Molecular Phylogenetics and Evolution 99 (2016) 182–193

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic distribution and evolutionary pattern of an α-proteobacterial small RNA gene that controls polyhydroxybutyrate accumulation in *Sinorhizobium meliloti*

Antonio Lagares Jr., Indra Roux, Claudio Valverde *

Laboratorio de Bioquímica, Microbiología e Interacciones Biológicas en el Suelo, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Sáenz Peña 352, Bernal B1876BXD, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 13 January 2016
Revised 9 March 2016
Accepted 21 March 2016
Available online 24 March 2016

Keywords:
Non-coding small RNA
mmgR
α-proteobacteria
Sinorhizobium meliloti
Polyhydroxybutyrate
RNA family

ABSTRACT

It has become clear that sRNAs play relevant regulatory functions in bacteria. However, a comprehensive understanding of their biological roles considering evolutionary aspects has not been achieved for most of them. Thus, we have characterized the evolutionary and phylogenetic aspects of the *Sinorhizobium meliloti mmgR* gene encoding the small RNA MmgR, which has been recently reported to be involved in the regulation of polyhydroxybutyrate accumulation in this bacterium. We constructed a covariance model from a multiple sequence and structure alignment of *mmgR* close homologs that allowed us to extend the search and to detect further remote homologs of the sRNA gene. From our results, *mmgR* seemed to evolve from a common ancestor of the α-proteobacteria that diverged from the order of Rickettsiales. We have found *mmgR* homologs in most current species of α-proteobacteria, with a few exceptions in which genomic reduction events or gene rearrangements seem to explain its absence. Furthermore, a strong microsyntenic relationship was found between a large set of *mmgR* homologs and homologs of a gene encoding a putative N-formyl glutamate amidohydrolase (NFGAH) that allowed us to trace back the evolutionary path of this group of *mmgR* orthologs. Among them, structure and sequence traits have been completely conserved throughout evolution, namely a Rho-independent terminator and a 10-mer (5′-UUUCCUCCCU-3′) that is predicted to remain in a single-stranded region of the sRNA. We thus propose the definition of the new family of α-proteobacterial sRNAs αr8, as well as the subfamily αr8s1 which encompass *S. meliloti mmgR* orthologs physically linked with the downstream open reading frame encoding a putative NFGAH. So far, *mmgR* is the *trans*-encoded small RNA with the widest phylogenetic distribution of well recognized orthologs among α-proteobacteria. Expression of the expected MmgR transcript in rhizobiales other than *S. meliloti* (*Sinorhizobium fredii*, *Rhizobium leguminosarum* and *Rhizobium etli*) was confirmed by Northern blot. These findings will contribute to the understanding of the biological role(s) of *mmgR* in the α-proteobacteria.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Regulation of gene expression occurs at multiple levels along the flow of genetic information, but post-transcriptional control of mRNA translation and/or stability is a key step subject to the influence of small non-coding regulatory RNA molecules, which are collectively regarded as sRNAs (Wagner and Romby, 2015). Typically, sRNAs are transcripts having 50–250 nt of average length, and are encoded by single transcriptional units in between open reading frames. When their regulatory targets are mRNAs originated elsewhere in the replicons, these regulatory sRNAs are classified as *trans*-encoded sRNAs. Most of the characterized sRNAs act by antisense base-pairing with their target mRNAs along stretches of partial complementarity of at least 6–7 base pairs. A few other sRNA types act as molecular mimics that titer out RNA binding proteins from the target mRNAs (Marzi and Romby, 2012). Whatever their molecular mechanism, sRNA transcripts have been detected experimentally in multiple prokaryotic taxons, often in number of hundreds of different sRNA genes that are expressed at a certain physiological condition. Thus, sRNA-based control of gene expression seems to be a deeply conserved trait in the prokaryote domain (Wagner and Romby, 2015).

As for other types of cellular RNA molecules, the sRNA function is tightly related to their sequence and structural features, a fact that profoundly impacts in the evolutionary faith of different sRNA genes

* Corresponding author.
E-mail addresses: cvalver@unq.edu.ar, valverdecl@hotmail.com (C. Valverde).

http://dx.doi.org/10.1016/j.ympev.2016.03.026
1055-7903/© 2016 Elsevier Inc. All rights reserved.

(Updegrove et al., 2015). However, in contrast to the tight linkage existing between protein function and protein sequence across widely divergent taxa, sRNA function may be conserved in different taxonomic groups without significant sequence conservation (Updegrove et al., 2015). This naturally hampers identification of sRNA homolog genes based on simple sequence homology searches. Thus, in a broad sense, RNA families are rather defined on the base of the methods used to search for homolog genes and the corresponding criteria to infer homology, which are largely dependent on sequence/structural considerations (Reinkensmeier et al., 2011).

The α-proteobacteria are a highly diverse class of microorganisms with many important biological roles, and that can adopt as contrasting lifestyles as plant intracellular symbionts or animal intracellular pathogens (Williams et al., 2007). Consequently, different α-proteobacterial lineages have suffered diverse paths of independent genome evolution like reduction, expansion or even fragmentation into several replicons (Williams et al., 2007). In association with their diverse niche and lifestyle, α-proteobacteria present a wide range of metabolic capabilities, cellular morphologies and developmental programs, controlled by multiple regulatory systems (Williams et al., 2007). In this regard, different computational and experimental approaches have identified hundreds of RNA transcripts with possible regulatory functions in several α-proteobacterial species (Landt et al., 2008; Madhugiri et al., 2012; Schluter et al., 2010; Valverde et al., 2008). This is the case for *Sinorhizobium meliloti*, the N_2-fixing symbiont of *Medicago* spp. plant roots, for which the non-coding RNome has been thoroughly characterized recently (Schluter et al., 2013, 2010; Torres-Quesada et al., 2014). *S. meliloti* expresses over 400 *trans*-encoded RNAs, and comparative genomic analysis allowed definition of dozens of α-proteobacterial RNA families with a wide range of phylogenetic distribution (del Val et al., 2012; Reinkensmeier and Giegerich, 2015; Reinkensmeier et al., 2011). Nonetheless, only a few sRNAs have been characterized in terms of their biological function (Baumgardt et al., 2015; Robledo et al., 2015; Torres-Quesada et al., 2013).

We have recently found that the transcript encoded by the *mmgR* gene negatively controls accumulation of intracellular storage granules in *S. meliloti* strain 2011 under conditions of surplus carbon source (Lagares (Jr) et al., 2016). The first bibliographic report related with *mmgR* was the sequence/structural RNA motif *suhB* found in a diverse group of α-proteobacterial species (Corbino et al., 2005). Corbino and colleagues proposed that the motif would encode a putative regulatory RNA molecule with conserved sequence and structure elements of possibly biological relevance. That report gave rise to the RFAM entry *suhB* (RF00519), which has been fed with a significant number of new sequences derived from accelerated genome sequencing (Burge et al., 2013). Several subsequent studies provided evidence that the *suhB* element was indeed an active gene expressed as a non-coding small transcript in *S. meliloti* and other α-proteobacteria (Landt et al., 2008; Madhugiri et al., 2012; Schluter et al., 2010; Valverde et al., 2008). The *suhB* transcript of *S. meliloti* was first designated as the Sm8 RNA upon detection by Northern Blot (Valverde et al., 2008) and later referred to as SmelC689 RNA following the nomenclature introduced in an RNAseq study (Schluter et al., 2010). Upon characterization of its mutant phenotype, the *S. meliloti suhB* element = *sm8* = *smelC689*, has been recently renamed as *mmgR* because it encodes an RNA regulator whose mutation results in a strain that "makes more granules"; i.e., the cytoplasm of the mutant strain lacking MmgR RNA contains a higher content of polyhydroxybutyrate (PHB) storage granules than the wild type strain (Lagares (Jr) et al., 2016).

The information available about phylogeny and evolution of functionally characterized bacterial sRNAs is certainly scarce (Peer and Margalit, 2014; Skippington and Ragan, 2012), particularly outside the γ-proteobacteria (Chen et al., 2014). In α-proteobacteria, three recent studies have catalogued and defined multiple sRNA families in rhizobiales on the basis of their phylogenetic distribution and evolution (del Val et al., 2012; Reinkensmeier and Giegerich, 2015; Reinkensmeier et al., 2011), but none of them has considered *mmgR*. Here, we have carried out a detailed analysis of the evolutionary and phylogenetic aspects of *mmgR*, based on computational tools considering sequence, structural and microsynteny features, and have confirmed expression of the sRNA in rhizobial species other than *S. meliloti*. Based on the results, we propose the definition of the new family of α-proteobacterial sRNAs αr8, as well as the subfamily αr8s1 encompassing *S. meliloti mmgR* homologs physically linked with a downstream open reading frame encoding a putative N-formyl glutamate amidohydrolase.

2. Materials and methods

2.1. Gene sequences used in this study

The reference sequence for the proposed sRNA family was that of the *mmgR* gene from *S. meliloti* strain 1021, whose 5′ and 3′ end coordinates were considered to be 3046712 and 3046789, respectively, based on the consensus of experimental and bioinformatics data (Schluter et al., 2013). The corresponding *mmgR* coordinates in the genome of the closely related strain 2011 are 3046287 and 3046363 (Fig. 1a). Additionally, we gathered the complete genome sequences from all eubacterial species that were available at the NCBI FTP site (ftp.ncbi.nlm.nih.gov) in October 2014. The corresponding 16S rDNA sequences of α-proteobacterial species were retrieved from the Silva server (Pruesse et al., 2007; Quast et al., 2013; Yilmaz et al., 2014).

2.2. Computational workflow for identification of mmgR homologs and construction of the RNA family

The complete workflow is summarized in Fig. 1b. An initial set of putative *mmgR* homologs was obtained by sequence homology search using BlastN (NCBI, http://blast.ncbi.nlm.nih.gov/), run with default parameters and $E < 0.005$ on the complete set of non-redundant nucleotide sequences available in GenBank. The retrieved sequences were subjected to multiple alignments with LocARNA using default parameters (Smith et al., 2010). LocARNA considers both primary and secondary structure for data treatment. The multiple alignments allowed identification of the strongly conserved 28-mer core 5′-CCCTCCTTGGGTGT TTCCTCCCTAGACT-3′ (nucleotide positions 3046733–3046760 of *S. meliloti* strain 1021 genome), which was subsequently used as the query to repeat the sequence homology search in BlastN with default parameters and $E < 0.005$. This second group of homolog sequences was inspected for the presence of upstream promoter and downstream terminator elements. Putative prokaryotic promoter elements were predicted on the sequence immediately upstream of the conserved core with the neural network tool BDGP (Reese, 2001), with a minimum promoter score of 0.55. *Rho*-independent terminator elements were predicted on the sequence immediately downstream of the conserved core with the ARNold server, using default parameters (Naville et al., 2011). Thus, all those homolog sequence elements bearing both predicted promoter and terminator elements surrounding the conserved core were considered *mmgR* homologs, and were subjected to a second round of multiple alignment and secondary structure information analysis using LocARNA (Smith et al., 2010). The structural alignment served to create a covariance model (CM) with the Infernal package (Nawrocki and Eddy, 2013). The covariance profile was

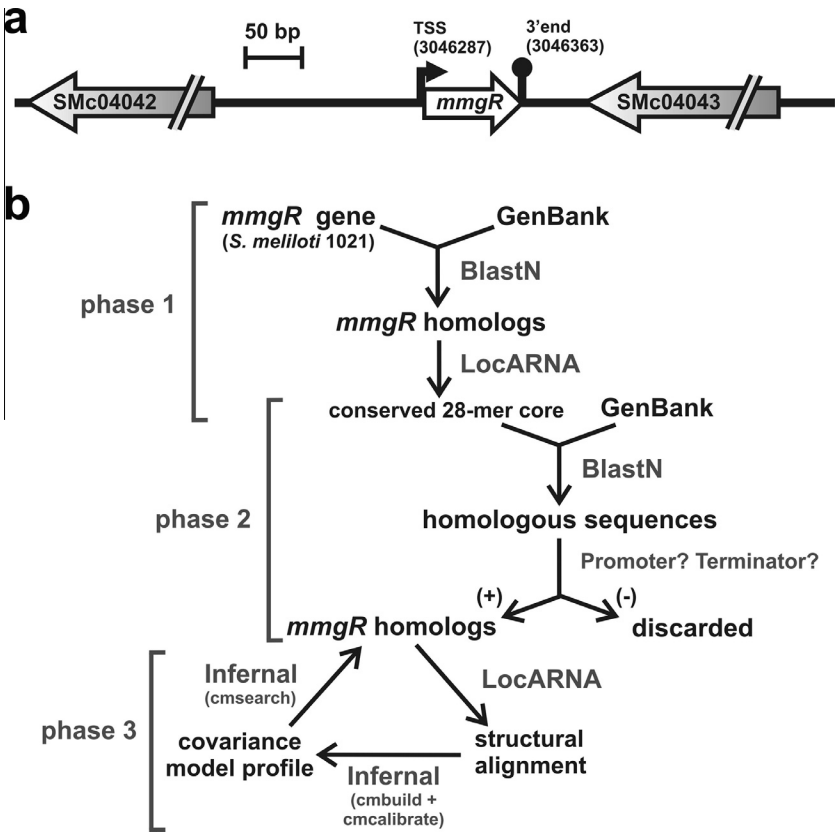


Fig. 1. (a) The *mmgR* locus of *S. meliloti* 2011 chromosome. The absolute chromosomal coordinates of the *mmgR* transcriptional start site (TSS) and 3′-end are indicated. The ORFs flanking *mmgR* are designated with their synonym gene names from strain 1021 (not drawn to scale). (b) Overview of the workflow performed for the construction of the covariance model for the αr8 sRNA family.

then utilized to perform a profile search on all eubacterial available genomes using BlastN. All hits showing $E < 10^{-4}$ were retained for the next round of analysis. The steps of model construction, search and homolog identification were repeated upon incorporation of the newly identified homologs. The process stopped when the number of new hits reached a plateau. Consensus secondary structure was deduced from the multiple sequence alignment with the RNAalifold program (Bernhart et al., 2008). Microsynteny was studied through the online servers Microbial Genomic Context Viewer (Overmars et al., 2013) and SyntTax (Oberto, 2013), which served to determine the identity of the open reading frames flanking each *mmgR* homolog in its respective replicons.

2.3. Phylogenetic analysis of α-proteobacterial species

Phylogenetic analysis was carried out with the MEGA v6 package (Tamura et al., 2013). The α-proteobacterial 16S rDNA sequence alignment generated with ClustalW2 (McWilliam et al., 2013) was used to infer the phylogenetic tree by the Maximum Likelihood method. We used the Tamura–Nei nucleotide substitution model and allowed for invariant positions in the alignment. Confidence analyses were undertaken using 500 bootstrap replicates. The phylogenetic relationships were summarized as a consensus cladogram only with nodes having over 50% of representativeness in the bootstrap analysis.

2.4. Expression analysis of sRNA transcripts in rhizobial strains

The bacterial strains with available sequenced genomes that were used in this study are: *S. meliloti* 2011 (Sallet et al., 2013), *Sinorhizobium fredii* NGR234 (Schmeisser et al., 2009), *Rhizobium leguminosarum* bv. *viciae* 3841 (Young et al., 2006), and *Rhizobium etli* CIAT652 (Gonzalez et al., 2010). The strains were grown aerobically in rhizobial defined medium (Vincent, 1970). RNA preparations and Northern blotting were done essentially as described previously (Valverde et al., 2008). Total RNA was extracted from cells collected in stationary phase. Cells were resuspended in Trizol (Invitrogen, Life Technologies), disrupted in a mini-bead beater during 30 s (BioSpec, USA), and extracted with 0.2 volumes of chloroform. The RNA present in the aqueous phase was precipitated with 1 volume of isopropanol, washed with 70% v/v ethanol, air-dried and resuspended in 20 μl of DEPC-treated water. RNA quality was assessed by denaturing agarose electrophoresis and RNA concentration was estimated by UV spectrophotometry (Sambrook et al., 1989). Northern blot hybridization was carried out after transfer of urea-PAGE fractionated RNA (2 μg per sample) onto nylon membranes, using a PCR generated DIG-labeled DNA fragment covering the entire *S. meliloti* MmgR RNA gene (Valverde et al., 2008). The probe hybridization temperature was 40 °C, in order to facilitate annealing with the MmgR transcript of rhizobial species having slightly divergent sequences (*R. leguminosarum* and *R. etli*). Membranes were stripped by incubating them twice with a boiling 0.1% SDS solution for 30 min. The pre-hybridization, hybridization (at 50 °C), and developing steps were repeated using a dsDNA probe targeting the *S. meliloti* 5S rDNA, in order to provide an indication of total RNA load.

3. Results

3.1. Search of mmgR homolog sequences in eubacterial genomes

An initial search for sequence homology allows the identification of a set of defined homolog genes with high specificity, which in turn facilitates the search for remote homologs by using

complementary and more sensitive strategies (Fig. 1). In the first phase of the workflow (Fig. 1b), we queried the NCBI non redundant nucleotide sequence database using BlastN with the *S. meliloti mmgR* gene sequence (Fig. 1a), and retrieved a set of 33 sequence hits encompassing 16 species from the *Rhizobiaceae* family (among 32 sequenced genomes), with a high statistical confidence ($E < 0.001$) supporting the homology with the *mmgR* sRNA gene. A single gene copy was detected in 31 out of 32 genomes. Interestingly, all homolog genes showed microsynteny, in contrast with what has been formerly anticipated for the related sequential/structural motif *suhB* (Corbino et al., 2005). The predicted consensus secondary structure for this group of *mmgR* homologs showed 2 conserved stem-loops and a third conserved and typical *Rho*-independent terminator stem-loop (Fig. 2b). The structural alignment of this first set of *mmgR* homologs from *Rhizobiaceae* resulted in the identification of an internal 28-mer stretch that is fully conserved (Fig. 2a; alignment nucleotide positions 22–49).

Based on the higher relative variability displayed by the terminator region in comparison to the central core (Fig. 2a), which may reduce the sensitivity of the sequence homology search with the BlastN algorithm, we repeated the search of *mmgR* homolog genes by restricting the query sequence to that of the conserved 28-mer (Fig. 1b; phase 2). In order to reduce the number of false positive hits that a shorter query sequence may generate, we applied a stringent filtering by selecting only those BlastN sequence hits located in between open reading frames and that contained both a predicted promoter and a predicted *Rho*-independent terminator flanking the conserved core such that they define a transcriptional unit coherent with a small non-coding transcript. This procedure notably expanded the number of *mmgR* homolog genes to 76 sequences among 67 sequenced genomes, distributed throughout 43 species within the order Rhizobiales (Fig. 3). Upon multiple alignment of this new set of sequences, it was evident that the evolutionary distance among homolog genes was associated with a higher variability in the sequence of the first stem-loop, although preserving its secondary structure. Provided that secondary structure conservation has functional relevance in RNA-based regulatory molecules (Updegrove et al., 2015), we hypothesized that it would be possible to identify additional *mmgR* homologs with lower sequence conservation but sharing similar structural motifs by querying GenBank using a sequence profile based on a covariance model (CM) (del Val et al., 2012; Nawrocki and Eddy, 2013; Reinkensmeier et al., 2011) (Fig. 1b).

3.2. Expanded search of mmgR homologs with structural covariance models defines the novel αr8 RNA family

CM represents a statistical tool based on the generation and utilization of multiple alignments of RNA sequences with structural considerations that allows searching and identifying homolog RNA genes in nucleotide databases by considering the

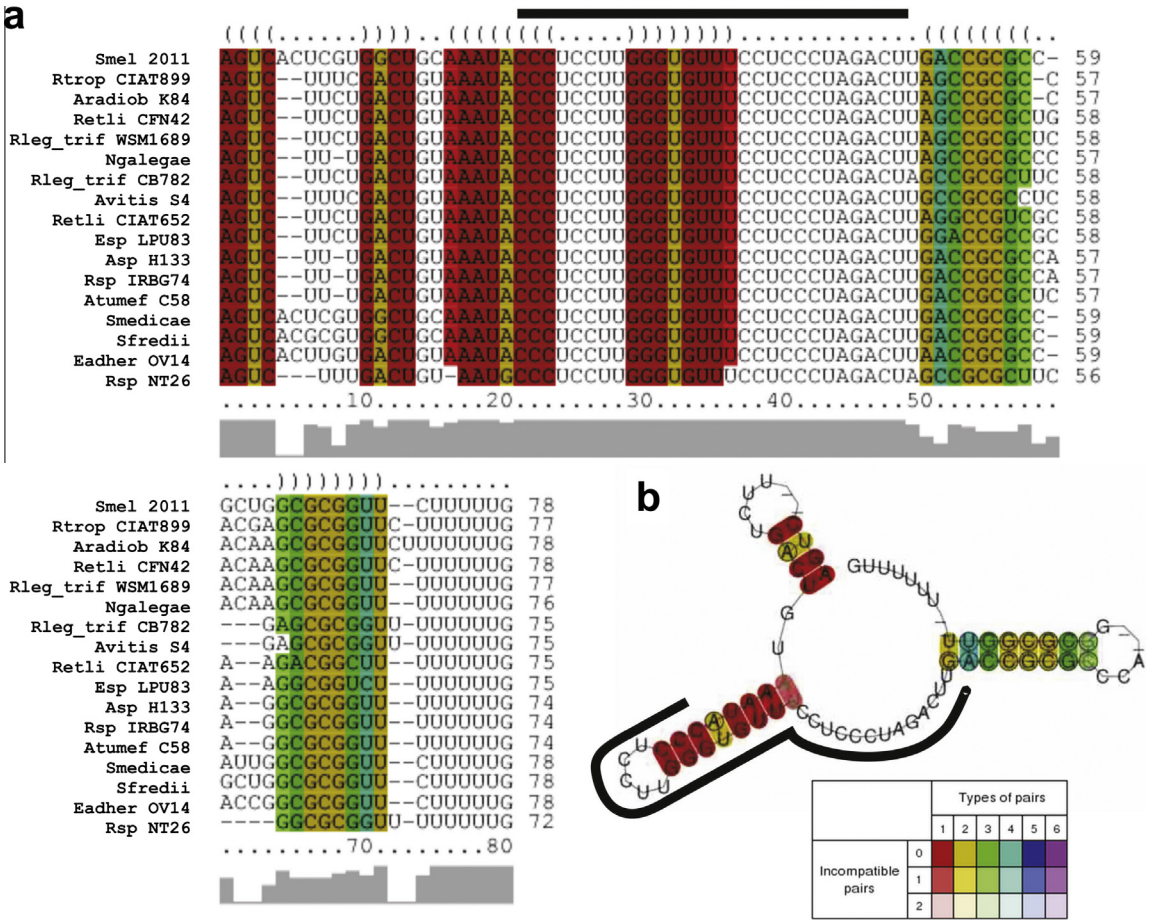


Fig. 2. (a) LocARNA based multiple sequence and structure alignment of BlastN hits for *mmgR* gene. Identical sequences from closely related strains of the same species were not included for simplification of the resulting alignment. (b) Consensus secondary structure predicted for *mmgR* close homologs after the LocARNA alignment shown in (a). The fully conserved sequence core is indicated with a black line. Compatible base pairs are colored, where the hue shows the number of different types C-G, G-C, A-U, U-A, G-U or U-G of compatible base pairs in the corresponding columns. Thus, the hue shows sequence conservation of the base pair. The saturation decreases with the number of incompatible base pairs, and indicates the structural conservation of the base pair (Smith et al., 2010).

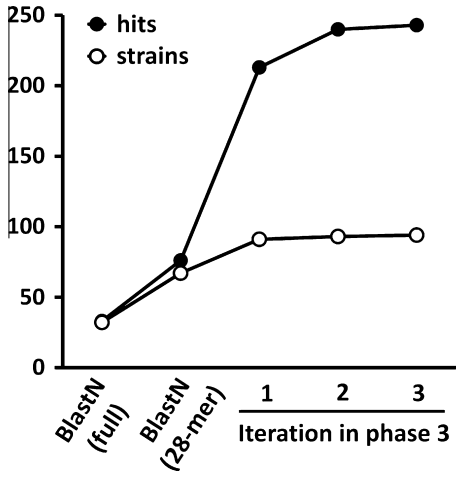


Fig. 3. Number of *mmgR* homologs and number of strains in which they were distributed that were detected throughout the workflow.

structural conservation factor (Nawrocki and Eddy, 2013). A profile containing sequence and structural information was then generated upon construction of a CM of the 76 homolog genes retrieved in phase 2, by means of the Infernal package (Nawrocki and Eddy, 2013). The resulting profile served as query to search for homolog sequences in all eubacterial genomes using BlastN (Fig. 1b, phase 3). The first run with the CM profile resulted in a total of 213 hits with significant similarity ($E < 10^{-4}$; Fig. 3), including all sequence homologs within the set of 76 sequences used to generate the CM profile. Thus, 137 novel homologs were identified, representing a significant near 2-fold increment in the number of *mmgR* homologs that were detected only by sequence homology (Fig. 3).

Next, we optimized the CM profile by incorporating all the new detected sequences in the structural alignment, and iterated the process aiming to collect novel and remotely related homolog *mmgR* sequences (Fig. 1b, phase 3). The iterative runs were stopped when the number of new sequences at the cycle i, was higher than the number of new sequences rendered by the cycle $(i - 1)$. Thus, for $i = 3$, we ended up with a total of 243 genes distributed in 70 species and 95 sequenced strains (Fig. 3). Notably, the CM profile-based search allowed the identification of *mmgR* sRNA homolog genes in 32 additional species within the class α-proteobacteria and belonging to the genera *Acidiphilium*, *Asticcaculis*, *Azospirillum*, *Beijerinckia*, *Bradyrhizobium*, *Brevundimonas*, *Caulobacter*, *Chelatovirans*, *Gluconacetobacter*, *Granulibacter*, *Hyphomonas*, *Magnetospirillum*, *Methylobacterium*, *Methylocystis*, *Ochrobactrum*, *Parvibaculum*, *Parvularcula*, *Phenylobacterium*, *Pseudovibrio*, *Rhodospirillum* and *Tistrella*. Within the Rhodobacterales clade, the search phase 3 allowed identification of an *mmgR* homolog only for the *Ruegeria* sp. genome (Fig. 4). Inspection of the corresponding genome loci in the rest of the Rhodobacterales genomes suggests the presence of loosely related sequences resembling *mmgR*, although homology could not be inferred with an adequate statistical support. Hence, *mmgR* genes are not only present within the Rhizobiales, but they are rather distributed among all α-proteobacteria, with the exception of the order Rickettsiales (Fig. 4).

The complete set of detected *mmgR* homolog genes found with the described workflow (Fig. 1b) establishes a novel RNA family that we designated αr8 (for α-proteobacterial RNA family 8). Fig. 4 shows the phylogenetic distribution of all members of this family over the 16S rDNA-based evolutionary tree of the α-proteobacterial class. Whereas most genomes contain a single *mmgR* gene copy, there are certain genera that contain multiple copies of *mmgR* homologs, such as *Methylobacterium* spp. (Fig. 4). Our analyses primarily suggest that the emergence of the *mmgR* sRNA gene within the α-proteobacteria took place in the common ancestor that immediately diverges from the order Rickettsiales. Certainly, that is the branch in which the sRNA gene has been preserved to reach the current wide taxonomic distribution (Fig. 4). Nevertheless, considering that species within the Rickettsiales have suffered genomic reduction due to their adaptation to the intracellular environment (Sallstrom and Andersson, 2005), it cannot be ruled out that *mmgR* may have originated earlier in the root of the alpha-proteobacterial branch and was subsequently lost in the Rickettsiales as a consequence of the release of the selection pressure to preserve the sRNA gene.

3.3. Microsynteny analysis and definition of the αr8s1 RNA subfamily

We carried out a detailed analysis of the flanking sequences of every *mmgR* homolog sequence retrieved by the iterative process of CM profile search (Fig. 1b, phase 3) and detected a strong physical linkage among a subgroup of *mmgR* homologs and homologs of the *S. meliloti SMc04043* gene (Fig. 1a, Supp. Fig. 2). *SMc04043* encodes a putative N-formyl glutamate amidohydrolase (NFGAH). We then seized the whole set of *mmgR* homolog sequences to retain all those showing the aforementioned microsynteny. In addition, considering that some homolog sRNA genes may have escaped the CM profile-based search due to a slightly higher sequence/structure divergence than that defined by the cut-off, we also inspected the neighbouring sequences of sequence hits arising from iteration 3 and having *E* values slightly higher than 10^{-4}. By doing so, we detected a novel group of candidate *mmgR* homolog genes in the genera *Novosphingobium*, *Erytrobacter*, *Sphingopyxis* and *Sphingomonas*, all belonging to the order Sphingomonadales. The genomic location of the putative sRNA gene adjacent to the NFGAH ORF confirms its common evolutionary origin and diversification along with the rest of the *mmgR* genes (Fig. 5).

Based on the above and on the recently introduced nomenclature for α-proteobacterial sRNAs (del Val et al., 2012), we propose the definition of the αr8s1 subfamily of sRNA genes, which comprise the set of *S. meliloti mmgR* orthologs that are linked by their microsynteny with the NFGAH gene (Fig. 5). Notably, within the evolutionary branch that bears *mmgR* homologs (Fig. 4), the sRNA gene conserves its microsynteny with *SMc04043* orthologs in 92% of the genomes with at least one *mmgR* copy (Fig. 5). We have not found *mmgR* genes within a rather limited number of genera, like *Maricaulis* spp. (order Caulobacterales), *Zymomonas mobilis* (the most recent evolutionary branch within the order Sphingomonadales) (Supp. Fig. 3), and *Brucella* spp. and *Bartonella* spp., two members of the Rhizobiales order that have diverged their evolutionary path from the closely related plant associated rhizobia and have become animal intracellular pathogens (Batut et al., 2004; del Val et al., 2012) (Fig. 4).

3.4. Sequence and structural properties of the αr8s1 subfamily

Fig. 6 shows the multiple sequence alignment of representative members from the αr8s1 subfamily. Identical sequences from closely related strains of the same species were not included for simplification of the resulting alignment. We detected gene duplication events in the genomes of *Tistrella*, *Beijerinckia*, *Hypomicrobium*, *Methylocella* and *Xhantobacter* species that derived in the presence of multiple copies of the *mmgR* gene in the same intergenic region, which blurred the possibility to recognize their nature as ortholog or paralog copies; these sequences were neither included in the sequence/structural alignment shown in Fig. 6.

The sRNA genes comprised within the here proposed αr8s1 subfamily, they all share conserved sequence and structural elements,

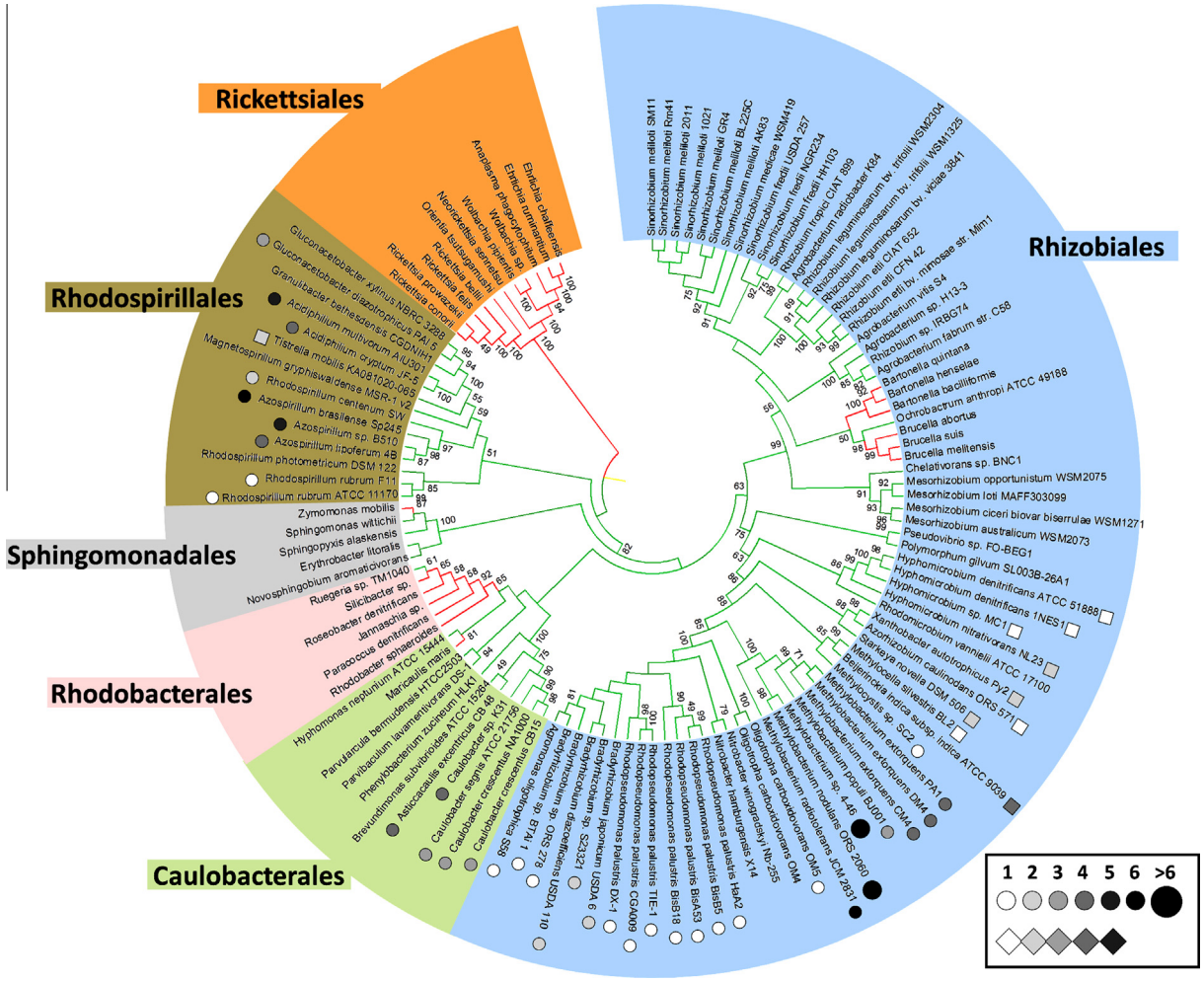


Fig. 4. Phylogenetic distribution of αr8 family members. Green or red lines indicate strains in which *mmgR* homologs were or were not found, respectively. The yellow lines indicate uncertainty about the presence of *mmgR* homologs in ancestral strains. Circles or squares located directly after a strain name indicate that multiple homologs were found in the corresponding genome, either in the same inter-ORF region or at different genomic loci, respectively. Greyscale shadowing of circles and squares indicate the numbers of additional homolog copies that were found, as denoted in the legend on the bottom of the figure. The consensus cladogram represents the phylogeny of the species inferred using the corresponding 16S rRNA gene sequences, as described in Materials and Methods. Tree structure correlates well with the tree obtained after the phylogenetic analysis of a concatenated alignment of more than 100 α-proteobacterial protein families (Williams et al., 2007). The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown at the nodes when values are higher than 50%. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

i.e. a *Rho*-independent transcriptional terminator and an internal and fully conserved core of 10 nucleotides encoding the RNA sequence motif 5′-UUUCCUCCCU-3′ (Fig. 6a; alignment nucleotide positions 49–58), which is additionally predicted to be located in a single-stranded region of the transcript (Fig. 6, panels b–e). In contrast, the sequence region lying immediately upstream the conserved core, is predicted to be folded into a stem-loop of limited stability and showing higher sequence divergence among the members of the subfamily (Fig. 6a). The orthologs of the order Sphingomonadales show a higher sequence variability around the conserved core, which in spite of being slightly shorter (5′-UCCUCCCU-3′), it is still predicted to be exposed between one stem-loop and the transcriptional terminator (Supp. Fig. 1), in agreement with the overall features of the members of the here proposed αr8s1 subfamily (Fig. 6).

3.5. Expression of mmgR homologs in Rhizobiales species other than S. meliloti

The MmgR transcript was originally detected by Northern blot in free-living *S. meliloti* strain 2011 as a discrete 77-nt RNA (Sobrero and Valverde, 2011; Valverde et al., 2008), and then recently confirmed by RNAseq (Schluter et al., 2013, 2010). MmgR is also expressed in the nitrogen-fixing *S. meliloti* bacteroids within nodules of *Medicago sativa* (alfalfa) roots (Roux et al., 2014). The presence of transcripts corresponding to *mmgR* homologs was detected by Northern blot in three other members of the Rhizobiales; i.e. *S. fredii* NGR234, *R. leguminosarum* bv *viciae* 3841 and *R. etli* CIAT652 (Fig. 7). In all cases, only one transcript with the expected size was detected in stationary phase cells.

4. Discussion

4.1. Sequence and structural conservation-based search of mmgR homologs and definition of the α-proteobacterial RNA family αr8

Small non-coding regulatory RNAs of the *trans*-acting class (sRNAs) are ubiquitous molecules in prokaryotes of relevant importance for control of gene expression (Wagner and Romby, 2015). As it happens with family proteins, a better understanding of the origin and evolution of a sRNA gene and its genetic context offers a valuable reference frame for interpreting the biological meaning of experimental data from gene characterization. However, unlike protein families, sRNA families rely relatively less on

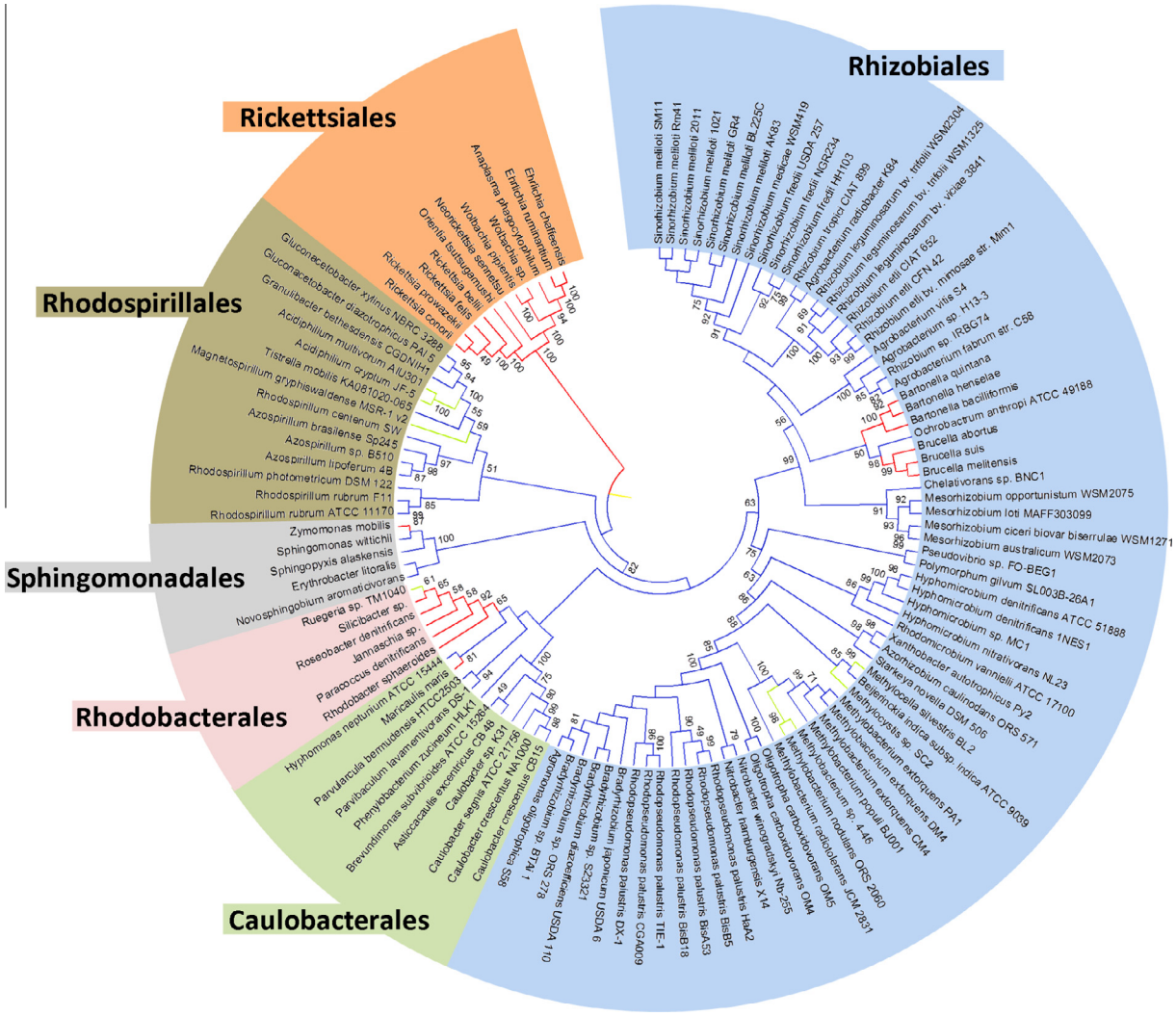


Fig. 5. Phylogenetic distribution of αr8s1 subfamily members. Blue lines indicate that at least one *mmgR* ortholog showing microsynteny with a NFGAH gene homolog was found. Green lines indicate that none of the *mmgR* homologs that were found show microsynteny with a NFGAH homolog. Red lines indicate strains in which no *mmgR* homologs were found. Cladogram construction parameters were the same as those described in Fig. 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sequence conservation during their evolutionary path, and the introduction of structural conservation among the criteria used to search and compare sRNA genes may significantly improve the identification of remote sRNA gene orthologs (Gruber et al., 2008).

The *mmgR* gene of *S. meliloti* strain 2011 encodes a sRNA that fine-tunes storage of PHB under conditions of carbon excess (Lagares (Jr) et al., 2016). We show here that *S. meliloti mmgR* is member of the proposed novel family of sRNA gene orthologs αr8, which most likely has evolved from a eubacterial ancestor that diverged from the order Rickettsiales to give rise to the rest of the species from the class α-proteobacteria (Fig. 4). Noteworthy, we have not detected *mmgR* homologs outside the α-proteobacteria. Through a combination of sequence and structural conservation-based search, we have so far identified a group of 243 *mmgR* genes that are present in 95 sequenced strains covering 70 species. The number of collected *mmgR* homologs increased notably after the first iteration upon introduction of the CM profile-based search (Fig. 3). However, the increment in the number of bacterial species was not of comparable magnitude, thus indicating that most of the *mmgR* sequences identified by using the CM profile were due to recognition of additional copies of the *mmgR* gene in the same genomes in which it was first detected. Data show that the current set of *mmgR* genes are evolutionary linked by a series of speciation and gene duplication events, resulting in subgroups of orthologs and paralogs of recent emergence. Thus, the αr8 family represents the set of sRNA homologs with the widest phylogenetic distribution and the strongest conservation among those non-coding RNA families that have been described in α-proteobacteria (del Val et al., 2012; Reinkensmeier and Giegerich, 2015; Reinkensmeier et al., 2011). In this regard, it must be stressed here that the RFAM family that was originally defined upon identification of the *suhB* element (Corbino et al., 2005), comprised 5 presumed *mmgR* homologs in the genome of *S. meliloti*, including the sRNA genes *abcR1* and *abcR2*, which were reclassified later as members of the separate family αr15 (del Val et al., 2012). Our analysis confirms that *mmgR* homologs have sequence and predicted structural features that clearly distinguish both families, although it does not rule out the possibility that ancient members of the αr8 and αr15 families were remote paralogs with enough similarities to be grouped together by the filtering criteria of RFAM at the time the number of gene sequences available was significantly less than at the moment of this study.

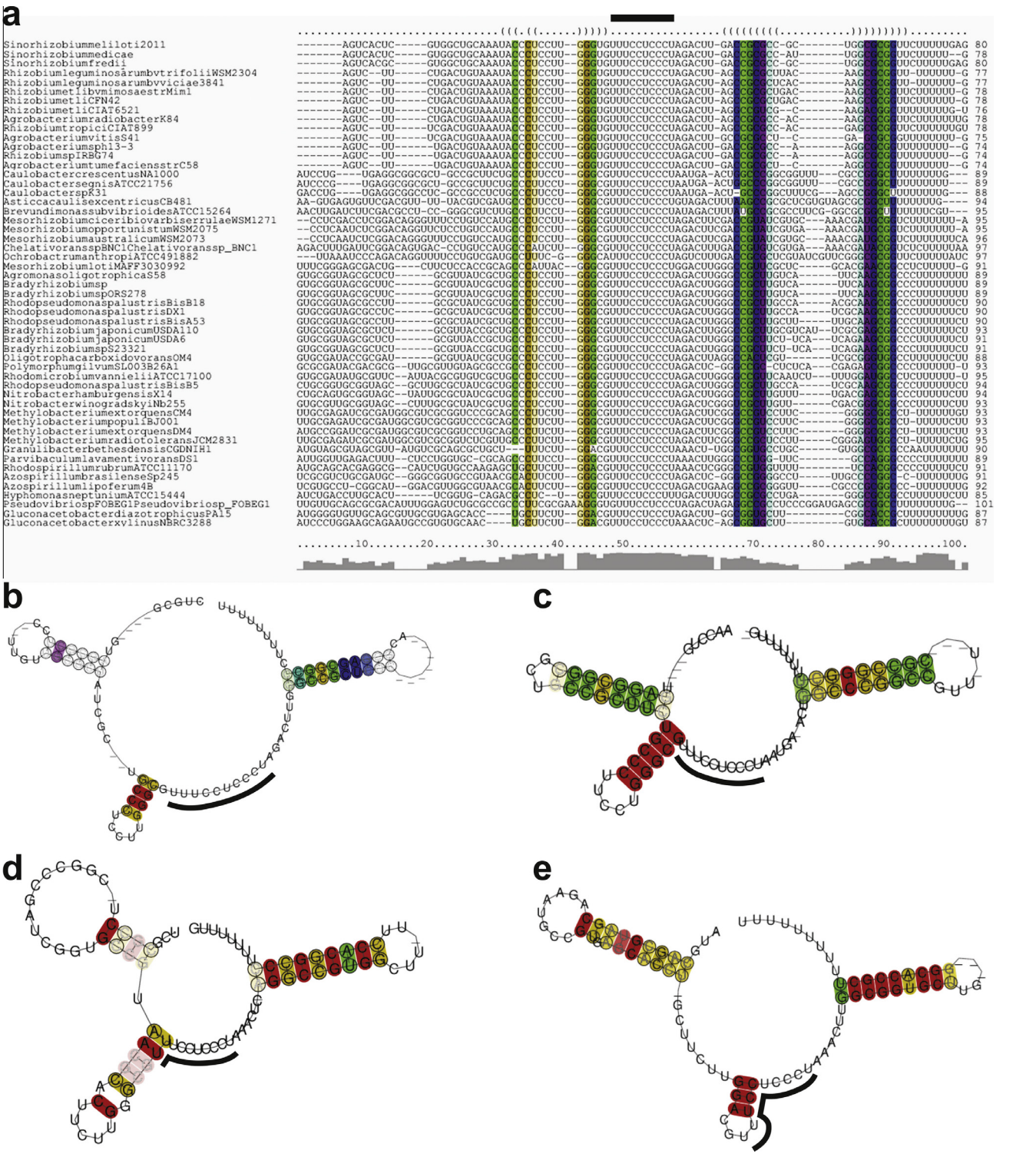


Fig. 6. (a) Multiple sequence/structure alignment of αr8s1 subfamily members and consensus predicted secondary structures for *mmgR* orthologs belonging to the order of Rhizobiales (b), Caulobacterales (c), and the families *Rhodospirillaceae* (d) and *Acetobactereaceae* (e) of the order of Rhodospirilalles.

4.2. Most mmgR alleles show extensive microsynteny

Microsynteny of sRNA genes and neighbouring protein-encoding genes has been reported in several cases (Gimpel et al., 2012; Heeb et al., 2002; Horler and Vanderpool, 2009), and may be a helpful criterion for identification of sRNA homologs having less conserved sequence and/or structural features. In fact, we have retrieved additional members of the αr8 family among the Sphingomonadales that had not been detected by the main workflow with the defined statistical cut-off. On the other hand, by analyzing the microsynteny of the *mmgR* loci within the whole αr8 family, we found that over 90% of the sRNA homologs are physically linked to a gene encoding a putative N-formyl glutamate amidohydrolase protein (Fig. 5). This remarkable loci conservation prompted us to define the αr8s1 subfamily of *mmgR* homologs, as those sRNA genes of the αr8 set that are encoded next to homologs of the *S.*

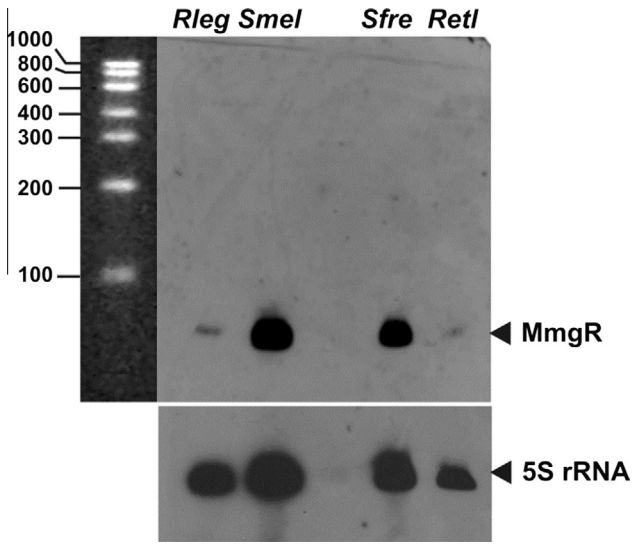


Fig. 7. Detection of MmgR transcripts in stationary phase cells from members of the Rhizobiales (within the αr8s1 subfamily). *Rleg*, *Rhizobium leguminosarum* bv. *viciae* 3841; *Smel*, *S. meliloti* 2011; *Sfre*, *S. fredii* NGR234; *Retl*, *Rhizobium etli* CIAT652. The size (nt) of standard RNA molecules is indicated. The exposition of the hybridized membrane was adjusted to maximize detection of the MmgR transcripts having slight sequence divergence with the utilized DNA probe. The lower panel shows the result of the Northern blot upon membrane striping and hybridization with a dsDNA probe targeting the *S. meliloti* 2011 5S rDNA gene (see Materials and methods for details). The signal intensity of each band cannot be compared quantitatively among the different strains due to the differential sequence identity between the DNA probes generated on the basis of *S. meliloti* genes and the homolog sequences of the other rhizobia.

meliloti SMc04043 gene. Indeed, the *mmgR*–NFGAH ORF physical linkage has a long evolutionary existence within the α-proteobacteria (Fig. 5), and there are a few cases of syntenic loss that have occurred in recently emerged phylogenetic groups as for the genomes of *Ruegeria*, *Methylocella* and *Magnetospirillum* genera. Microsynteny may reflect a functional relationship among physically linked genes, although this is not necessarily verified in all cases. The biological linkage of *mmgR* and the flanking NFGAH ORF remains to be determined experimentally for members of the αr8s1 subfamily.

4.3. Members of the αr8 family share an ultraconserved single-stranded nucleotide sequence stretch (5′-UUUCCUCCCU-3′)

In general, the *mmgR* sRNA genes of the different α-proteobacterial clades display highly conserved elements of primary and secondary structure, such as the fully conserved 10-mer 5′-UUUCCUCCCU-3′, one or two preceding stem-loops, and a downstream typical *Rho*-independent transcriptional terminator. The high sequence conservation of this internal region let us theorize about its involvement in the biological function of the MmgR transcript, assuming that it would be particularly determined by the sequence. This hypothesis is supported by the fact that both flanking regions have relatively less conserved sequences. The evolutionary distance among α-proteobacterial species is such that it has resulted in a significant loss of sequence identity, for instance, at the level of the transcriptional terminator, whose function is known to be determined primarily by its structure. The conserved internal 10-mer has a high probability to be exposed in a single stranded region of the *in silico* consensus structure and it resembles an anti-ribosome binding site motif (anti-RBS), strongly suggesting that it could be involved in translational control of target mRNAs by means of antisense base-pairing interactions, a canonical mechanism of *trans*-acting sRNAs (Storz et al., 2011). It is tempting to speculate that the sequence lying immediately upstream the conserved 10-mer, which is predicted to fold into a rather weak stem-loop, would contribute to determine the specificity of the interaction with a target mRNA. In this model, the common factor among interactions would be the mRNA–sRNA pairing of the RBS/anti-RBS type, thus acting as a seed pairing (Papenfort et al., 2010); this would explain the strong evolutionary pressure to keep the MmgR internal 10-mer invariant in sequence. By contrast, the upstream flanking region that would contribute to specificity of pairing could be subject to compensatory nucleotide replacements to ensure mRNA–sRNA interaction, without compromising other molecular phenomena. This proposed modular structure of the MmgR "active site" might explain the observed difference in the substitution rate among the conserved core and the 5′ flanking region of the *mmgR* orthologs. Indeed, the functional characterization of the members of αr8 family and of their molecular mechanisms will provide solid information on the requirements of sequence and/or structural motifs that may be critical for the activity of these sRNA orthologs (Updegrove et al., 2015). Such experimental support would also reinforce the criteria to search additional members of the family with more remote homology.

4.4. Evolutionary pattern of mmgR RNA genes

The high congruence observed when comparing the occurrence of *mmgR* alleles and the 16S rDNA phylogenetic structure of the corresponding strains (Fig. 4), as well as the marked physical linkage to the neighbouring gene encoding a putative NFGAH (Fig. 5), strongly suggest that the main inheritance mechanism underlying the architecture of the αr8 family might have been vertical gene transfer from an ancestral lineage deeply rooted in the α-proteobacterial tree. Other α-proteobacterial sRNAs with a similar evolutionary pattern are the members of the rhizobial αr7, αr9, αr45 and cuckoo RNA families (del Val et al., 2012; Reinkensmeier and Giegerich, 2015). However, other mechanisms, including gene duplication and lateral transfer, likely contributed in a minor extent to shape the current distribution of *mmgR*. Certain species have suffered genomic rearrangements that resulted in acquisition of multiple copies of members of the αr8 gene family (e.g., *Methylobacterium* spp.) (Fig. 4). The biological implications of this observation are not clear yet for the *mmgR* family, although the presence of functional sRNA paralogs has been reported in other bacterial species (Deng et al., 2012; Kay et al., 2005). Gene duplication may serve to generate paralogs that evolve independently to adopt diverse sequence and structure, and even to achieve differential regulation of its own expression (Updegrove et al., 2015). Interestingly, the genome of *R. etli* CIAT 652 revealed two *mmgR* copies, both separated by a 10-Mb DNA fragment with reminiscence of a mobile genetic element; one of the copies corresponds to the expressed *mmgR* homolog, whereas the second one has lost its promoter region, most probably due to the insertional mechanism of the mobile element (data not shown).

On the other hand, despite the apparently ancient origin of the sRNA gene and its current wide distribution within the α-proteobacteria, we have detected the absence of *mmgR* homologs in certain species, most likely as a direct consequence of a massive genome reduction or due to local sequence rearrangement following a loss of selective pressure to preserve the gene. The first could be the case of *Z. mobilis*, the most recent evolutionary branch within the Sphingomonadales that has experienced a 65% reduction on its genome size compared to the closest species still bearing an *mmgR* gene copy (*Sphingomonas wittichii*). The second explanation for the absence of *mmgR* may apply to *Miracaulis maris*, which belongs to the Caulobacterales and is a close

phylogenetic relative of *Hyphomonas neptunium*, a species that still bears an *mmgR* copy (Supp. Fig. 3). Of special interest results the cases of *Brucella* and *Bartonella* species, both of which have lost *mmgR* (Supp. Fig. 3). These genera have diverted their evolutionary path from that of rhizobia, and established as facultative intracellular animal pathogens. This has been claimed to be the causal origin of the genomic reduction that these microorganisms have suffered (Batut et al., 2004; Boussau et al., 2004; Sallstrom and Andersson, 2005). However, whereas *Bartonella* has completely lost the *mmgR* locus and flanking genes, *Brucella* still shows microsynteny of both neighbouring ORFs in *S. meliloti* but the sRNA gene has disappeared and the homolog intergenic region is 50% shorter than that of the third close relative genus *Ochrobactrum antrophi*, which still bears an *mmgR* ortholog gene (Supp. Fig. 3). *Ochrobactrum anthropi* is a rhizospheric species reported as an opportunistic animal pathogen, and has a genome that is 45% larger than that of *Brucella* spp. most likely due to acquisition of several genomic islands following genome size reduction as *Brucella* (Chain et al., 2011). Interestingly, *O. antrophi* lacks the genes that encode the enzymes involved in PHB metabolism, thus it is tempting to speculate that MmgR might have a different function in this species to that reported for in *S. meliloti*. In this context, three close phylogenetic relative genera show different evolutionary pictures concerning *mmgR*, suggesting that the function(s) of *mmgR* would be related to cellular or physiological processes that have been lost, or whose regulation has become irrelevant upon divergence of their lifestyles.

4.5. On the co-evolution of mmgR alleles and PHB metabolism in α-proteobacteria

mmgR alleles are broadly distributed within α-proteobacteria (Fig. 4) and thus represent an example of a core RNA gene; i.e. it is conserved in most genomes of a phylogenetic group (Skippington and Ragan, 2012). For several characterized core *trans*-encoded sRNAs, it has been observed that certain mRNAs reoccur as their targets in a range of bacteria, which points to a co-evolutionary process (Updegrove et al., 2015). In *S. meliloti*, the MmgR sRNA limits accumulation of PHB (Lagares (Jr) et al., 2016). Interestingly, the genetic capacity to synthesize and store PHB seems to be also broadly distributed in α-proteobacteria, provided representatives from all orders – except for the Rickettsiales – have been reported to metabolize PHB (Kim et al., 1996; Klüttermann et al., 2002; Okon and Itzigsohn, 1995; Van Gool et al., 1971; Wu et al., 2016; Yang et al., 2006). Thus, the occurrence of *mmgR* alleles parallels the phylogenetic distribution of PHB metabolism within all orders in α-proteobacteria, and suggests a deep co-evolutionary linkage between *mmgR* and PHB genetic elements. However, additional experimental evidences about the relevance of *mmgR* genes in the control of PHB accumulation and the identification of the direct target mRNAs in species other than *S. meliloti* are required to provide support to this hypothesis.

4.6. Expression of mmgR in other Rhizobiales

The deduced sequences of *mmgR* orthologs predict RNA transcripts of 70–80 nt, depending on the species (Fig. 2). Northern blot analysis of cells from stationary growth phase grown in defined medium confirmed expression of single transcripts having sizes that match the identified *mmgR* ortholog genes of the αr8s1 subfamily (i.e., *S. fredii*, *R. etli* y *R. leguminosarum* bv *viciae*) (Fig. 7). Previously, MmgR RNAs had been detected in other rhizobia like *S. meliloti* (Valverde et al., 2008) and *Bradyrhizobium japonicum* (Madhugiri et al., 2012), and in other α-proteobacteria like *Caulobacter crescentus* (Landt et al., 2008). In all cases, the size of the MmgR transcripts was coherent with the sequence signals of the corresponding sRNA loci. It would be interesting to decipher the biological relevance of the expression of *mmgR* orthologs in these members of the α-proteobacteria that possess different life styles and interacting hosts, and particularly, to verify if storage of intracellular polyhydroxybutyrate is also a control target of the MmgR sRNA as shown for *S. meliloti* (Lagares (Jr) et al., 2016).

4.7. Evolution of trans-encoded sRNAs and the case of mmgR (αr8 family)

A significant body of experimental evidences support the relevant role of *trans*-encoded sRNAs as regulators of diverse biological processes in bacteria; however, little is known about their evolution, and evolution of their interactions with target mRNAs, within bacterial lineages (Peer and Margalit, 2014). It appears that lineage-specific accumulation of sRNAs, as we observed for *mmgR* within the α-proteobacteria (Fig. 4), is a rather common feature of *trans*-encoded sRNAs (Chen et al., 2014; Peer and Margalit, 2014; Reinkensmeier and Giegerich, 2015; Skippington and Ragan, 2012). For enterobacteria, it has been suggested that the current distribution of sRNA genes might have been associated with the evolution of the protein Hfq, provided interactions between sRNAs and their targets is promoted by this chaperone partner (Peer and Margalit, 2014). With regards to the *mmgR* distribution, all α-proteobacterial lineages have Hfq homologs, with the sole exception of members of the Rickettsiales, which –coincidentally – do not have *mmgR* alleles (Fig. 4). Thus, the presence of Hfq seems to have a strong correlation with the dispersal of the *mmgR* gene in α-proteobacteria, and those lineages that have lost *mmgR* alleles (Fig. 4), may still require Hfq to support other relevant regulatory sRNA–mRNA interactions.

A comprehensive genomic-based analysis of the evolution of enterobacterial sRNAs has recently allowed the definition of two types of sRNA genes: (1) core sRNAs, which are characterized by their broad distribution and a central role in regulatory networks through interaction with multiple target that ultimately constrains the evolution of these sRNAs; (2) variable sRNAs, which have a rather dispersed occurrence pattern, a more peripheral role in the interaction network (one or a few targets), and more keen to be lost during evolution (Skippington and Ragan, 2012). Based on its phylogenetic distribution and its likely vertical inheritance, *mmgR* might be categorized as a core α-proteobacterial sRNA. Nevertheless, the network of MmgR interactors is yet to be determined experimentally.

For a *trans*-encoded sRNA to be maintained in evolution, it is expected that there will be at least one interaction with a target mRNA that confers a selective advantage to the microorganism and imposes a selective pressure. In enterobacteria, most sRNA genes preceded the emergence of sRNA binding sites in mRNA targets; this evolutionary observation is consistent with the hypothesis that the establishment of a productive sRNA–mRNA regulatory pair does impose a selective pressure to keep the sRNA, and further allows the subsequent emergence of additional targets (Peer and Margalit, 2014). In such a co-evolutionary context, it is expected that mRNA binding sites evolve faster the sRNA binding sites, because a single sRNA may regulate multiple mRNAs (Peer and Margalit, 2014). This could be the case of *mmgR* alleles, which display a strikingly conserved internal sequence core that is highly reminiscent of an anti-RBS site (Fig. 6). Current undergoing work to elucidate the set of target mRNAs that are regulated by MmgR

in *S. meliloti* will provide further elements to understand the evolutionary constraints of members of the αr8 family.

4.8. Concluding remarks

To conclude, by contrasting the structure of the α-proteobacterial evolutionary tree with the distribution of *mmgR* gene orthologs, we were able to trace back the *mmgR* evolutionary path at the interspecific level, and in parallel, to recognize orthologs with no microsynteny with the NFGAH encoding gene as lineage variants that have recently emerged in the most recent phylogenetic clades. Altogether, the wide distribution and the high sequence/structure conservation of *mmgR* across the α-proteobacterial class, which includes a marked diversity of lifestyles and shows an important evolutionary distance, suggests that members of the αr8 family have key roles in genetic control of central microbial processes that would only be dispensable in a few species due to yet unknown reasons.

Acknowledgements

This work was supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica, Argentina (PICT 2007 no. 700 and PICT 2012 no. 1777), from CONICET, Argentina (PIP 201101 n° 00031), and from Universidad Nacional de Quilmes, Argentina (PUNQ 1009/11 and 1411/15). A.L. was supported by CONICET fellowships; C.V. is a member of CONICET.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.03.026.

References

Batut, J., Andersson, S.G., O'Callaghan, D., 2004. The evolution of chronic infection strategies in the alpha-proteobacteria. Nat. Rev. Microbiol. 2, 933–945.

Baumgardt, K., Smidova, K., Rahn, H., Lochnit, G., Robledo, M., Evguenieva-Hackenberg, E., 2015. The stress-related, rhizobial small RNA RcsR1 destabilizes the autoinducer synthase encoding mRNA *sinI* in *Sinorhizobium meliloti*. RNA Biol. http://dx.doi.org/10.1080/15476286.2015.1110673 (in press).

Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., Stadler, P.F., 2008. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9, 474.

Boussau, B., Karlberg, E.O., Frank, A.C., Legault, B.A., Andersson, S.G., 2004. Computational inference of scenarios for alpha-proteobacterial genome evolution. Proc. Natl. Acad. Sci. U. S. A. 101, 9722–9727.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., Bateman, A., 2013. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 41, D226–D232.

Corbino, K.A., Barrick, J.E., Lim, J., Welz, R., Tucker, B.J., Puskarz, I., Mandal, M., Rudnick, N.D., Breaker, R.R., 2005. Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. Genome Biol. 6, R70.

Chain, P.S., Lang, D.M., Comerci, D.J., Malfatti, S.A., Vergez, L.M., Shin, M., Ugalde, R. A., Garcia, E., Tolmasky, M.E., 2011. Genome of *Ochrobactrum anthropi* ATCC 49188 T, a versatile opportunistic pathogen and symbiont of several eukaryotic hosts. J. Bacteriol. 193, 4274–4275.

Chen, I.C., Griesenauer, B., Yu, Y.T., Velicer, G.J., 2014. A recent evolutionary origin of a bacterial small RNA that controls multicellular fruiting body development. Mol. Phylogenet. Evol. 73, 1–9.

del Val, C., Romero-Zaliz, R., Torres-Quesada, O., Peregrina, A., Toro, N., Jimenez-Zurdo, J.I., 2012. A survey of sRNA families in alpha-proteobacteria. RNA Biol. 9, 119–129.

Deng, Z., Meng, X., Su, S., Liu, Z., Ji, X., Zhang, Y., Zhao, X., Wang, X., Yang, R., Han, Y., 2012. Two sRNA RyhB homologs from *Yersinia pestis* biovar *microtus* expressed *in vivo* have differential Hfq-dependent stability. Res. Microbiol. 163, 413–418.

Gimpel, M., Preis, H., Barth, E., Gramzow, L., Brantl, S., 2012. SR1–a small RNA with two remarkably conserved functions. Nucleic Acids Res. 40, 11659–11672.

Gonzalez, V., Acosta, J.L., Santamaria, R.I., Bustos, P., Fernandez, J.L., Hernandez Gonzalez, I.L., Diaz, R., Flores, M., Palacios, R., Mora, J., Davila, G., 2010. Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. Appl. Environ. Microbiol. 76, 1604–1614.

Gruber, A.R., Bernhart, S.H., Hofacker, I.L., Washietl, S., 2008. Strategies for measuring evolutionary conservation of RNA secondary structures. BMC Bioinformatics 9, 122.

Heeb, S., Blumer, C., Haas, D., 2002. Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. J. Bacteriol. 184, 1046–1056.

Horler, R.S., Vanderpool, C.K., 2009. Homologs of the small RNA SgrS are broadly distributed in enteric bacteria but have diverged in size and sequence. Nucleic Acids Res. 37, 5465–5476.

Kay, E., Dubuis, C., Haas, D., 2005. Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens* CHA0. Proc. Natl. Acad. Sci. U. S. A. 102, 17136–17141.

Kim, S.W., Kim, P., Lee, H.S., Kim, J.H., 1996. High production of Poly-β-hydroxybutyrate (PHB) from *Methylobacterium organophilum* under potassium limitation. Biotechnol. Lett. 18, 25–30.

Klüttermann, K., Tauchert, H., Kleber, H.P., 2002. Synthesis of Poly-β-hydroxybutyrate by *Agrobacterium radiobacter* after Growth on D-Carnitine. Acta Biotechnol. 22, 261–269.

Lagares, Jr., A., 2015. The Biological Role of the Small Regulatory RNA Sm8 in Free-Living and Symbiotic *Sinorhizobium Melilioti* Ph.D. Thesis; Universidad Nacional de La Plata, Argentina.

Landt, S.G., Abeliuk, E., McGrath, P.T., Lesley, J.A., McAdams, H.H., Shapiro, L., 2008. Small non-coding RNAs in *Caulobacter crescentus*. Mol. Microbiol. 68, 600–614.

Madhugiri, R., Pessi, G., Voss, B., Hahn, J., Sharma, C.M., Reinhardt, R., Vogel, J., Hess, W.R., Fischer, H.M., Evguenieva-Hackenberg, E., 2012. Small RNAs of the *Bradyrhizobium*/*Rhodopseudomonas* lineage and their analysis. RNA Biol. 9, 47–58.

Marzi, S., Romby, P., 2012. RNA mimicry, a decoy for regulatory proteins. Mol. Microbiol. 83, 1–6.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., Lopez, R., 2013. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 41, W597–W600.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., Gautheret, D., 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol. 8, 11–13.

Nawrocki, E.P., Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933–2935.

Oberto, J., 2013. SyntTax: a web server linking synteny to prokaryotic taxonomy. BMC Bioinformatics 14, 4.

Okon, Y., Itzigsohn, R., 1995. The development of *Azospirillum* as a commercial inoculant for improving crop yields. Biotechnol. Adv. 13, 415–424.

Overmars, L., Kerkhoven, R., Siezen, R.J., Francke, C., 2013. MGcV: the microbial genomic context viewer for comparative genome analysis. BMC Genom. 14, 209.

Papenfort, K., Bouvier, M., Mika, F., Sharma, C.M., Vogel, J., 2010. Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA. Proc. Natl. Acad. Sci. U.S.A. 107, 20435–20440.

Peer, A., Margalit, H., 2014. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA 20, 994–1003.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glockner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glockner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596.

Reese, M.G., 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput. Chem. 26, 51–56.

Reinkensmeier, J., Giegerich, R., 2015. Thermodynamic matchers for the construction of the cuckoo RNA family. RNA Biol. 12, 197–207.

Reinkensmeier, J., Schluter, J.P., Giegerich, R., Becker, A., 2011. Conservation and occurrence of trans-encoded sRNAs in the rhizobiales. Genes 2, 925–956.

Robledo, M., Frage, B., Wright, P.R., Becker, A., 2015. A stress-induced small RNA modulates alpha-rhizobial cell cycle progression. PLoS Genet. 11, e1005153.

Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., Carrere, S., Sallet, E., Courcelle, E., Moreau, S., Debelle, F., Capela, D., de Carvalho-Niebel, F., Gouzy, J., Bruand, C., Gamas, P., 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. 77, 817–837.

Sallet, E., Roux, B., Sauviac, L., Jardinaud, M.F., Carrere, S., Faraut, T., de Carvalho-Niebel, F., Gouzy, J., Gamas, P., Capela, D., Bruand, C., Schiex, T., 2013. Next-generation annotation of prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. DNA Res. 20, 339–354.

Sallstrom, B., Andersson, S.G., 2005. Genome reduction in the alpha-Proteobacteria. Curr. Opin. Microbiol. 8, 579–585.

Sambrook, J., Fritsch, E., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

Schluter, J.P., Reinkensmeier, J., Barnett, M.J., Lang, C., Krol, E., Giegerich, R., Long, S. R., Becker, A., 2013. Global mapping of transcription start sites and promoter motifs in the symbiotic alpha-proteobacterium *Sinorhizobium meliloti* 1021. BMC Genom. 14, 156.

Schluter, J.P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Janicke, S., Becker, J.D., Giegerich, R., Becker, A., 2010. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. BMC Genom. 11, 245.

Schmeisser, C., Liesegang, H., Krysciak, D., Bakkou, N., Le Quere, A., Wollherr, A., Heinemeyer, I., Morgenstern, B., Pommerening-Roser, A., Flores, M., Palacios, R., Brenner, S., Gottschalk, G., Schmitz, R.A., Broughton, W.J., Perret, X., Strittmatter, A.W., Streit, W.R., 2009. Rhizobium sp. strain NGR234 possesses a remarkable number of secretion systems. Appl. Environ. Microbiol. 75, 4035–4045.

Skippington, E., Ragan, M.A., 2012. Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella* genomes. Genome Biol. Evol. 4, 330–345.

Smith, C., Heyne, S., Richter, A.S., Will, S., Backofen, R., 2010. Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res. 38, W373–W377.

Sobrero, P., Valverde, C., 2011. Evidences of autoregulation of *hfq* expression in *Sinorhizobium meliloti* strain 2011. Arch. Microbiol. 193, 629–639.

Storz, G., Vogel, J., Wassarman, K.M., 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell 43, 880–891.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Torres-Quesada, O., Millan, V., Nisa-Martinez, R., Bardou, F., Crespi, M., Toro, N., Jimenez-Zurdo, J.I., 2013. Independent activity of the homologous small regulatory RNAs AbcR1 and AbcR2 in the legume symbiont *Sinorhizobium meliloti*. PLoS ONE 8, e68147.

Torres-Quesada, O., Reinkensmeier, J., Schluter, J.P., Robledo, M., Peregrina, A., Giegerich, R., Toro, N., Becker, A., Jimenez-Zurdo, J.I., 2014. Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in *Sinorhizobium meliloti*. RNA Biol., 11.

Updegrove, T.B., Shabalina, S.A., Storz, G., 2015. How do base-pairing small RNAs evolve? FEMS Microbiol. Rev. 39, 379–391.

Valverde, C., Livny, J., Schluter, J.P., Reinkensmeier, J., Becker, A., Parisi, G., 2008. Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. BMC Genom. 9, 416.

Van Gool, A.P., Tobback, P.P., Fischer, I., 1971. Autotrophic growth and synthesis of reserve polymers in *Nitrobacter winogradskyi*. Arch. Mikrobiol. 76, 252–264.

Vincent, J.M., 1970. A Manual for the Practical Study of Root Nodule Bacteria. Blackwell Scientific Publications, Oxford.

Wagner, E.G., Romby, P., 2015. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv. Genet. 90, 133–208.

Williams, K.P., Sobral, B.W., Dickerman, A.W., 2007. A robust species tree for the alphaproteobacteria. J. Bacteriol. 189, 4578–4586.

Wu, M., Li, G., Huang, H., Chen, S., Luo, Y., Zhang, W., Li, K., Zhou, J., Ma, T., 2016. The simultaneous production of sphingan Ss and poly(R-3-hydroxybutyrate) in *Sphingomonas sanxanigenens* NX02. Int. J. Biol. Macromol. 82, 361–368.

Yang, M.K., Lin, Y.C., Shen, C.H., 2006. Identification of two gene loci involved in poly-beta-hydroxybutyrate production in *Rhodobacter sphaeroides* FJ1. J. Microbiol. Immunol. Infect 39, 18–27.

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glockner, F.O., 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42, D643–D648.

Young, J.P., Crossman, L.C., Johnston, A.W., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., Wexler, M., Curson, A.R., Todd, J.D., Poole, P.S., Mauchline, T.H., East, A.K., Quail, M.A., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., Davis, P., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., Whitehead, S., Parkhill, J., 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. 7, R34.