35B and τ35/τ35B could be regarded as a kind of "proteobacteria-specific structural Legos" which could accommodate autonomous 5′ domains to create functionally diverse sRNAs.

We have also shown that the *S. meliloti* Smr45C mRNA and its downstream mRNA containing the co-regulatory element speF are detected as different RNA species across phylogenetically related species. However, single-copy genes hardly represent 5% of the total gene content of the *S. meliloti* genome. The genomes of plant-interacting bacteria usually evidence high levels of paralogy suggesting that their expansion through gene duplication has been little constrained during the evolution, facilitating the acquisition of new adaptive functions for life in the soil and within plant cells. The τ14 and τ15 family members occur in multiple copies as "the individual genomes". Multiple sRNA copies are not unique in bacteria although the physiological/ecological advantages of these reiterated have been only investigated in a subset of cases. Seemingly homologous sRNAs could act either redundantly, serving as backups in critical pathways, additively sensing different stimuli to integrate diverse environmental signals, independently, regulating different set of genes or hierarchically upon each other. In this work we have investigated the expression of free-living bacteria of the *Smr45* and *Smr15* genes identified by the respective covariance models in *S. meliloti* 1021. Northern experiments, promoter predictions and reported RNA-Seq data provide evidences for the differential regulation of these genes. In particular, the opposite expression patterns of Smr15C1 and Smr15C2 contrast with those of their *A. tumefaciens* homologs, which encoding genes are similarly arranged in tandem in the circular chromosome of this bacterium but showed identical expression profiles.

Interestingly, Smr15C1 retained its accumulation pattern in a *S. meliloti* Δsmr15C2 derivative and vice versa suggesting that these sRNAs act independently or additively rather than hierarchically as riboregulators in *S. meliloti* (O. Torres-Quispeza and J.J. Jiménez-Zurdo, unpublished). On the other hand, the undetectable expression of some transcripts in our assays, particularly of those grouped within the τ14 sRNA family suggests that they could be only expressed under not tested specific biological conditions to different adaptative functions in this bacterium.

In summary, our findings provide a baseline for the forthcoming investigation of the full-length sRNA in *S. meliloti* and related plant-interacting bacteria.

### Materials and Methods

Computational tools and methods. In a first step the *smr* gene sequences were BLASTed with default parameters against all currently available bacterial genomes (1,615 sequences at 20 April 2011; www.ncbi.nlm.nih.gov). The regions exhibiting signalScan homologies to the query sequence (78–89% similarity) were used to generate annotated Infernal alignment files for each family. This initial alignment was hand-curated and manually inspected to deduce a consensus secondary structure for each family. The consensus structure was also independently predicted with the program locARNATE in an automatic manner and differences reconciled giving priority to the structural conservation. Given the initial hand-curated structural alignment of close homologs Infernal was used to interrogate the same set of bacterial genomes, searching for new members of the models. The alignment process was repeated during three iterations. The candidates obtained with the Infernal models were selected as members of a given family if their Infernal E-value was ≤10^-5 or lower, or after manual inspection for those with higher Infernal E-values. The hierarchical cluster-tree for each family is derived by WPGMA clustering of the pairwise alignment distances and the optimal number of clusters was calculated from the tree using RNAclust (www.hennet.uni-leipzig.de/~kristin/Software/RNAclust/). A Stockholm format text file of each family alignment is provided in the links to the family wiki.
In order to study the microsinteny of each family, we located and extracted the flanking genes of their respective members. Non-annotated ORFs were further annotated using Blast2GO,34 35 or Smart 36 and the high-throughput pipeline ProWeep,37 and DomainSweep.33 The obtained coordinates of the sequence stretches complementary to each probe in the S. meliloti 1021 genome are listed in Table 2.

## References

1. Wang XH, Stee G. Regulatory RNA in bacteria. Cell 2009; 136:617-28; PMID:19239884; http://dx.doi.org/10.1016/j.cell.2009.01.063

2. Betz J, Andrews GC, O’Callaghan D. The evolution of short RNA elements in the γ-proteobacteria. Nat Rev Microbiol 2006; 4:235-43; PMID:16550893; http://dx.doi.org/10.1038/nrmicro1726

3. Jones KM, Kabara H, Dairi WW, Tajk AK, Walker GC. How shortlived sytems result plants the Sinorhizobium meliloti model. Nat Rev Microbiol 2007; 5:55-63; PMID:17052375; http://dx.doi.org/10.1038/nrmicro1705

4. Gahlin F, Franin TM, Long SR, Polish A, Milia P, Anpe F, et al. The compote genome of the legume symbiont Sinorhizobium meliloti. Science 2001; 293:64-72; PMID:11434066; http://dx.doi.org/10.1126/science.1064946

5. Hino SC, Goldbain BS, Grodd B, Seifull SC, Faraud SK, Nester EW. Genome sequence of the root nodule bacterium Rhizobium etli CEOC 8510. Nucleic Acids Res 2007; 35:9008-15; PMID:17830196; http://dx.doi.org/10.1093/nar/gkm1380

6. Glöckner FO, Rios E, Torres-Quesada O, Torres M, Jiménez-Zurdo J. Identification of differentially expressed small RNAs in the legume endosymbiotic Sinorhizobium meliloti by comparative genomics. Mol Microbiol 2007; 66:1080-91; PMID:17632573; http://dx.doi.org/10.1111/j.1365-2958.2007.05912a

7. Ueh VM, Scini EW, Chawla A, Brad-Miller F. Identification of chromosomal alpha-proteobacterial small RNAs by comparative genomic analysis and detection in Sinorhizobium meliloti strain 1021. BMC Genomics 2007; 8:467; PMID:17609952; http://dx.doi.org/10.1186/1471-2164-8-467

8. Veltri C, Loury J, Schlüter JP, Renkomchonok J, Becher A, Paris T. Prediction of small RNA genes and experimental detection in strain AGL2009. BMC Genomics 2009; 10:405; PMID:19783584; http://dx.doi.org/10.1186/1471-2164-10-405

9. Schlüter JP, Renkomchonok J, Pansecke S, Espagne-Hackenberg F, Jensen S, Jarizek J, et al. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium Sinorhizobium meliloti. BMC Genomics 2010; 11:245; PMID:20539413; http://dx.doi.org/10.1186/1471-2164-11-245

10. Zhou G, Ouyang F, Zhang X. Controlling alpha-sRNAs stability and translation with small, non-coding RNAs. Curr Opin Microbiol 2006; 9:44-53; PMID:16628690; http://dx.doi.org/10.1016/j.mib.2005.11.008

11. Gerdrup PP, Dush J, Tian J, Moore JL, Osaku M, Griffin-Lutes S, et al. Rfam is the ‘reference’ resource. Nucleic Acids Res 2011; 39:D141-5; PMID:21343988; http://dx.doi.org/10.1093/nar/gkq1042

12. Goude AR, Lorrain S, Bottrell SH, Newbold R, Holloway IE. The Virus RNA database. Nucleic Acids Res 2007; 35:W116-20; PMID:17444795; http://dx.doi.org/10.1093/nar/gkl081

13. Gauthier KA, Barrick JE, Jain J, Wile B, Tucker MJ, Pinkert S, et al. Evidence for a second class of Sinorhizobium meliloti chromosomal and other regulatory RNA motifs in alpha-proteobacteria. Genome Biol 2005; 6:R70; PMID:16054852; http://dx.doi.org/10.1186/gb-2005-6-7-r70

14. Schuchardt MH, Painter RJ, Miller T, Václavíková-Hartman J, Brinton RG. Structure of the photoreceptor transition helix and an ssRNA complex, a bacterial like the protein. EMBO J 2003; 22: 3546-56; PMID:12805757; http://dx.doi.org/10.1093/emboj/cdf322

15. Otaka H, Ishikawa M, Tani T, Akai H. Bacterial small RNAs by comparative genome analysis and detection in strain AGL2009. BMC Genomics 2010; 11:245; PMID:20539413; http://dx.doi.org/10.1186/gb-2009-10-8-r70

16. Sauer E, Weichenrieder O. Structural basis for RNA chaperone Hfq influences central carbon metabolism. Genome Res 2008; 18:950-5; PMID:18338188; http://dx.doi.org/10.1101/gr.079258.107

17. Sauer E, Weichenrieder O. Structural basis for RNA chaperone Hfq influences central carbon metabolism. EMBO J 2010; 29:3441-51; PMID:20602018; http://dx.doi.org/10.1038/emboj.c2010-108

18. Sauer E, Weichenrieder O. Structural basis for RNA chaperone Hfq influences central carbon metabolism. EMBO J 2010; 29:3441-51; PMID:20602018; http://dx.doi.org/10.1038/emboj.c2010-108

19. Tarrach O, Pennington K, O’Rourke P, John S, Hellen J, Roel F, et al. The Sinorhizobium meliloti ssRNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. BMC Microbiol 2010; 10:180; PMID:20529931; http://dx.doi.org/10.1186/1471-2180-10-17
18. MacLellan SR, MacLean AM, Finan TM. Promoter prediction in the rhizobia. Microbiology 2006; 152: 1751-63; PMID:16735738; http://dx.doi.org/10.1099/mic.0.28743-0
19. Wassarman KM. 6S RNA: a regulator of transcription. Mol Microbiol 2007; 65: 145-51; PMID:17714443; http://dx.doi.org/10.1111/j.1365-2958.2007.05894.x
20. Voss B, Holscher M, Baumgarth B, Kalbfleisch A, Kaya C, Hess WR, et al. Expression of small RNAs in Rhizobiales and protection of a small RNA and its degradation products by Hfq in flavobacteria outline. Biochem. Biophys. Res. Commun. 2009; 383:511-6; PMID:19084663; http://dx.doi.org/10.1016/j.bbrc.2009.01.125
21. Wilcsy C, Von R, Hess WR, Leichert LI, Niehrman F. Small RNA-mediated control of the Agrobacterium movement. GABA binding protein. Mol Microbiol. 2013; 89:455-66; PMID:23320835; http://dx.doi.org/10.1111/mmi.12459
22. Finkelstein R, Bevan M, Mike S, Harms CM, Vogel J. Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. Proc Natl Acad Sci USA 2010; 107:20365-69; PMID:21059983; http://dx.doi.org/10.1073/pnas.1009784107
23. Hunziker A, Tuboly C, Horvath P, Krishna S, Semsey S. Genetic flexibility of regulatory networks. Proc Natl Acad Sci USA 2010; 107:12998-3003; PMID:20615961; http://dx.doi.org/10.1073/pnas.0915003107
24. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jäger JG, Hüttenhofer A, et al. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res 2003; 31:6435-43; PMID:14062981; http://dx.doi.org/10.1093/nar/gkg867
25. Liu X, Damsgaard O, Grøntvedt J, Vaahtokari K, Tirmenstein T, Meftahi P, et al. A transactivating (thrombin) control element of the insulin regulatory PCR in a minigene promoter. Cell 2008, 137:795-90; PMID:18591405; http://dx.doi.org/10.1016/j.cell.2008.08.046
26. Luys DE, Meck KC, Lilly BN, Rathbun RV, Wangen NS, Rasker BL. The small RNA shavenbrog HP and multiple small RNAs control quorum sensing in Vibrio fischeri and Vibrio cholerae. Cell 2004; 116: 69-92; PMID:15242645; http://dx.doi.org/10.1016/j.cell.2004.09.125
27. Wu KC, Rasker BL. Multiple small RNAs act additively to integrate sensory information and control quorum sensing in Vibrio harveyi. Genes Dev 2007; 21: 331-6; PMID:17234887; http://dx.doi.org/10.1101/gad.147297
28. Urban JH, Vogel J. Two seemingly homologous non-coding RNAs act coordinately to activate gluT mRNA translation. PLoS Biol 2008; 6:e96; PMID:18351803; http://dx.doi.org/10.1371/journal.pbio.0060096
29. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. Bioinformatics 2009; 25:133-4; PMID:19076742; http://dx.doi.org/10.1093/bioinformatics/btp057
30. Will S, Ribeiro K, Hofacker I, Stadler PF, Badouch K. Selecting Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering. PLoS Comput. Biol. 2007; 3:e55; PMID:17527920; http://dx.doi.org/10.1371/journal.pcbi.0030055
31. Weis J, Nussbaumer J, Ermolaeva MD, Bork P. The small RNA database: sequence, structure and biological function. Nucleic Acids Res 2005; 33: D133-5; PMID:15364770; http://dx.doi.org/10.1093/nar/gki060