Small RNA Deep-Sequencing Analyses Reveal a New Regulator of Virulence in *Agrobacterium fabrum* C58

M. Dequivre,^{1,2,3} B. Diel,^{1,2,3,4,5} C. Villard,^{1,2,3} O. Sismeiro,⁶ M. Durot,^{7,8} J. Y. Coppée,⁶ X. Nesme,^{1,2,4,5} L. Vial,^{1,2,4,5} and F. Hommais^{1,2,3}

¹Université de Lyon, F-69622, Lyon, France; ²Université Lyon 1, F-69622 Villeurbanne, France; ³CNRS, UMR 5240 Microbiologie Adaptation et Pathogénie, F-69622 Villeurbanne, France; ⁴CNRS, UMR 5557 Ecologie Microbienne, F-69622 Villeurbanne, France; ⁵INRA, USC 1364 Ecologie Microbienne, F-69622 Villeurbanne, France; ⁶Plate-forme Transcriptome et Epigénome, Département Génomes et Génétique, Institut Pasteur, 25 rue du Dr. Roux, F75015 Paris, France; ⁷CEA/DSV/FAR/IG/Genoscope and CNRS UMR8030 Laboratoire d'Analyses Bioinformatiques en Métabolisme et Génomique, 2 rue Gaston Crémieux 91057 Evry cedex, France; ⁸Total New Energies USA, 5858 Horton Street, Emeryville, CA 94608, U.S.A.

Submitted 30 October 2014. Accepted 14 January 2015.

Novel ways of regulating Ti plasmid functions were investigated by studying small RNAs (sRNAs) that are known to act as posttranscriptional regulators in plant pathogenic bacteria. sRNA-seq analyses of *Agrobacterium fabrum* C58 allowed us to identify 1,108 small transcripts expressed in several growth conditions that could be sRNAs. A quarter of them were confirmed by bioinformatics or by biological experiments. Antisense RNAs represent 24% of the candidates and they are over-represented on the pTi (with 62% of pTi sRNAs), suggesting differences in the regulatory mechanisms between the essential and accessory replicons. Moreover, a large number of these pTi antisense RNAs are transcribed opposite to those genes involved in virulence. Others are 5'- and 3'-untranslated region RNAs and *trans*-encoded RNAs. We have validated, by rapid amplification of cDNA ends polymerase chain reaction, the transcription of 14 *trans*-encoded RNAs, among which RNA1111 is expressed from the pTiC58. Its deletion decreased the aggressiveness of *A. fabrum* C58 on tomatoes, tobaccos, and kalanchoe, suggesting that this sRNA activates virulence. The identification of its putative target mRNAs (6b gene, *virC2*, *virD3*, and *traA*) suggests that this sRNA may coordinate two of the major pTi functions, the infection of plants and its dissemination among bacteria.

While regulation by transcription factors is widespread in plant-bacteria interactions, there is growing evidence that riboregulators (sRNAs) could act as activators or repressors at a posttranscriptional level. The untranslated 5' ends of genes (5' UTR RNA), acting as riboswitches or RNA thermometers, are known to be involved in posttranscriptional regulation (Bastet et al. 2011). sRNAs usually range from 50 to 500 nucleotides (nt) in length and they mainly act by regulating mRNA translation or stability through base-pairing (Storz et al. 2011). Antisense RNAs (asRNAs) are encoded in the opposite strand of their target and *trans*-encoded RNAs are present in intergenic regions, showing limited complementarity with their

multiple targets localized elsewhere in the genome. Biocomputational predictions and experimental approaches, including high-throughput sequencing of cDNA (RNA-seq), have led to the discovery of many sRNAs, and hundreds have been shown to be transcribed in different bacteria (Storz et al. 2011). However, overlaps between data are restricted, probably because of differences between the methods or the growth conditions used, and work is needed to improve our knowledge of the sRNA content in bacterial species. Furthermore, attempts to functionally characterize putative sRNAs have been restricted to a small number. The characterization of *trans*-encoded RNA targets is intricate, since the posttranscriptional control by sRNAs needs only a limited complementary sequence (around only 10 nt), which is difficult to predict (Backofen et al. 2014; Li et al. 2012).

Using a combination of sRNA prediction and deep-sequencing analyses, the sRNA content of several plant-interacting bacteria has been studied (del Val et al. 2007; Lee et al. 2013; Schlüter et al. 2013; Schmidtke et al. 2012; Wilms et al. 2012a; Zeng and Sundin 2014) and several of these sRNAs are thought to be involved in plant-bacteria interactions. Indeed, various RNA-seq studies of *Sinorhizobium meliloti* revealed a plethora of small untranslated transcripts, suggesting major roles for riboregulators in the adaptive flexibility required for the complex lifestyle of the family *Rhizobiaceae* (Becker et al. 2014; del Val et al. 2007; Schlüter et al. 2010, 2013).

Agrobacterium is a bacterial genus commonly found in rhizospheres, in which its members exercise generally benign commensalism with plants. However, when harboring a virulence plasmid called the Ti plasmid (pTi), agrobacteria behave as plant pathogens able to induce the formation of tumors (crown galls) in almost all wounded dicots and gymnosperms and in some monocots. The strain *A. fabrum* C58 is widely used to study the pathogenicity of agrobacteria. C58 was named in former papers as *A. tumefaciens*, but this terminology is considered irrelevant because it has been given to several species, and it was recently renamed *A. fabrum* (Lassalle et al. 2011; Mousavi et al. 2014). The *A. fabrum* C58 genome is composed of four replicons: two chromosomes and two accessory plasmids, called pAtC58 and pTiC58 (Goodner et al. 2001; Wood et al. 2001), the latter encoding major virulence factors (Pitzschke and Hirt 2010). Pathogenic agrobacteria affect the host plant by exporting T-DNA, a DNA segment of the pTi, through a typical type IV secretion system (encoded by the *vir* region). The T-DNA is integrated into the nuclear plant genome and expression of its genes by transformed

Corresponding author: Florence Hommais;
E-mail: florence.hommais@univ-lyon1.fr

*The e-Xtra logo stands for “electronic extra” and indicates that five supplementary figures and seven supplementary tables are published online.

cells results in (i) the overproduction of phytohormones responsible for the tumor formation and (ii) the synthesis of opines. These compounds are amino-sugars used as nutrient sources by pTi-harboring agrobacteria. Opines also induce the conjugative transfer and dissemination of pTis. To induce virulence, agrobacteria must (i) perceive wound-released chemical signals from plants, (ii) move to the wound site in the plant, (iii) attach themselves to the plant cells, and (iv) transform them through a cascade of events.

Recently, two deep-sequencing studies revealed hundreds of sRNAs in *A. fabrum* C58 (Lee et al. 2013; Wilms et al. 2012a). Some of them were differentially expressed in the presence of acetosyringone and some others were characterized, namely the asRNA RepE, involved in the control of pTiR10 replication (Chai and Winans 2005), and the sRNA AbcR1, involved in the regulation of the GABA transporter in *A. fabrum* C58 (Overlöpfer et al. 2014; Wilms et al. 2011). Furthermore, the absence of the RNA chaperone Hfq reduced *A. fabrum* C58 virulence (Wilms et al. 2012b), suggesting an important role for sRNAs during the interaction of agrobacteria with plants. Here, using the *A. fabrum* C58 sRNA-seq analyses that specifically target small transcripts, we detected a plethora of putative small RNAs (sRNAs) and confirmed more than a quarter of them by bioinformatics predictions or with transcriptomic assays. Several sRNAs were identified, and one of them, expressed from pTiC58, was functionally characterized, showing its involvement in the control of virulence.

RESULTS

Identification of small transcripts.

Size-selected sRNAs were isolated from *A. fabrum* C58 cells cultivated in various growth conditions, using different media (rich and minimal) and growth phases (exponential and stationary). They were strand-specific sequenced on an Illumina GAI (Perkins et al. 2009), and coverage depths of the un-repeated genome sequences were calculated for each nucleotide position in the forward and the reverse strands. The resulting coverage plots were visualized on the Artemis genome browser (Rutherford et al. 2000), and genome portions with contiguous sequencing coverage were analyzed manually. Resulting regions with an average depth at least 10 times greater than the background (estimated by the average depth of the adjacent regions) were considered as transcribed. Finally, signals detected from each library were compared and those transcripts found in all the libraries were considered as putative sRNA candidates. These results were compared with novel sRNAs identified with the Rockhopper software system. A total of 1,108 putative sRNAs were selected and were named according to their position on the genome. Small transcripts ranged from 18 to 576 nt in length (median = 117 nt), in accordance with the size-selection performed. Among the 1,108 small transcripts, 602 were found on the circular chromosome, 291 on the linear one, 140 on pAtC58, and 75 on pTiC58, with respect to the size of replicons (Fig. 1). With the exception of the repeated regions, small transcripts were homogeneously distributed along genome sequences at a density of about one sRNA candidate per 5 kb (one sRNA for every five proteins) (Supplementary Fig. S1).

Putative sRNAs were also classified according to four categories: (i) 407 putative sRNAs were localized in intergenic regions (IGR) and correspond to putative *trans*-encoded RNAs, (ii) 402 were localized near or overlapped the translation start of the coding sequence (CDS) (5' UTR), (iii) 37 were localized near or overlapped the termination of the CDS (3' UTR), and (iv) 262 were localized in the opposite strand of the CDS, corresponding to the asRNAs (Fig. 1).

In silico prediction of sRNA candidates in *A. fabrum* C58.

Sequences of sRNA candidates were compared with the Rfam database and significant similarities were found for 31 candidates (Supplementary Table S1). As only a few sRNAs are in the Rfam database, we performed an in silico prediction of sRNAs present in the *A. fabrum* C58 genome using the sRNAPredict algorithm. Comparative genomics were performed between (i) the *A. fabrum* C58 genome, (ii) the genomes of seventeen *Agrobacterium* strains belonging to different genomospecies (AgrobacterScope database), and (iii) the genomes of *A. vitis* S4, *A. radiobacter* K84, and *S. meliloti* 1021 strains. Using such an approach, 805 regions (from 60 to 550 nt) were predicted to encode putative sRNAs: 469 were localized on the circular chromosome, 214 on the linear one, 91 on pAtC58, and 31 on pTiC58. A comparison with the expressed small transcripts identified above showed that there were 129 features in common (Fig. 2). Thus, about 15% of the transcribed sRNAs were correctly predicted.

Validation of sRNA candidates using comparative transcriptomics.

Putative sRNAs were also compared with those identified by previous RNA-seq analyses performed in *A. fabrum* C58 (Lee et al. 2013; Wilms et al. 2012a). A total of 102 had been identified by Wilms et al. (2012a) and 148 by Lee et al. (2013). Of these, 45% of the sRNAs identified by Wilms et al. (2012a) and 31% of the sRNAs identified by Lee et al. (2013) were recovered in our analysis (Fig. 2). Combining these three RNA-seq analyses, 1,560 small transcripts could be identified in *A. fabrum* C58, and among these, 54 were detected in all three of the RNA-seq analyses. They correspond to (i) the sRNAs largely conserved among bacteria phyla and already referred to the Rfam database, such as RNaseP, tmRNA, and ssrS (RNA454, RNA441, and RNA734, respectively), (ii) sRNAs already characterized in various agrobacteria, such as AbcR1 (RNA479) (Chai and Winans 2005; Wilms et al. 2011), or (iii) unknown sRNAs whose presence had been previously validated by Northern blot analyses or by 5' and 3' rapid amplification of

Fig. 1. Distribution of the 1,108 small transcripts among the *Agrobacterium fabrum* C58 genome according to sRNA categories. Histograms represent the percentage of each sRNA category per replicon (circular and linear chromosomes, and pAtC58 and pTiC58). Numbers correspond to the number of sRNAs found in each category. UTR = untranslated regions, asRNA = antisense RNA.

cDNA ends-polymerase chain reaction (RACE-PCR). Moreover, 12 candidates were similar to the sRNAs previously identified in *S. meliloti* or *Rhizobium* (Schlüter et al. 2010; Ulvé et al. 2007; Valverde et al. 2008). Finally, combining all these RNA-seq analyses, 196 of our 1,108 candidates (17.7%) had been found previously and 828 were identified here for the first time (Fig. 2), among those are the 21 shortest transcripts (length $\leq$ 30 nt) (Behrens et al. 2014; Liu et al. 2009).

Characterization of new putative sRNAs.

Among the novel small transcripts, 17 were arbitrarily chosen for validation. RACE-PCR analyses were performed on total RNAs. For every candidate, the 5' and 3' ends could be easily identified, suggesting that the regions of interest were actually expressed. Evaluation of the 5' and 3' ends and of the length of each transcript, highlighted (i) differences in 3' end determination (median of difference = 36 bases), likely due to the instability of RNA 3' end because of the 3' ribonuclease degradation (Behrens et al. 2014; Hui et al. 2014), (ii) a very good correlation between 5' end identification (differences $\leq$ 3 bases), and (iii) the small lengths of the resulting transcripts ($\leq$ 270 bases) (Table 1). Thus, the 17 candidates seem to be small transcripts and could be sRNAs.

To gain further insights into the characterization, a search for the presence of CDS and putative RBS (ribosome binding sites) was performed using ORFfinder. For 11 transcripts, no CDS could be identified, indicating that they are untranslated sRNAs. Partial open reading frames (without initiation or stop codons) could be identified for RNA535, RNA834, RNA846, RNA869, and a putative RBS could be predicted upstream from RNA846 and RNA869. Nevertheless, no significant similarities could be identified, except for long hypothetical proteins that showed significant similarities with both RNA869 (5' UTR of *atu4669*) and *atu4669*, suggesting a mistake in the annotation of translation initiation. For the last two transcripts, RNA868 and RNA960, a putative RBS and complete CDS were identified, suggesting that these two genes could encode peptides specific to *A. fabrum* C58, as no significant similarities were found in other bacterial sequences.

Secondary structures of these small transcripts were predicted using the RNAfold algorithm with the calculation of minimal free energy. Base pair probabilities were analyzed, as sRNAs are generally described as strongly structured. For at least 10 structures, base pair probabilities were close to 1, suggesting the presence of strong hairpins in most of the transcripts analyzed (Table 1). For the seven other structures, the average of the base pair probabilities was below 0.5; four of them had been

previously classified as 5' UTR RNAs and the three others were localized in the intergenic region (IGR sRNA). Sequence conservations of these putative sRNAs were then evaluated, and the results showed that their presence is restricted to *Rhizobiaceae*: two of them are largely conserved (RNA834 and RNA846), six seem to be restricted to *Agrobacterium* spp. (RNA115, RNA141, RNA154, RNA275, RNA387, and RNA869), and the others are occasionally present in some members of family *Rhizobiaceae* and some *Agrobacterium* spp.

Differential expression of sRNAs.

We identified the differentially expressed sRNAs using the Rockhopper software system (q value < 0.01) (McClure et al. 2013). We centered our analysis on the 297 sRNAs also predicted by bioinformatics, detected in other transcriptomic experiments, or validated here by RACE-PCR. We compared their expression levels in the stationary phase versus exponential phase and their growth in rich medium versus minimal medium. Of the total, 58 sRNAs were shown to be differentially expressed between growth-phases: six were down-regulated in the exponential phase and 52 were up-regulated. In addition, 56 sRNAs were differentially regulated between growth media: one was down-regulated in minimal medium and 55 were up-regulated (Supplementary Table S2). Finally, about 20% of the analyzed sRNAs were differentially expressed, either according to the growth phase or the growth media, and only five sRNAs (RNA1066, RNA1077, RNA1094, RNA1096, and RNA1113) were transcribed from pTiC58.

RNA1111, a small transcript expressed from the Ti plasmid, involved in virulence.

The RNA1111 is transcribed from the minus strand of the pTiC58, between nucleotide 205,448 and 205,620 (RACE-PCR data), and consequently, its genomic localization is between gene *atu6186* (*virE3*) and *atu6188* (*virE0*) (Supplementary Fig. S2). A conservation search showed significant conservation of this sRNA. Nine of the 15 Ti plasmids presented *rna1111* homologous genes completely identical to the C58 *rna1111* gene; the synteny was also conserved. All nine Ti plasmids are nopaline type. An additional Ti plasmid presented a *rna1111* homolog. It displays 71% identity and was found in the CFBP5473 strain, which harbors a chrysopine-type Ti plasmid (Supplementary Fig. S3). Since this gene is localized inside the *vir* region, its expression level was measured in virulence-mimicking conditions, but no significant changes were observed (data not shown). Its impact on bacterial virulence was also tested. Bacterial aggressiveness was evaluated on tomato plants for mutants deleted for or overexpressing the genomic region expressing RNA1111. Quantitative reverse transcription (qRT)-PCR analyses were performed to measure *rna1111* expression level in the different mutant strains. RNA1111 could not be detected in this deleted strain and, when compared with the wild-type strain harboring the empty vector, an increase of 6.5-fold was evaluated in the strain harboring *rna1111* cloned into the pBBR1MCS-5. A similar level of aggressiveness was observed both for the mutant overexpressing *rna1111* (AT199) and the wild-type strains (C58 and AT200, the C58 with the empty pBBR1MCS-5). Remarkably, the strain deleted for *rna1111* (AT192) showed a drastic decrease in aggressiveness on plants, with an average of two tumors per plant, whereas the wild-type presented more than 20 tumors per plant (Fig. 3). A significant increase in virulence symptoms was observed when the *rna1111* deleted strain was compared with the deleted strain complemented with *rna1111* carried alone in the plasmid pBBR1MCS-5 (9.5 vs. 2 tumors, on average, for complemented and deleted strains, respectively; P value < 0.003). Thus, RNA1111 alone is able to restore, at least partially, virulence. Similar results were also obtained on tobacco

Fig. 2. Comparison between the different RNA-seq identifications of sRNAs and sRNAs prediction in *Agrobacterium fabrum* C58. The Venn diagram represents sRNAs identified in common between the prediction data and the sRNA-seq data performed in this study and the RNA-seq data obtained previously (Lee et al. 2013; Wilms et al. 2012a).

Table 1. Validation of novel sRNAs by rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR)

sRNA	Strand	Position sRNA-seq	Position RACE-PCR	Size sRNA-seq	Size RACE-PCR	Secondary structure prediction (RNAfold)
Circular chromosome						
RNA115	–	534338-534448	534375-534449	110	74	
RNA141	+	693246-693312	693248-693381	66	133	
RNA154	+	762537-762638	762538-762623	101	85	
RNA275	+	1279489-1279713	1279489-1279716	224	227	
RNA387	–	1757820-1757974	1757829-1757973	154	144	
RNA535	+	2409060-2409165	2409062-2409193	105	131	
RNA576	–	2659267-2659331	2659260-2659323	64	63	
Linear chromosome						
RNA779	–	1168075-1168226	1168156-1168227	141	71	
RNA807	+	1310255-1310367	1310256-1310349	112	93	
RNA834	+	1583219-1583390	1583222-1583346	171	124	
RNA846	+	1647094-1647233	1647102-1647289	139	187	
RNA847	–	1648470-1648623	1648521-1648622	153	101	
RNA868	+	1824963-1825233	1824964-1825182	270	218	

(continued on next page)

^a secondary structure with base pair probabilities close to one

and *Kalanchoe daigremontiana* plants (data not shown). Taken together, these findings indicate the involvement of RNA1111 in *A. fabrum* C58 virulence on plants.

mRNA targets of RNA1111.

RNA1111 belongs to the *trans*-encoded RNA class. To further investigate its function, its putative mRNA targets were predicted using a combination of sRNA target prediction algorithms (RNApredator, sTarPicker, and IntaRNA). Annotated CDS and regions upstream of the translation start of CDS were analyzed. To reduce the number of false positives, we selected only the target genes found by every algorithm. This procedure led to the selection of eight target candidates; only the six

localized on pTiC58 were selected for further investigation (Supplementary Table S3). Putative regions of interaction are located 91 bases upstream from the start codon for *atu6014* (the *6b* gene), 32 bases upstream from the start codon for *atu6179* (*virC2*), 31 bases upstream from the start codon for *atu6039* (*trbD*), 88 bases downstream from the start codon for the gene *atu6072* (encoding an hypothetical protein), 64 bases downstream from the start codon for *atu6127* (*traA*), and 40 bases downstream from the start codon for *atu6183* (*virD3*) (Fig. 4). Hence, three candidates (the *6b* gene, *virC2*, and *virD3*) are related to virulence functions, and two other ones are related to the bacterial conjugative transfer of pTiC58 (*traA* and *trbD*). To validate these putative interactions, conservation searches were

Table 1. (continued on previous page)

sRNA	Strand	Position sRNA-seq	Position RACE-PCR	Size sRNA-seq	Size RACE-PCR	Secondary structure prediction (RNAfold)
RNA869	+	1830152-1830293	1830152-1830354	141	202	
At plasmid						
RNA960	+	214071-214308	214074-214288	237	214	
RNA1028	-	492222-492349	492186-492346	127	160	
Ti plasmid						
RNA1111	-	205455-205545	205448-205620	90	173	

Fig. 3. Tests of virulence on tomatoes for *rna1111* mutants. Wild-type strain (WT), a wild-type strain containing empty cloning vector pBBR1MCS-5 (AT200), strains deleted for *rna1111* (AT192), a strain overexpressing *rna1111* (AT199), and a deleted strain complemented with the pBBR1MCS-5 carrying *rna1111* alone (AT203) were all tested for their ability to induce tumors on 3-week-old tomato stems. Bacterial genotypes showing statistical differences (Kruskal-Wallis test) in their aggressiveness are noted with three stars (comparison with wild-type strain) or three hashtags (comparison with deleted strain), corresponding to a *P* value ≤ 0.003.

performed in strains harboring *rna1111* homolog genes (Supplementary Table S4). In addition, we performed qRT-PCR from bacteria cultivated in rich medium and in medium mimicking virulence, using primers localized downstream from the predicted region of interaction. In the tested conditions, *trbD* was not significantly expressed. No significant variation in expression was observed for the *atu6183* (*virD3*) and *atu6179* (*virC2*) genes whatever the strain analyzed, whereas decreases in expression of 12-, 6-, and 3.5-fold were observed for *atu6127* (*traA*), *atu6072* (hypothetical protein), and *atu6014* (the *6b* gene), respectively, when we compared the strain deleted for *rna1111* with the wild-type strain (Fig. 4). Moreover, their expression levels were completely restored in the *rna1111* deleted strain complemented with *rna1111* carried alone in a plasmid (Fig. 4). Conversely, the strain overexpressing *rna1111* (AT199) showed similar expression levels of target genes when compared with expression levels obtained with the wild-type strain harboring the empty vector (AT200) (data not shown). Taken together, these results suggest that the absence of RNA1111 destabilized, directly or indirectly, the mRNAs of these three genes.

DISCUSSION

Regulation by sRNAs is now widely recognized in bacteria but only a few examples concern pathogenic bacteria interacting with plants. Hence, attempts had been made to increase our knowledge on the *Agrobacterium* RNAomes (Lee et al. 2013; Möller et al. 2014; Wilms et al. 2011). Here, we undertook an exhaustive analysis of small transcripts in the model strain *A. fabrum* C58. Our approach led to the discovery of numerous small transcripts, including those expressed at low levels, among those are Ti plasmid sRNA candidates. In agreement with previous data (Wilms et al. 2012a), the latter are shown here to be six times lower than those of the chromosomes in *A. fabrum* C58. These pTi characteristics could significantly impede an automatic detection of transcripts and, consequently, necessitate manual detection and validation of the 5' and 3' ends of new transcripts. In addition, our data were found to be quantitative, as differential growth phase expression analyses were in accordance with previous quantifications (Lee et al. 2013; Wilms et al. 2012a) and relative

Fig. 4. Regions of interaction between predicted target mRNAs and RNA1111. **A**, Sequences around the start codon of predicted target genes are presented with the localization of the region of interaction with RNA1111. Fold changes, measured by quantitative reverse transcription-polymerase chain reaction (qRT-PCR) in the strain deleted for *rna1111* versus the wild-type strain, are indicated in brackets (culture in rich medium and in virulence induction medium). **B**, The predicted secondary of RNA1111 is presented with the localization of the predicted regions of interaction for each putative target mRNA. ND = not determined. **C**, Measurement of differential expression level by qRT-PCR for *atu6014*, *atu6072*, *atu6127*, *atu6179*, and *atu6184* compared with the wild-type strain (C58 or AT200, respectively). Fold changes between the deleted strain $\Delta rna1111$ (AT192) and the wild-type strain C58 are represented in gray and fold changes between the complemented strain $\Delta rna1111$ /pBBR1MCSS::*rna1111* (AT203) and the wild-type strain (AT200) in black. Fold change was calculated with the $2^{-\Delta\Delta CT}$ method. Shown here are the mean values and standard error of at least two biological replicates. Asterisks (*) indicate $P < 0.01$.

quantifications agreed with those performed by Northern blot analyses on total RNA (Supplementary Figs. S4 and S5).

The 1,108 small transcripts identified might be sRNAs, but they could also be processed or degraded RNAs (even if small transcripts from internal mRNA were not taken into account by our procedure). Hence, their validation by other approaches was essential. We compared our sRNA-seq data with sRNAPredict data and with previous RNA-seq data. These comparisons allowed the confirmation of a quarter of sRNA candidates. However, small overlaps between results were observed (Fig. 2). The small overlap with predicted data likely reflects limitations of RNA prediction algorithms, which generally fail to predict a large portion of sRNAs and can also present false positives (Soutourina et al. 2013). Novel small transcripts identified could, therefore, be useful for improving prediction algorithms. Furthermore, the different RNAseq analyses were performed from agrobacteria cultivated in different growth conditions and with different sequencing approaches (Lee et al. 2013; Wilms et al. 2011), and unrecovered sRNAs could be due to their absence of expression in some of the growth conditions tested. Despite these features, a large proportion of previous identified sRNAs (>31%) could be validated by our analysis, illustrating the quality of our approach.

One third of putative sRNAs are localized upstream of a CDS (5' UTR) and most of them correspond to 5' UTR longer than 100 nt. The presence of these small transcripts has been previously described in other bacteria (Gómez-Lozano et al. 2012; Kawano et al. 2005). Long 5' UTR are not unusual in *Rhizobiaceae* species (Schlüter et al. 2013). These long mRNA leader sequences may correspond to (i) processed *cis*-regulatory elements of longer mRNAs, such as metabolite-controlled riboswitches and RNA thermometers (Bastet et al. 2011) or high-affinity protein-binding sites that protect the transcripts from exoribonucleolytic degradation (for example, Hfq), (ii) *trans*-encoded sRNAs resulting from parallel transcriptional output (Loh et al. 2009), or (iii) artifacts in the sequencing procedures. RACE-PCR experiments on total RNAs showed that, at least, RNA275, RNA387, and RNA868 are not small transcript artifacts (Table 1) and could encode short peptides, known to be a characteristic feature in a mechanism called transcriptional attenuation (Mentz et al. 2013). Generally, small open reading frames (ORF) are found upstream of the CDS of metabolic proteins and 20% of CDS with a long 5' UTR are actually involved in metabolism, five harbored a putative ORF and three present an ORF on the opposite strand, suggesting that putative sRNAs could also be the asRNAs of those peptide-encoding transcripts (RNA111, RNA573, and RNA575).

The *cis*-encoded RNAs, asRNAs, represented a quarter of the small transcripts (Fig. 1). They were found on the opposite strand of approximately 5% of the CDS. Surprisingly, the expression of *repE* asRNA (Chai and Winans 2005) was not detected because it was below the threshold in the growth conditions tested, suggesting *cis*-acting RNAs are underestimated. As already observed (Schlüter et al. 2013; Wilms et al. 2011), asRNAs are over-represented in pTiC58 (Fig. 1) and 44% were found to be opposite to genes involved in virulence. Hence, the asRNA regulatory mechanisms might be preferred for these functional categories.

A ratio of 13.8 CDS per *trans*-encoded RNA candidate was obtained in *A. fabrum* C58, in agreement with the previous data (Mitschke et al. 2011; Schlüter et al. 2013). Sixteen could encode peptides, since they presented a putative ORF with significant similarities to the already-predicted proteins in *Rhizobiaceae* species. In addition, the conservation of sRNAs seems to be restricted to related alpha-proteobacterial species, as previously observed in *S. meliloti*. However, the primary nucleotide sequence analyses are probably not sufficient for performing an

accurate conservation search and the use of sRNA classification according to the secondary structure domains might be more relevant. Among the novel and unpredicted *trans*-encoded sRNA candidates, the transcription as *trans*-encoded RNAs of fourteen of them was, nevertheless, validated by RACE-PCR (Table 1), and for at least nine of them (>64%), the absence of CDS and their very probable strong secondary structures suggest that they are true *trans*-encoded RNAs.

Interestingly, one *trans*-encoded RNA, RNA1111, was validated by RACE-PCR and further characterized for its involvement in virulence. A longer transcript, named pTi_205576R and corresponding to *ma1111*, has been previously identified, and in agreement with our data, its expression has previously been shown as not modulated in condition-mimicking virulence (Lee et al. 2013). Nevertheless, we showed a drastic decrease in virulence symptoms on several plants using the strain deleted for *ma1111*, and the complemented strain restored the virulence. This suggested the involvement of RNA1111 in the virulence. However, as the *vir* box of *virE0* is inside the *ma1111* gene (Steck et al. 1988), the deletion also removed this *vir* box and the virulence decrease could have been due to a polar effect on *virE0*. Nonetheless, in *trans* complementation with the sole *ma1111* gene is enough to restore—at least partially—the virulence (Fig. 3). This establishes the actual involvement of RNA1111 in the virulence phenotype. Moreover, in *trans* complementation with the sole *ma1111* gene led to recovery of an expression level of the putative target genes similar to those of the wild type (Fig. 4C). We also noticed that, in addition to the absence of modification in virulence phenotype for the strain overexpressing RNA1111, any significant increase of mRNA target expression levels could be observed in this strain when compared with the wild-type strain. This suggests that, in the wild-type strain, the RNA1111 copy number could already be sufficient to regulate mRNA targets. Recently, Möller et al. (2014) showed an enrichment of several mRNAs by Hfq3xFlag, including *virC2*, *virD3*, and *traA* mRNAs. Interestingly, we presently suspect these mRNAs to be putative RNA1111 targets (Fig. 4), strongly suggesting their interactions with the Hfq protein and their putative regulations by sRNA and, potentially, by RNA1111. Among the predicted RNA1111 mRNA targets, three candidates are connected with virulence but are not essential for tumorigenicity, i.e., two *vir* genes and the 6b gene (*atu6014*). *VirD3* (*Atu6183*) was not tested for the modulation of aggressiveness (Vogel and Das 1992). *VirC2* stimulates the translocation of T-DNA into plant cells, and its DNA-binding function has been shown to be important for bacterial aggressiveness in different plants (Lu et al. 2009). The 6b protein is involved in tumor development, with the accumulation of sugars and phenolic compounds (Clément et al. 2007), and the interaction with nuclear proteins (Kitakura et al. 2002; Terakura et al. 2007). It also modulates auxin transport (Kakiuchi et al. 2006) and the microRNA pathway in plants (Wang et al. 2011). The 6b protein is produced from transformed plant cells, and its regulation in the bacterial background by RNA1111 is surprising. However, the translocation of sRNAs into infected plant cells has been suggested previously, and we might speculate that this is the case for RNA1111. Although we cannot rule out that RNA1111 controls other virulence factors, we can assume that the regulation of all these targets together by RNA1111 could have a synergic impact on virulence. The regions of interaction with RNA1111 are localized upstream or downstream of their putative RBS and could facilitate the binding of ribosomes by modifying mRNA secondary structures and, thus, facilitating the translation. Modifying the interaction between one sRNA and its mRNA target could lead to degradation or the absence of degradation of mRNA (Desnoyers et al. 2013). Based on qRT-PCR data, we suspect that RNA1111 could bind to *virD3* and *virC2*

mRNA and, thus, could stimulate their translation without affecting their stability, whereas the putative binding of RNA1111 on *atu6014* mRNA stabilizes this transcript. However, to definitively validate a direct regulation of RNA1111 on mRNA targets, RNA-RNA interaction experiments are still required. *traA* mRNA has also been shown to be downregulated in the absence of RNA1111. This gene is part of the *tra* operon encoding the type IV secretion system for the conjugative transfer of pTiC58, suggesting that RNA1111 also regulates the conjugation. Noticeably, the virulence and the conjugation of Ti plasmid were actually shown to be regulated in a coordinated way by TraM (Cho and Winans 2005). Our work highlights a potential new level of coordinated regulation of the essential pTi functions by a small *trans*-encoded RNA that is, furthermore, largely conserved among Ti plasmids. Thus, this suggests that RNA1111 and, possibly, sRNAs in general play a major role in the infection of plants and in the dissemination of Ti plasmids.

MATERIALS AND METHODS

Bacterial strains and growth conditions.

Bacterial strains used in this study are the *A. fabrum* C58 wild-type strain and its derivatives, which are listed in Supplementary Table S5. Bacteria were grown with shaking (150 rpm) at 28°C in YPG rich media (per liter: 5 g of yeast extract, 5 g of peptone, 10 g of glucose, pH 7.2) or in AT minimal media (80 mM KH₂PO₄; 0.65 mM MgSO₄; 7 H₂O; 18 µM FeSO₄; 7 H₂O; 70 µM CaCl₂; 2 H₂O; MnCl₂; 10 µM 4 H₂O; pH 7.2) supplemented with 10 mM succinate as the carbon source and 10 mM ammonium sulfate as the nitrogen source (Petit et al. 1978). Conditions mimicking virulence are those previously described by (Lee et al. 2013). Media were supplemented, when needed, with appropriate antibiotics (per milliliter: 25 µg of gentamicin, 25 µg of neomycin, and 25 µg of kanamycin). *Escherichia coli* DH5α strain was used for cloning experiments in classical growth conditions.

Vector constructions.

The Ti plasmid DNA fragment from 205,367 bp to 205,626 bp (corresponding to *rna1111*) was amplified by PCR using primers 186 and 187 (Supplementary Table S6). Fragments were purified and digested using *SalI* and *BamHI* enzymes and were then cloned into a pBBR1MCS-5 plasmid giving rise to pARA005.

Construction of mutant strain deleted for *rna1111*.

The *A. fabrum* C58 strain was used to construct a mutant, as previously described (Lassalle et al. 2011). The genomic region corresponding to transcript RNA1111 was deleted and replaced by the neo-kan resistance gene *nptII*, resulting in the strain AT192.

sRNA selection and abundant sRNA depletions.

Total RNAs were extracted as previously described (Hommais et al. 2008), and DNA contamination was eliminated with a DNase I treatment (Roche). Total RNAs (100 µg) were separated on a 15% denaturing polyacrylamide gel (7 M urea, 1× tris-borate-EDTA), and RNAs with sizes ranging from 25 to 500 nt were extracted (representing 1.4% of total RNAs). Briefly, a piece of gel containing RNAs was broken mechanically and was incubated for 4 h in 500 µl of 10 mM Tris-HCl, 1 mM EDTA, 300 mM NaCl buffer with soft shaking at 15°C. Size-selected RNAs were washed twice and were ethanol-precipitated. 5S RNA and tRNA depletions were performed from 2.5 µg of size-selected RNAs by reverse transcription treatment using the First Strand cDNA synthesis kit (Life Biosciences), with 100 ng of primer 5S and 50 ng of each of the other primers representing the tRNAs. An RNase H treatment was then performed, and newly synthesized DNAs were eliminated with a DNase I

treatment (DNA-free kit; Ambion). Between each step, RNA quantities and qualities were checked, using a Nanodrop 2000 spectrophotometer (Thermo Scientific) and a 2100 bioanalyzer with an RNA 6000 Nano kit (Agilent). RNAs were stored at -80°C until further use.

Library preparation and sequencing.

The library preparation was an adaptation of the Illumina directional mRNA-Seq protocol, using the DGE-Small RNA sample preparation kit. Selected sRNAs (200 ng) were treated with Antarctic phosphatase (New England Biolabs) and T4 PolyNucleotide kinase (New England Biolabs) and were purified on an RNeasy MinElute spin column (Qiagen). Ligations with Illumina v1.5 sRNA 3' and sRNA 5' adapters, reverse transcription, PCR amplification, and purification with Agencourt AMPure beads (Beckman) were carried out as described in the Illumina directional mRNA-Seq protocol (Illumina). One sequence run per library was performed, using the Illumina Genome Analyzer II (Perkins et al. 2009). An average of 24,137,549 reads of 36 bases per run were obtained, corresponding to a total of 193,100 million reads (Supplementary Table S7).

Data analyses.

Raw sequence files were first converted to standard fastq format with Sanger quality encoding, and then, summary statistics of read quality were generated for each sequencing lane, using random samples of 500,000 reads. Read quality of all experiments were found to be sufficiently high for the sequence reads to be directly processed by the mapping software. Reads were mapped on the *A. fabrum* C58 genome (RefSeq accession: NC_003062, NC_003063, NC_003064, and NC_003065) using SSAHA2 software, with the '-diff' parameter set to 0 to ensure that only reads mapping to a single location in the genome were retained (uniquely mapped reads [UMR]) (Ning et al. 2001). Depending on the sequencing run, 95.24 to 99.2% of reads were mapped at least once on the genome. The coverage of UMR (14.71%) along the genome and the number of UMR located in each genomic object (encompassing CDS, rRNA, tRNA, and predicted sRNA) were then computed using R/Bioconductor packages Rsamtools and GenomicRanges (Gentleman et al. 2004) and were visualized in the Agrobacteroscope database (Vallenet et al. 2013). Replicon repartition showed that 72.5% of UMR mapped on the circular chromosome, 25% on the linear one, 1.6% on the pAt, and 0.9% on the pTi. Rockhopper software was used for the detection of novel small transcripts and the differential expression analyses (McClure et al. 2013). The data followed MIAME guidelines and were registered at ArrayExpress under accession number E-MTAB-2947.

Determination of 5' and 3' ends by RACE-PCR.

Phenol/chloroform extraction followed by ethanol precipitation was performed between each step of the following protocol, and enzymatic reactions were performed according to the manufacturer's recommendations. Cleaved monophosphate RNAs—isolated from bacteria grown in rich medium—were eliminated from 12 µg of total DNA-free RNAs, using the XRN-I exoribonuclease (Biolabs). The resulting triphosphate RNAs were treated with tobacco acid pyrophosphatase (Epicentre) and were ligated using T4 RNA ligase (Biolabs). Circular RNAs were reverse transcribed using SuperScript III reverse transcription (First Strand cDNA synthesis; Life Technologies) and primers specific to the gene of interest. Then, the 5'- and 3'-end junction was specifically amplified by PCR. The resulting RACE-PCR products were cloned into pGEMT-Easy (Promega) and were sequenced using M13fwd primer (Beckman Coulter Genomics). An average of five independent clones per RNA was analyzed.

qRT-PCR analysis.

SuperScript III (Life Technologies) was used with random hexamer primers to reverse transcribe 2.5 µg of the total RNAs, according to the manufacturer's protocol (First Strand cDNA synthesis). The DyNAmo Flash SYBR green qPCR kit (Thermo Scientific) was used with 0.1 µl of the cDNA reaction mixture and gene-specific primers. The thermal cycling reactions were performed using the LC480 Lightcycler from Roche according to the following conditions: an initial step at 95°C for 7 min, followed by 45 cycles at 95°C for 10 s, 58°C for 15 s, and 72°C for 20 s. Normalization of data were performed using *pAW* and genes *atu0231* and *atu1924*, as their expressions are constant in the growth conditions tested. The specificity of the PCR primers was verified with a melting curve analysis.

Virulence assays on plants.

The plants inoculated with *A. fabrum* C58 were three-week-old tomatoes (Marmande) grown in controlled environment chambers with 70% hygrometry and a 16-h photoperiod at 25°C followed by 8 h at 22°C. Virulence assays were performed as previously described (Haudecoeur et al. 2009). Briefly, plants were wounded at the top of the stem just under the second leaves, and were infected with around 10⁶ bacterial cells. Tumors were observed and counted 3 weeks postinfection. Ten plants were used for each strain, and each plant was classified according to severity symptom categories. Each virulence assay was performed in three independent experiments. Kruskal-Wallis statistic tests were applied to determine significant differences between the virulence assays. These assays were also performed on 6- to 8-week-old *Nicotiana tabacum* burley and 2-month-old *Kalanchoe daigemontiana*, and symptoms were evaluated 5 weeks and 3 to 4 weeks postinfection, respectively.

Prediction of bacterial sRNA targets.

mRNA target genes of sRNAs were predicted from the *A. fabrum* C58 genome (accession number NC_003062, NC_003063, NC_003064, and NC_003065) using IntaRNA (Busch et al. 2008), sTarPicker (Ying et al. 2011), and RNApredator (Eggenhofer et al. 2011) algorithms. We arbitrarily selected the 350 most significant mRNA targets provided by each algorithm. The mRNA targets found by the three algorithms were chosen for further analyses.

Secondary structure predictions and similarity searches.

RNA secondary structures were predicted using the RNAfold algorithm. Basic options were chosen for folding. Sequence similarity searches were performed with the BLAST algorithm on the MAGE database (Vallenet et al. 2013).

ACKNOWLEDGMENTS

This study was supported by the EcoGenome project of the French Agence Nationale de la Recherche [grant number ANR-BLAN-08-0090], PEPS INSB and PEPII INSB-INEE-INSMI-INS2I "Bio-Math-Info" funding from the Centre National de la Recherche Scientifique. M. Dequivre received a doctoral grant from the French Ministère de l'Éducation nationale de l'Enseignement Supérieur et de la Recherche. The authors thank J. Baude, T. Campillo, and N. Poussereau from UMR5240, C. Oger from the Rhône-Alpes Bioinformatics Center, and the LABGeM/MicroScope for skillful technical assistance. We are grateful for the stimulating discussions with and comments from W. Nasser, S. Reverchon and C. Lavire.

LITERATURE CITED

Backofen, R., Amman, F., Costa, F., Findeiß, S., Richter, A. S., and Stadler, P. F. 2014. Bioinformatics of prokaryotic RNAs. *RNA Biol.* 11:470-483.

Bastet, L., Dubé, A., Massé, E., and Lafontaine, D. A. 2011. New insights into riboswitch regulation mechanisms. *Mol. Microbiol.* 80:1148-1154.

Becker, A., Overlöpfer, A., Schlüter, J. P., Reinkensmeier, J., Robledo, M., Giegerich, R., Narberhaus, F., Evgueniev-Hackenberg, E., and Narberhaus, F. 2014. Riboregulation in plant-associated α -proteobacteria. *RNA Biol.* 11:550-562.

Behrens, S., Widder, S., Mannala, G. K., Qing, X., Madhugiri, R., Kefer, N., Abu Mraheil, M., Rattei, T., and Hain, T. 2014. Ultra deep sequencing of *Listeria monocytogenes* sRNA transcriptome revealed new antisense RNAs. *PLoS ONE* 9:e83979.

Busch, A., Richter, A. S., and Backofen, R. 2008. IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24:2849-2856.

Chai, Y., and Winans, S. C. 2005. A small antisense RNA downregulates expression of an essential replicase protein of an *Agrobacterium tumefaciens* Ti plasmid. *Mol. Microbiol.* 56:1574-1585.

Cho, H., and Winans, S. C. 2005. VirA and VirG activate the Ti plasmid repABC operon, elevating plasmid copy number in response to wound-released chemical signals. *Proc. Natl. Acad. Sci. U.S.A.* 102:14843-14848.

Clément, B., Perot, J., Geoffroy, P., Legrand, M., Zon, J., and Otten, L. 2007. Abnormal accumulation of sugars and phenolics in tobacco roots expressing the *Agrobacterium* T-6b oncogene and the role of these compounds in 6b-induced growth. *Mol. Plant Microbe Interact.* 20:53-62.

del Val, C., Rivas, E., Torres-Quesada, O., Toro, N., and Jiménez-Zurdo, J. I. 2007. Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. *Mol. Microbiol.* 66:1080-1091.

Desnoyers, G., Bouchard, M. P., and Massé, E. 2013. New insights into small RNA-dependent translational regulation in prokaryotes. *Trends Genet.* 29:92-98.

Eggenhofer, F., Tafer, H., Stadler, P. F., and Hofacker, I. L. 2011. RNApredator: Fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res.* 39:W149-W154.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80.

Gómez-Lozano, M., Marvig, R. L., Molin, S., and Long, K. S. 2012. Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. *Environ. Microbiol.* 14:2006-2016.

Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B. S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gursen, J., Lomo, C., Sear, C., Strub, G., Cielo, C., and Slater, S. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* 294:2323-2328.

Haudecoeur, E., Tannières, M., Cirou, A., Raffoux, A., Dessaux, Y., and Faure, D. 2009. Different regulation and roles of lactonases AiiB and AttM in *Agrobacterium tumefaciens* C58. *Mol. Plant Microbe Interact.* 22:529-537.

Hommais, F., Oger-Desfeux, C., Van Gijsegem, F., Castang, S., Ligori, S., Expert, D., Nasser, W., and Reverchon, S. 2008. PecS is a global regulator of the symptomatic phase in the phytopathogenic bacterium *Erwinia chrysanthemi* 3937. *J. Bacteriol.* 190:7508-7522.

Hui, M. P., Foley, P. L., and Belasco, J. G. 2014. Messenger RNA degradation in bacterial cells. *Annu. Rev. Genet.* 48:537-559.

Kakiuchi, Y., Gális, I., Tamogami, S., and Wabiko, H. 2006. Reduction of polar auxin transport in tobacco by the tumorigenic *Agrobacterium tumefaciens* AK-6b gene. *Planta* 223:237-247.

Kawano, M., Storz, G., Rao, B. S., Rosner, J. L., and Martin, R. G. 2005. Detection of low-level promoter activity within open reading frame sequences of *Escherichia coli*. *Nucleic Acids Res.* 33:6268-6276.

Kitakura, S., Fujita, T., Ueno, Y., Terakura, S., Wabiko, H., and Machida, Y. 2002. The protein encoded by oncogene 6b from *Agrobacterium tumefaciens* interacts with a nuclear protein of tobacco. *Plant Cell* 14:451-463.

Lassalle, F., Campillo, T., Vial, L., Baude, J., Costechareyre, D., Chapulliot, D., Shams, M., Abrouk, D., Lavire, C., Oger-Desfeux, C., Hommais, F., Guéguen, L., Daubin, V., Muller, D., and Nesme, X. 2011. Genomic species are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. *Genome Biol. Evol.* 3:762-781.

Lee, K., Huang, X., Yang, C., Lee, D., Ho, V., Nobuta, K., Fan, J. B., and Wang, K. 2013. A genome-wide survey of highly expressed non-coding RNAs and biological validation of selected candidates in *Agrobacterium tumefaciens*. *PLoS ONE* 8:e70720.

Li, W., Ying, X., Lu, Q., and Chen, L. 2012. Predicting sRNAs and their targets in bacteria. *Genomics Proteomics Bioinformatics* 10:276-284.

Liu, J. M., Livny, J., Lawrence, M. S., Kimball, M. D., Waldor, M. K., and Camilli, A. 2009. Experimental discovery of sRNAs in *Vibrio cholerae*

- by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.* 37:e46.
- Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. 2009. A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* 139:770-779.
- Lu, J., den Dulk-Ras, A., Hooykaas, P. J., and Glover, J. N. 2009. *Agrobacterium tumefaciens* VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold. *Proc. Natl. Acad. Sci. U.S.A.* 106:9643-9648.
- McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., Vanderpool, C. K., and Tjaden, B. 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* 41:e140.
- Mentz, A., Neshat, A., Pfeifer-Sancar, K., Pühler, A., Rückert, C., and Kalinowski, J. 2013. Comprehensive discovery and characterization of small RNAs in *Corynebacterium glutamicum* ATCC 13032. *BMC Genomics* 14:714.
- Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J., and Hess, W. R. 2011. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* 108:2124-2129.
- Möller, P., Overlöpfer, A., Förstner, K. U., Wen, T. N., Sharma, C. M., Lai, E. M., and Narberhaus, F. 2014. Profound impact of Hfq on nutrient acquisition, metabolism and motility in the plant pathogen *Agrobacterium tumefaciens*. *PLoS ONE* 9:e110427.
- Mousavi, S. A., Osterman, J., Wahlberg, N., Nesme, X., Lavire, C., Vial, L., Paulin, L., de Lajudie, P., and Lindström, K. 2014. Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. *Syst. Appl. Microbiol.* 37:208-215.
- Ning, Z., Cox, A. J., and Mullikin, J. C. 2001. SSAHA: A fast search method for large DNA databases. *Genome Res.* 11:1725-1729.
- Overlöpfer, A., Kraus, A., Gurski, R., Wright, P. R., Georg, J., Hess, W. R., and Narberhaus, F. 2014. Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region. *RNA Biol.* 11:624-640.
- Perkins, T. T., Kingsley, R. A., Fookes, M. C., Gardner, P. P., James, K. D., Yu, L., Assefa, S. A., He, M., Croucher, N. J., Pickard, D. J., Maskell, D. J., Parkhill, J., Choudhary, J., Thomson, N. R., and Dougan, G. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet.* 5:e1000569.
- Petit, A., Tempe, J., Kerr, A., Holsters, M., Vanmontagu, M., and Schell, J. 1978. Substrate induction of conjugative activity of *Agrobacterium tumefaciens* ti plasmids. *Nature* 271:570-572.
- Pitzschke, A., and Hirt, H. 2010. New insights into an old story: *Agrobacterium*-induced tumour formation in plants by plant transformation. *EMBO (Eur. Mol. Biol. Organ.) J.* 29:1021-1032.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. 2000. Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944-945.
- Schlüter, J. P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Jänicke, S., Becker, J. D., Giegerich, R., and Becker, A. 2010. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics* 11:245.
- Schlüter, J. P., Reinkensmeier, J., Barnett, M. J., Lang, C., Krol, E., Giegerich, R., Long, S. R., and Becker, A. 2013. Global mapping of transcription start sites and promoter motifs in the symbiotic α -proteobacterium *Sinorhizobium meliloti* 1021. *BMC Genomics* 14:156.
- Schmidtke, C., Findeiss, S., Sharma, C. M., Kuhfuss, J., Hoffmann, S., Vogel, J., Stadler, P. F., and Bonas, U. 2012. Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. *Nucleic Acids Res.* 40:2020-2031.
- Soutourina, O. A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., Le Bouguenec, C., Coppée, J. Y., Dupuy, B., and Martin-Verstraete, I. 2013. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genet.* 9:e1003493.
- Steck, T. R., Morel, P., and Kado, C. I. 1988. Vir box sequences in *Agrobacterium tumefaciens* pTiC58 and A6. *Nucleic Acids Res.* 16:8736.
- Storz, G., Vogel, J., and Wassarman, K. M. 2011. Regulation by small RNAs in bacteria: Expanding frontiers. *Mol. Cell* 43:880-891.
- Terakura, S., Ueno, Y., Tagami, H., Kitakura, S., Machida, C., Wabiko, H., Aiba, H., Otten, L., Tsukagoshi, H., Nakamura, K., and Machida, Y. 2007. An oncoprotein from the plant pathogen *Agrobacterium* has histone chaperone-like activity. *Plant Cell* 19:2855-2865.
- Ulvé, V. M., Sevin, E. W., Chéron, A., and Barloy-Hubler, F. 2007. Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in *Sinorhizobium meliloti* strain 1021. *BMC Genomics* 8:467.
- Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fèvre, F., Longin, C., Mornico, D., Roche, D., Rouy, Z., Salvignol, G., Scarpelli, C., Thil Smith, A. A., Weiman, M., and Médigue, C. 2013. MicroScope—An integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* 41:D636-D647.
- Valverde, C., Livny, J., Schlüter, J. P., Reinkensmeier, J., Becker, A., and Parisi, G. 2008. Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 1021. *BMC Genomics* 9:416.
- Vogel, A. M., and Das, A. 1992. The *Agrobacterium tumefaciens* virD3 gene is not essential for tumorigenicity on plants. *J. Bacteriol.* 174:5161-5164.
- Wang, M., Soyano, T., Machida, S., Yang, J. Y., Jung, C., Chua, N. H., and Yuan, Y. A. 2011. Molecular insights into plant cell proliferation disturbance by *Agrobacterium* protein 6b. *Genes Dev.* 25:64-76.
- Wilms, I., Voss, B., Hess, W. R., Leichert, L. I., and Narberhaus, F. 2011. Small RNA-mediated control of the *Agrobacterium tumefaciens* GABA binding protein. *Mol. Microbiol.* 80:492-506.
- Wilms, I., Overlöpfer, A., Nowrousian, M., Sharma, C. M., and Narberhaus, F. 2012a. Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. *RNA Biol.* 9:446-457.
- Wilms, I., Möller, P., Stock, A. M., Gurski, R., Lai, E. M., and Narberhaus, F. 2012b. Hfq influences multiple transport systems and virulence in the plant pathogen *Agrobacterium tumefaciens*. *J. Bacteriol.* 194:5209-5217.
- Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G. E., Almeida, N. F., Jr., Woo, L., Chen, Y., Paulsen, I. T., Eisen, J. A., Karp, P. D., Bovee, D., Sr., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutayavin, T., Levy, R., Li, M. J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z. Y., Dolan, M., Chumley, F., Tingey, S. V., Tomb, J. F., Gordon, M. P., Olson, M. V., and Nester, E. W. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317-2323.
- Ying, X., Cao, Y., Wu, J., Liu, Q., Cha, L., and Li, W. 2011. sTarPicker: A method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. *PLoS ONE* 6:e22705.
- Zeng, Q., and Sundin, G. W. 2014. Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen *Erwinia amylovora* discovered small RNAs with virulence regulatory function. *BMC Genomics* 15:414.

AUTHOR-RECOMMENDED INTERNET RESOURCES

- European Molecular Biology Laboratory-European Bioinformatics Institute
Rfam database: rfam.xfam.org
IntaRNA server: rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp
Agrobacterscope database:
www.genoscope.cns.fr/agc/microscope/home/index.php
National Center for Biotechnology Information ORFfinder database:
www.ncbi.nlm.nih.gov/gorf/gorf.html
RNAfold web server: rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi
RNApredator target prediction webserver:
rna.tbi.univie.ac.at/RNApredator2/target_search.cgi
sTarPicker prediction server: ccb.bmi.ac.cn/starpicker/prediction.php