Supplementary Materials for

Bacterial antisense RNAs are mainly the product of transcriptional noise

Verónica Lloréns-Rico, Jaime Cano, Tjerko Kamminga, Rosario Gil, Amparo Latorre, Wei-Hua Chen, Peer Bork, John I. Glass, Luis Serrano, Maria Lluch-Senar

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at advances.sciencemag.org/cgi/content/full/2/3/e1501363/DC1)

Table S6 (Microsoft Excel format). Shotgun proteomics results of the whole proteome of the nine clones of *M. pneumoniae* overexpressing asRNAs.
Table S7 (Microsoft Excel format). RNA-seq results of the whole transcriptome of the nine clones of *M. pneumoniae* overexpressing asRNAs.

Supplementary table descriptions:

Table S1: sRNA annotation of *B. aphidicola*. There are columns for the sRNA strand, unique ID, Start and Stop positions.

Table S2: sRNA annotation of *M. hyopneumoniae*. There are columns for the sRNA strand, unique ID, Start and Stop positions.

Table S3: sRNA annotation of *M. mycoides*. There are columns for the sRNA strand, unique ID, Start and Stop positions.

Table S4: Bacterial strains used in this study. Data for each of the bacterial species used in this study is presented. Information shown includes genome size, number of ORFs, number of total sRNAs (only asRNAs and trans-encoded sRNAs; cis-encoded sense sRNAs are not included because their transcription levels are difficult to distinguish from those derived of overlapping mRNA), number of trans-encoded sRNAs, number of asRNAs, genomic AT content and data source(*43, 49-62*). For bacteria with more than one replicon, only the largest one was considered. In *E. coli* and in the *A. thaliana* chloroplast, trans-encoded sRNAs were not reported in the source publications.

Table S4: Bacterial strains used in this study

Species	Genome size (bps)	ORFs	Total sRNAs	Trans-encoded sRNAs	asRNAs	AT content	Source
Buchnera aphidicola BCc	416380	397	364	30	334	79.8	*De novo* annotation
Mycoplasma hyopneumoniae 11	892758	727	629	50	579	71.4	*De novo* annotation
Synthetic Mycoplasma mycoides JCVI-syn1.0	1078809	910	1043	37	1006	76.1	*De novo* annotation
Mycoplasma genitalium MG137	580076	547	473	25	448	68.3	Manuscript in preparation
Mycoplasma pneumoniae M129	816394	780	251	47	204	60	Publication[22]
Escherichia coli K12 MG1655	4641652	4607	1005	-	1005	49.2	Publication[43]
Mycobacterium tuberculosis H37Rv	4411532	4112	456	21	435	34.4	Publication[49]
Arabidopsis thaliana chloroplast	154478	173	107	53	54	63.7	Publication[50]
Bacillus subtilis subsp. subtilis 168	4215606	4421	1583	1123	460	56.5	Publication[51]
Salmonella typhimurium SL1344	4857432	4631	629	160	469	47.8	Publication[52]
Pseudomonas aeruginosa PAO1	6264404	5682	743	510	233	33.4	Publication[53,54]
Helicobacter pylori 26695	1667867	1561	803	78	725	61.1	Publication[55]
Streptomyces avermitilis MA-4680	9025608	7670	712	176	536	29.3	Publication[56]
Streptomyces coelicolor M145	8667507	7910	785	105	680	27.9	Publication[56]
Streptomyces venezuelae 10712	8226158	7536	791	199	592	27.6	Publication[56]
Vibrio cholerae El Tor O1	3961149	2690	519	412	107	52.3	Publication[57]
Corynebacterium glutamicum 13032	3282708	3138	730	43	687	46.2	Publication[58]
Campylobacter jejuni RM1221	1777831	1940	1424	27	1397	69.7	Publication[59]
Sinorhizobium meliloti 2011	3657276	4448	459	356	103	37.3	Publication[60]
Synechocystis sp. PCC 6803	3573470	3229	1391	629	762	52.3	Publication[61]
Agrobacterium tumefaciens C58	2841490	2789	221	65	156	40.6	Publication[62]

Table S5: asRNAs overexpressed in *M. pneumoni*ae. The table indicates the overlapping gene of each of the asRNAs, region of overlap, and expression levels for the asRNAs in wild-type control conditions and in the overexpression of the asRNAs.

Table S5: overexpressed asRNAs and effect in their overlapping genes

asRNA	Overlapping mRNA	Overlapping region	CONTROL: asRNA expression (log2(CPKM))	OVEREXPRESSION: asRNA expression (log2(CPKM))	Mean change in asRNA expression
ncMPN090	MPN492	3'	3.574	6.921	3.347
ncMPN098	MPN545	Centre	1.054	6.677	5.623
ncMPN007	MPN050	3'	3.322	7.582	4.260
ncMPN230	MPN055	3'	1.648	5.261	3.613
ncMPN230	MPN056	5'	1.648	5.261	3.613
ncMPN283	MPN211	Centre	1.158	7.354	6.196
ncMPN289	MPN247	5'	0.795	5.201	4.406
ncMPN306	MPN305	3'	2.826	7.562	4.736
ncMPN312	MPN310	Centre	3.717	7.364	3.647
ncMPN323	MPN342	Centre	3.309	6.757	3.448

Table S6: Shotgun proteomics results of the whole proteome of the nine clones of *M. pneumoniae* overexpressing asRNAs. Table indicates the fold-changes and multiple-test corrected p-values of the protein levels with respect to the wild-type conditions.

Table S7: RNA-seq results of the whole transcriptome of the nine clones of *M. pneumoniae* overexpressing asRNAs. Table indicates the fold-changes and multiple-test corrected p-values of the RNA levels with respect to the wild-type conditions.

Table S8: Parameters and initial conditions used in the simulations of the asRNA effects. The majority of the parameters were determined experimentally specifically for the modeled bacterium *M. pneumoniae*. The following parameters were extracted from previous publications: *M. pneumoniae* volume (*63*), binding and degradation rate for the duplex (*64*), dissociation rate of the duplex (*65*), binding rate of the ribosome (*42*), protein translation and degradation rate (*42*), initial protein and ribosome copy number (*42*). The parameter β (binding rate mRNA-ribosome) was adapted from the publication to match the different copy numbers observed in our experiments. The mRNA-asRNA duplex decay rate was set to be differentially smaller than the single RNA decay rates. The rest of the parameters were determined using experimental data (see Methods).

Table S8: Parameters and initial conditions used for the simulations

Parameter	Meaning	Value	Units	Source
Vol	*M. pneumoniae* volume	5.50e-17	L	Publication[63]
α_m	mRNA transcription rate	3.95e-10 - 3.83e-9	M/min	Experimental
k_m	mRNA degradation rate	0.127	1/min	Experimental
k_{on}	binding/degradation rate of the duplex	0 - 6e+7	1/M*min	Publication[64]
k_{off}	dissociation rate of the duplex	0.84	1/min	Publication[65]
β	bindinf rate of mRNA-ribosome	1.34e-4	1/M*min	Publication[42]
α_s	asRNA transcription rate	2.33e-10 - 1.31e-9	M/min	Experimental
k_s	asRNA degradation rate	0.137	1/min	Experimental
α_p	protein translation rate	1.2	1/min	Publication[42]
k_p	protein degradation rate	5.02e-5	1/min	Publication[42]
k_{dup}	duplex degradation rate	6e-2	1/min	-
m_0	initial mRNA copy number	0	-	-
s_0	initial asRNA copy number	0	-	-
p_0	initial protein copy number	167	-	Publication[42]
rib_0	initial ribosome copy number	200	-	Publication[42]
$mrib_0$	initial mRNA-ribosome complex number	0	-	-
dup_0	initial mRNA-asRNA duplex	0	-	-

Table S9: Primers used in this study to clone the asRNAs. Forward and reverse primer sequences are indicated.
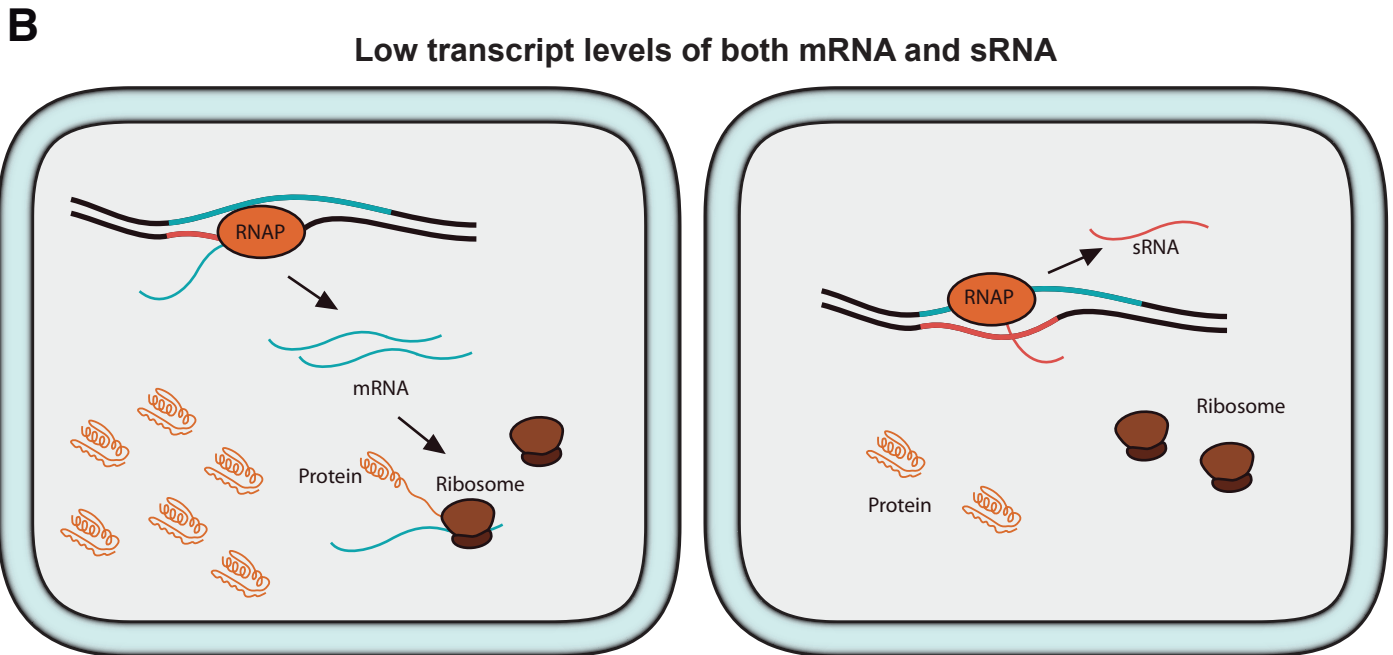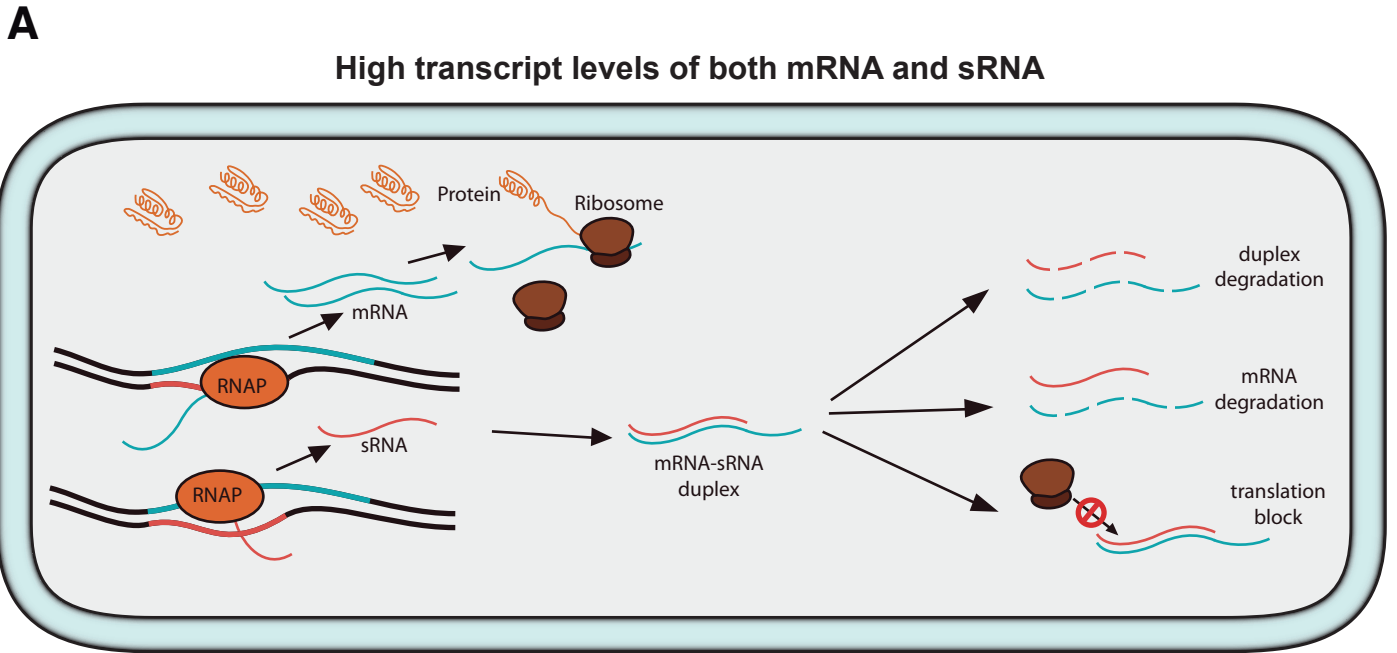
Table S9: Primers used in this study to clone asRNAs

asRNA	Primer Forward
ncMPN007-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTTGATTTAAAAAAATTTTTTTCTCT
ncMPN090-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTTGTTGGTACATACTGGTGAAT
ncMPN098-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTTCTTCGACAAATTGATTCGCT
ncMPN230-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTACATGAAGGGGGTGAACAGGTAGA
ncMPN283-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTCATCAGCGGGACCTAACTTAA
ncMPN289-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTACTAAGGTAGTCGCCATTCGC
ncMPN306-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTAAAGAAAAATTAAAGTGGTTT
ncMPN312-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTAACCGCTTCCTTGCGTCTTTG
ncMPN323-Fwd	TATAATTGTGTAAAAGGGCCCCCCCCTCGAGTAGTATTTAGAATTAATAAAGTTCAATAGAGTGATTCAAAACC
ncMPN007-Rev	ATACTTTATTAATTCTAAATACTAGACTGGAATCGCTAAGCAGTTCA
ncMPN090-Rev	ATACTTTATTAATTCTAAATACTAGACACAGAAACAGAACGCAAT
ncMPN098-Rev	TTATTAATTCTAAATACTAGTTAGGTAAAGGAATGCCGTATAC
ncMPN230-Rev	ATACTTTATTAATTCTAAATACTAGCAAAAGATTCACCAACAAAG
ncMPN283-Rev	ATACTTTATTAATTCTAAATACTAGTTAAAAACGCAGTTAATTCA
ncMPN289-Rev	ATACTTTATTAATTCTAAATACTAGTTTTTACCAAGTCCTTTACA
ncMPN306-Rev	ATACTTTATTAATTCTAAATACTAGTTGAAACCCACTTTGATGGC
ncMPN312-Rev	ATACTTTATTAATTCTAAATACTAGCTGCAACACAAGTATGCTAA
ncMPN323-Rev	ATACTTTATTAATTCTAAATACTAGTTAATAGCAAAAAAGAAGAA

Supplementary figure descriptions:

Figure S1. Different regulatory mechanisms of sRNAs. A) sRNAs usually act via complementary base pairing with an mRNA. This base pairing can result in altering the stability of the duplex, via degrading it or stabilizing the complex. It also can result in the degradation of the mRNA only, if the sRNA has some catalytic activity (alone or binding a protein, as it occurs in the eukaryotic RISC system. Finally, the sRNA can alter the accessibility of the ribosome-binding site of the mRNA, altering translation. For these mechanisms to occur, both mRNA and sRNA need to be present simultaneously at the same cell. B) In cases in which the transcript levels are low, the probability of the mRNA and sRNA to be simultaneously present at one cell will be very small and the pairing will be unlikely to occur in a large percentage of the population cells. Therefore, some cells will have few copies of the mRNA and others will have few copies of the sRNA, thus at the population level no effect will be observed.
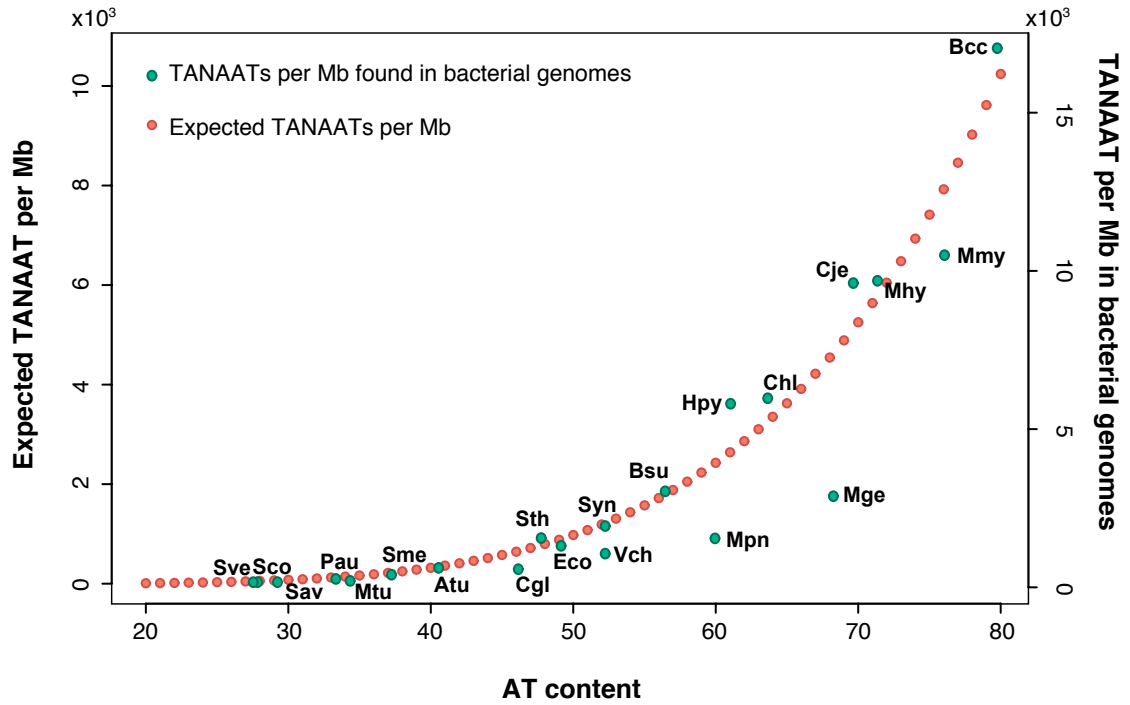
Figure S1

A

High transcript levels of both mRNA and sRNA

Protein
Ribosome
mRNA
RNAP
sRNA
RNAP
mRNA-sRNA duplex
duplex degradation
mRNA degradation
translation block

B

Low transcript levels of both mRNA and sRNA

RNAP
mRNA
Protein
Ribosome

RNAP
sRNA
Ribosome
Protein

Figure S2: Theoretical and real numbers of TANAAT boxes in bacteria. Genomes represented: Atu: *Agrobacterium tumefasciens*; Bcc: *Buchnera aphidicola* (str Cc); Bsu: *Bacillus subtilis*; Cgl: *Corynebacterium glutamicum*; Chl: Chloroplast (*Arabidopsis thaliana*); Cje: *Campylobacter jejuni*; Eco: *Esherichia coli*; Hpy: *Helicobacter pylori*; Mge: *Mycoplasma genitalium*; Mhy: *Mycoplasma hyopneumoniae*; Mmy: *Mycoplasma mycoides*; Mpn: *Mycoplasma pneumoniae*; Mtu: *Mycobacterium tuberculosis*; Pau: *Pseudomonas aeruginosa*; Sav: *Streptomyces avermitilis*; Sco: *Streptomyces coelicolor*; Sme: *Sinorhizobium meliloti*; Sth: *Salmonella typhimurium*; Sve: *Streptomyces venezuelae*; Syn: *Synechocystis* spp, Vch: *Vibrio cholerae*. A) Theoretical and real numbers of TANAAT boxes expected from a given AT content. The theoretical trend is similar to the one observed in bacteria, but the expected TANAAT boxes outnumber the detected ones in different genomes. This may be due to a joint effect of negative selection and the natural constraints imposed by protein-coding sequences. B) Proportion of ORFs and sRNAs that have a canonical TANAAT box upstream. We considered the transcriptional start sites in 4 species for which the data was available (*M. genitalium* (Chen et al, manuscript in preparation), *M. pneumoniae* (*22*), *M. tuberculosis* (*66*) and *E. coli* (*67*)), and the translation start sites for the rest (Fig. S2b).
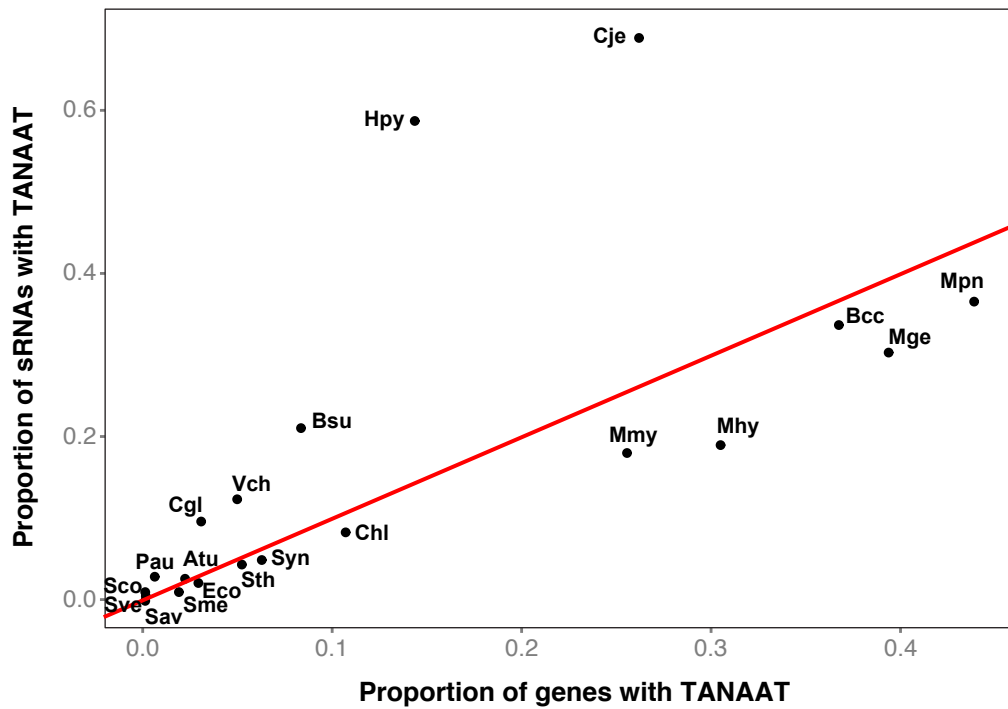
Red line represents slope 1 in the graph. In the majority of species, this proportion is similar. However, in some of them, this proportion is much higher for sRNAs. This could be due to a less strong regulation for the transcription of sRNAs.
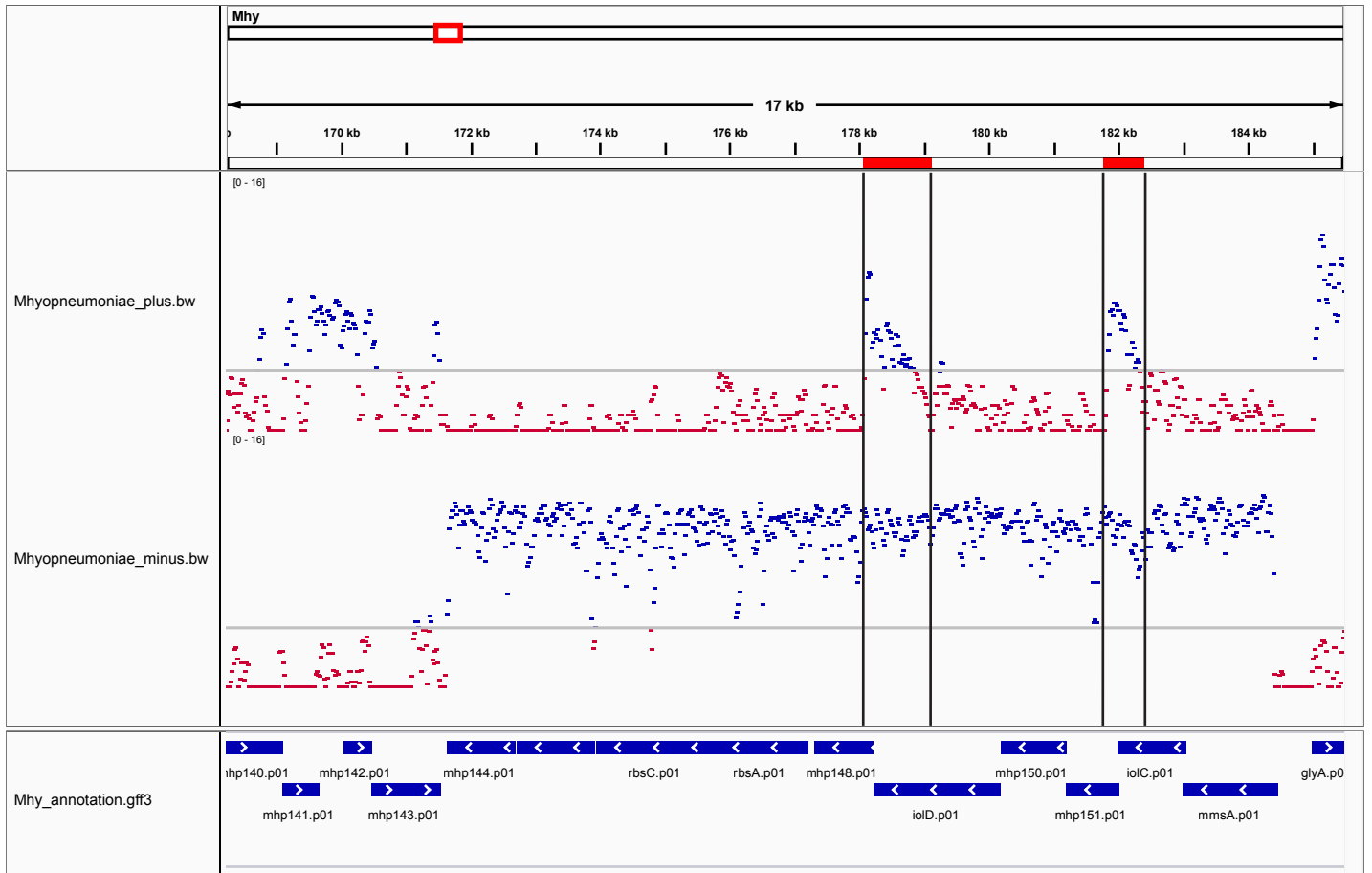
Figure S2

A

B

Figure S3: Manual annotation of sRNAs in *M. hyopneumoniae*. Expression data from bacteria whose sRNAs were *de novo* annotated was visualized on the Integrative Genomics Viewer (IGV, (*41*)). Pileups of the RNA-seq data were converted to IGV tracks and visualized to perform manual annotation. The figure shows two data tracks, for the RNA expression of plus and minus strands; and an annotation track with the gene annotation of *M. hyopneumoniae*. A threshold was calculated for each bacterium, based on the lower 5% quantile of protein-coding gene expression. This threshold is represented by a grey horizontal line in both data tracks. Transcripts whose expression was below this threshold were not considered as experimental noise, and thus not annotated as sRNAs. Points below this threshold are represented in red, whilst points above are depicted in blue. Vertical black lines delimit two asRNAs in the plus strand overlapping a polycistronic transcript in the minus strand; these were annotated as MHYs_63 and MHYs_64.
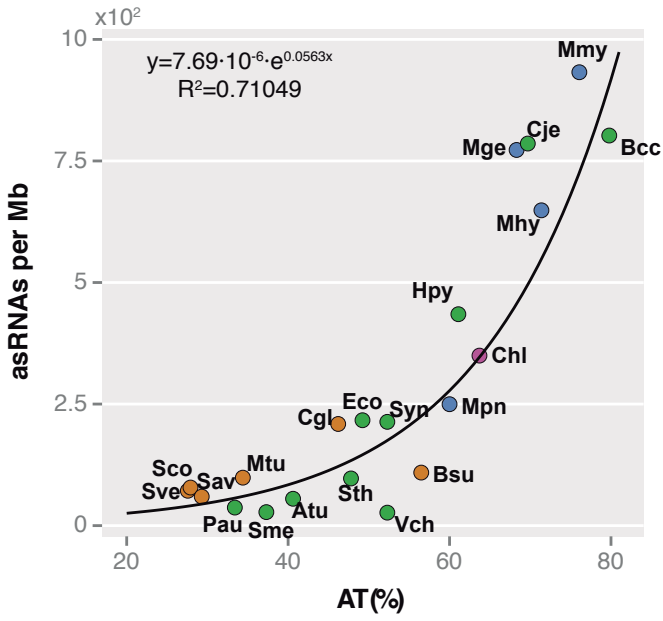
Figure S3

Figure S4: Dependency on the AT content of different types of sRNAs. A) asRNAs, normalized by genome size, versus AT content. Similarly to the number of total sRNAs, the number of asRNAs in the different bacteria shows an exponential dependency on the AT content ($R^2=0.71$), whilst it does not correlate with genome size (R=0.08). B) Trans-encoded sRNAs, normalized by genome size, versus AT content. Trans-encoded sRNAs (intergenic) were compared normalizing by the intergenic genome size of the different species, and considering the AT content of these intergenic regions. These sRNAs show a trend more similar to that of ORFs. Their number is independent of the AT content, but it does not correlate with the intergenic genome size (R=0.23).
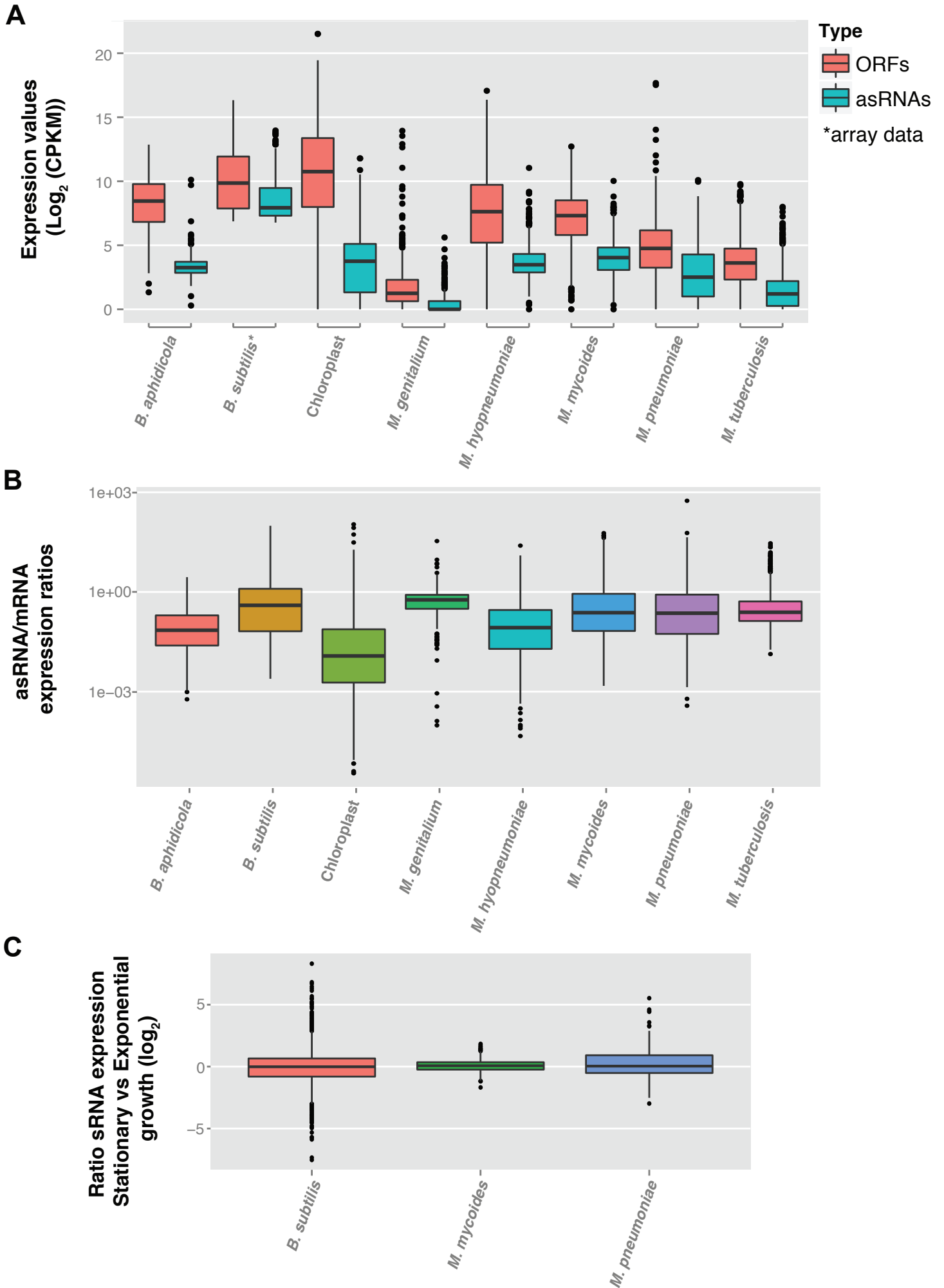
Figure S4
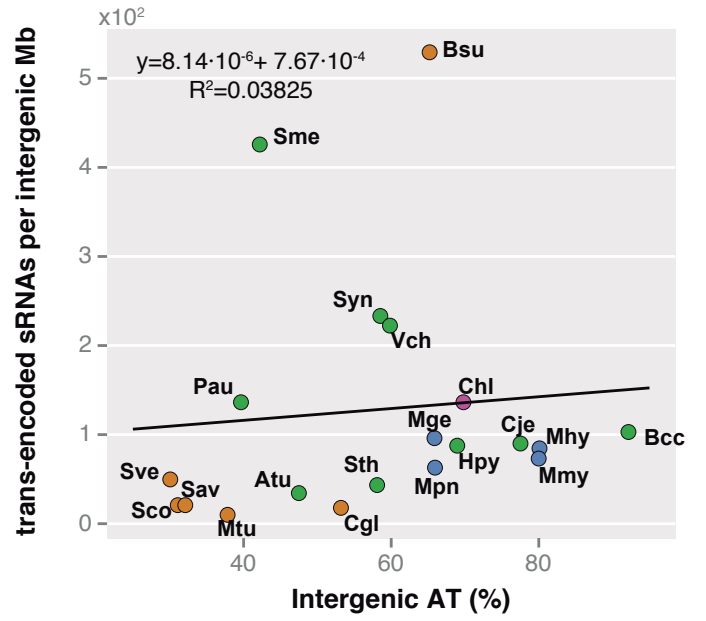
Figure S5: Transcript levels of asRNAs and mRNAs in different bacteria. A) In 7 bacterial species (*B. aphidicola, M. genitalium, M. hyopneumoniae, M. mycoides* and *M. pneumoniae, A. thaliana* chloroplast(*50*) and *M. tuberculosis*(*49*)), transcript abundance was characterized by RNA-seq, and in *B. subtilis*(*51*), it was determined using expression arrays. All bacterial species but *B. aphidicola* were studied in exponential growth conditions. Transcript abundances of all samples are shown as Tukey box-and-whisker plots. In all samples analyzed, asRNAs showed expression values that are on average lower than those of mRNAs. B) Expression ratios of overlapping asRNA-mRNA. Expression ratios were calculated by dividing the expression of each asRNA by the expression of its overlapping mRNA, for the same bacteria as in panel A). In the majority of cases, the ratio is below 1. C) Expression of asRNAs in exponential and stationary phases. For three bacteria in our study (*B. subtilis (51)*, *M. mycoides* and *M. pneumoniae*) expression was compared at exponential and stationary phase. Box-and-whiskers plots show that there is not a global accumulation of antisense transcripts throughout the growth of these bacteria.
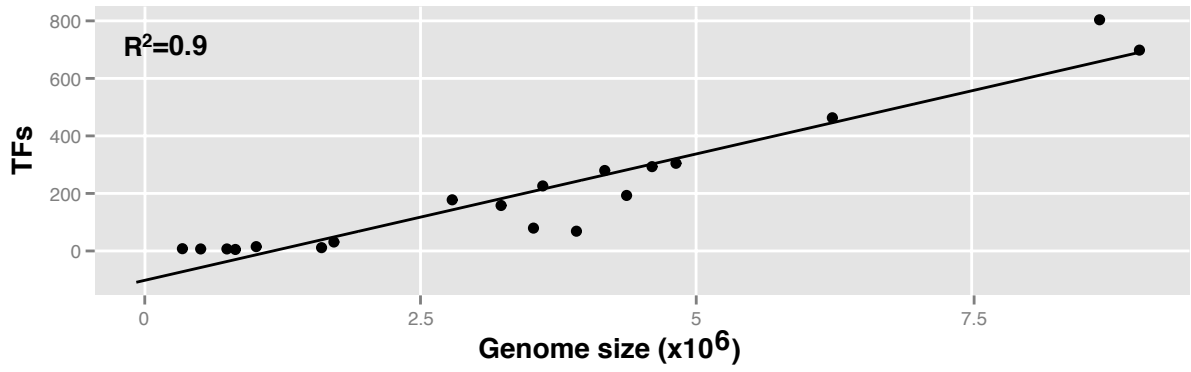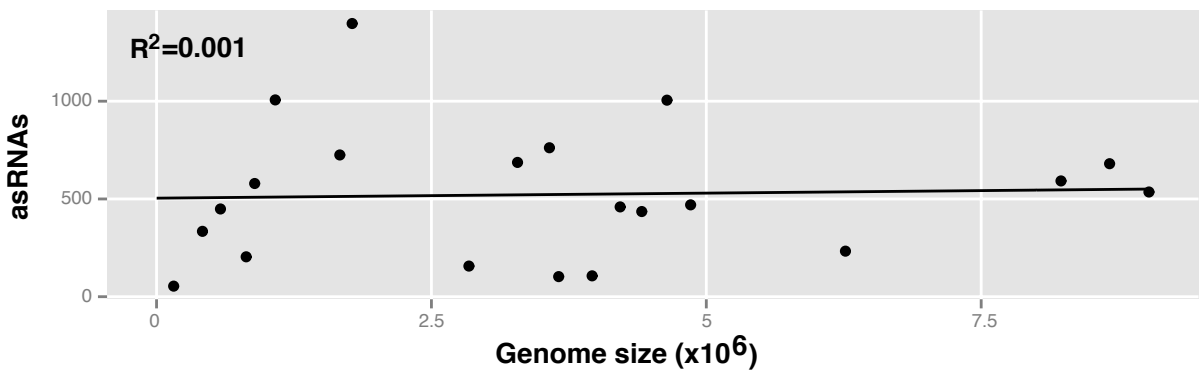
Figure S5

Figure S6: Relationship between asRNAs and transcription factors in bacteria A) Number of transcription factors versus genome size. There is a linear relationship between the number of transcription factors and genome size in bacteria: R^2=0.9. B) Number of asRNAs versus genome size. The number of asRNAs does not correlate to genome size R^2=0.001, as it is dependent on the AT content of bacterial genomes. C) Relationship between number of transcription factors and number of asRNAs in bacteria. We compared the number of asRNAs and transcription factors in bacteria with similar AT content (Low AT content: ≤40%; Mid AT content: 40-60%; High AT content: >60%). A negative correlation in between the two was not observed in any case. In fact, an unexpected positive correlation was found for high AT genomes (R=0.94). D) Relationship between number of genes and number of asRNAs in bacteria with similar AT content. For bacteria with similar AT content, the number of asRNAs correlates with the number of genes, which is an indicative of the genome size.
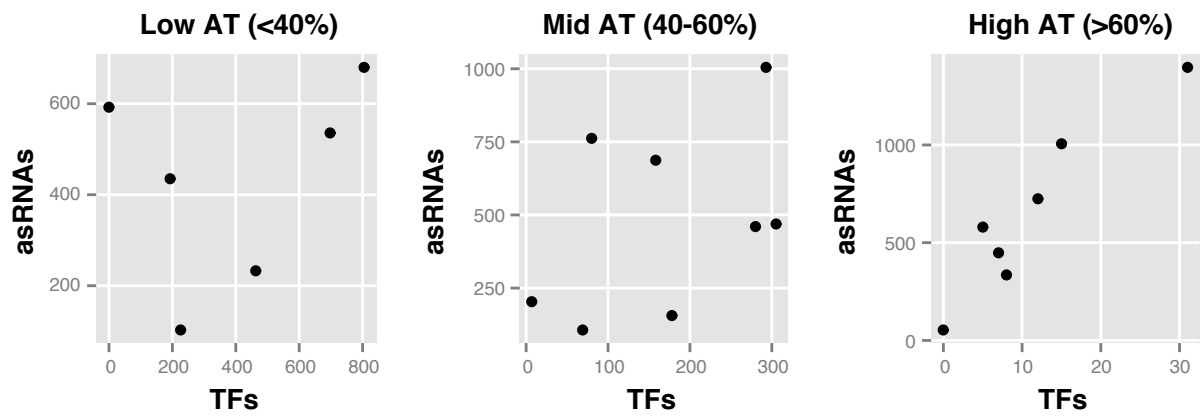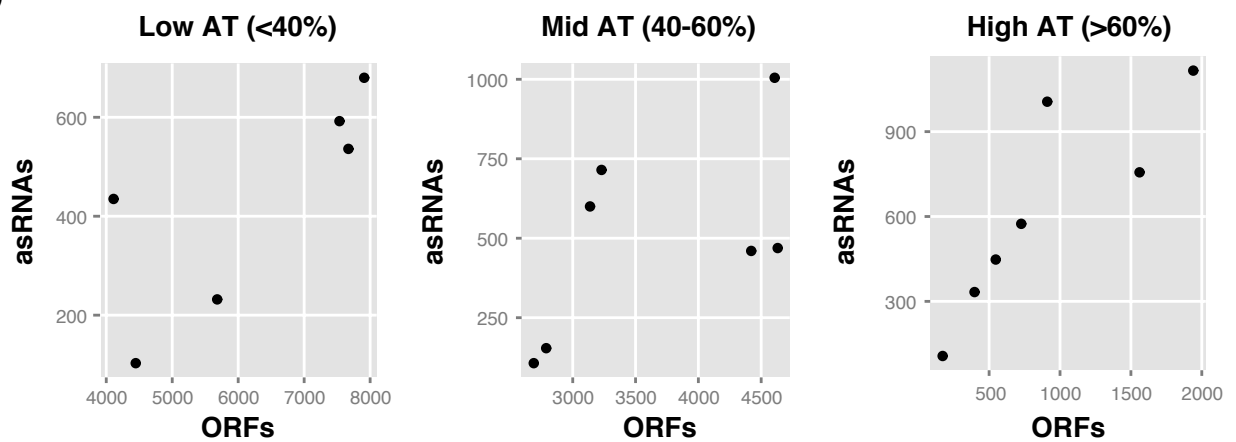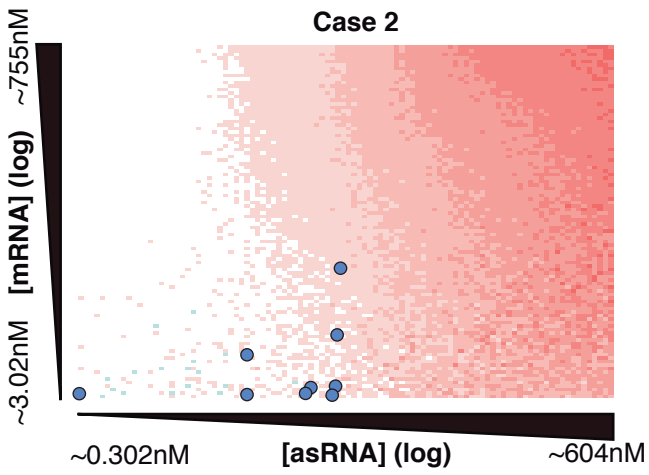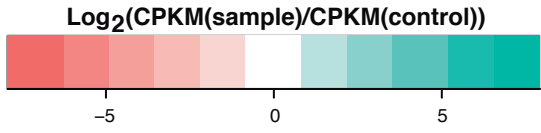
Figure S6

Figure S7: Simulation of the effect of the asRNAs, assuming that the pairing asRNA-mRNA causes mRNA degradation (case 2) or translation inhibition (case 3). Parameters for the simulations are detailed in the Supplementary Information. Each point of the heatmaps represents the average change of the mRNA concentration for 100 simulations of 1000 minutes each, for specific parameters of asRNA and mRNA transcription rates. The remaining parameters remain constant for all the simulations. The axes represent the mRNA and asRNA concentration in the control experiments for the corresponding transcription rates scanned. A) Changes in the mRNA concentration after 1000 minutes of simulation. Blue circles represent experimental data from the overexpression of asRNAs in *M. pneumoniae*. B) Changes in the protein concentration after 1000 minutes of simulation. Blue circles represent experimental data from the overexpression of asRNAs in *M. pneumoniae*.

Figure S7

● *M. pneumoniae* asRNA overexpression

A Log$_2$(CPKM(sample)/CPKM(control))

B Log$_2$(Protein sample/Protein control)