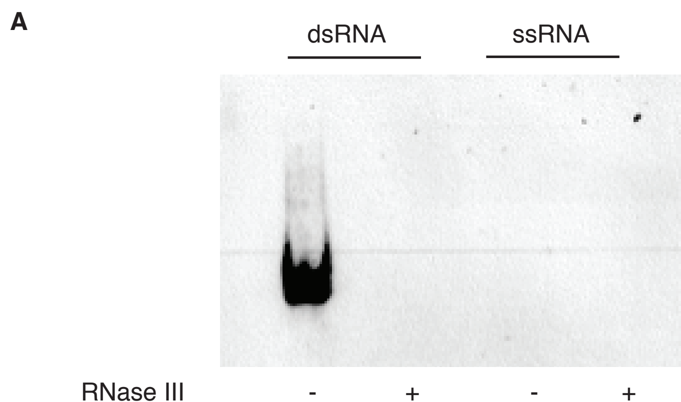
A



B

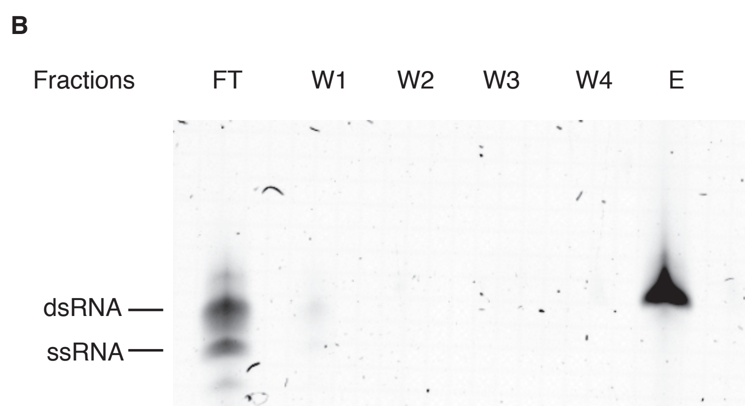
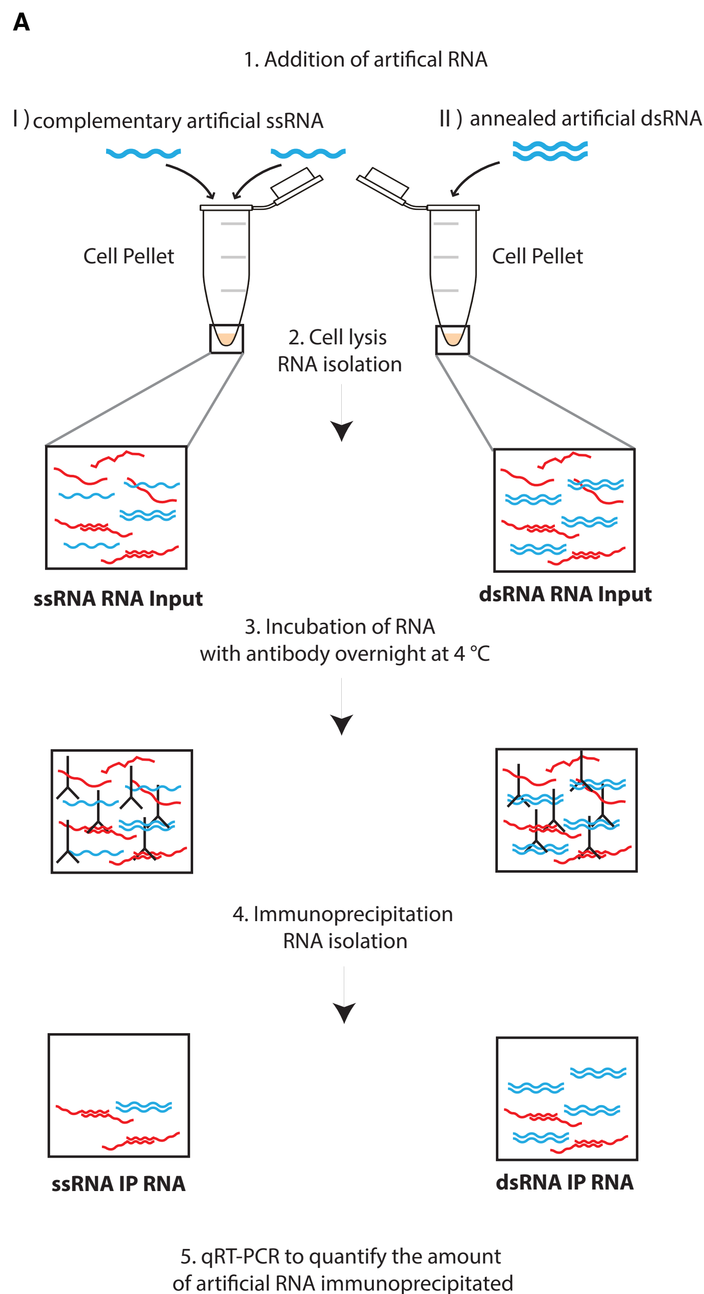


Figure S1. Characterization of the J2 dsRNA antibody. (A) An RNA immunoblot of artificial dsRNA and ssRNA using the J2 dsRNA antibody. *In vitro* transcribed dsRNA and ssRNA were separated on an 8% native polyacrylamide gel, electroblotted to a nylon + membrane and hybridized with the J2 monoclonal antibody. (B) *In vitro* transcribed dsRNA and ssRNAs were mixed together and immunoprecipitated with the J2 monoclonal antibody. The fractions from the immunoprecipitation flow through (FT), wash (W) and elution (E) were separated on an 8% denaturing polyacrylamide gel and visualized with SYBR Safe staining.

A

B

Lybecker_Fig. S2



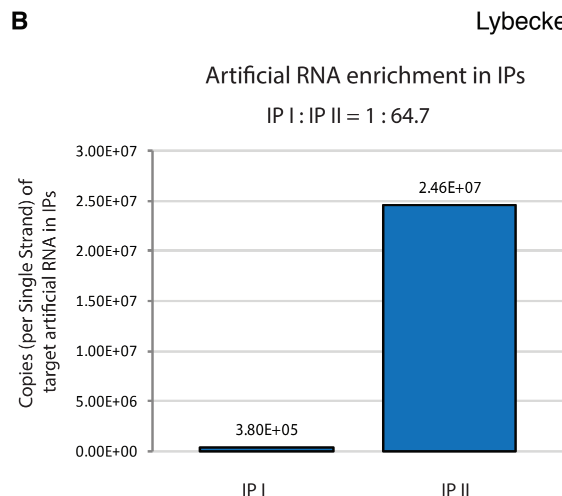


Figure S2. Quantification of artificial double-stranded and single stranded RNA substrates in the IP protocol. (A) A schematic representation of the experimental procedure. The resulting IP-RNA from both samples were subjected to qRT-PCR to quantify the amounts of dsRNA immunoprecipitated. (B) qRT-PCR results for the artificial RNA enrichment in IP-dsRNAs. Copies of the artificial transcript in each immuniprecipitation were calculated per single strand using the standard curve method (described in the Supplemental Materials and Methods) and normalized to the initial ammounts in inputs.The sample spiked with pre-formed artificial dsRNAs (II) showed more than 64-fold enrichment of immunoprecipitated artificial transcript compared to the sample spiked with artificial RNA in form of complementary single strands (I).

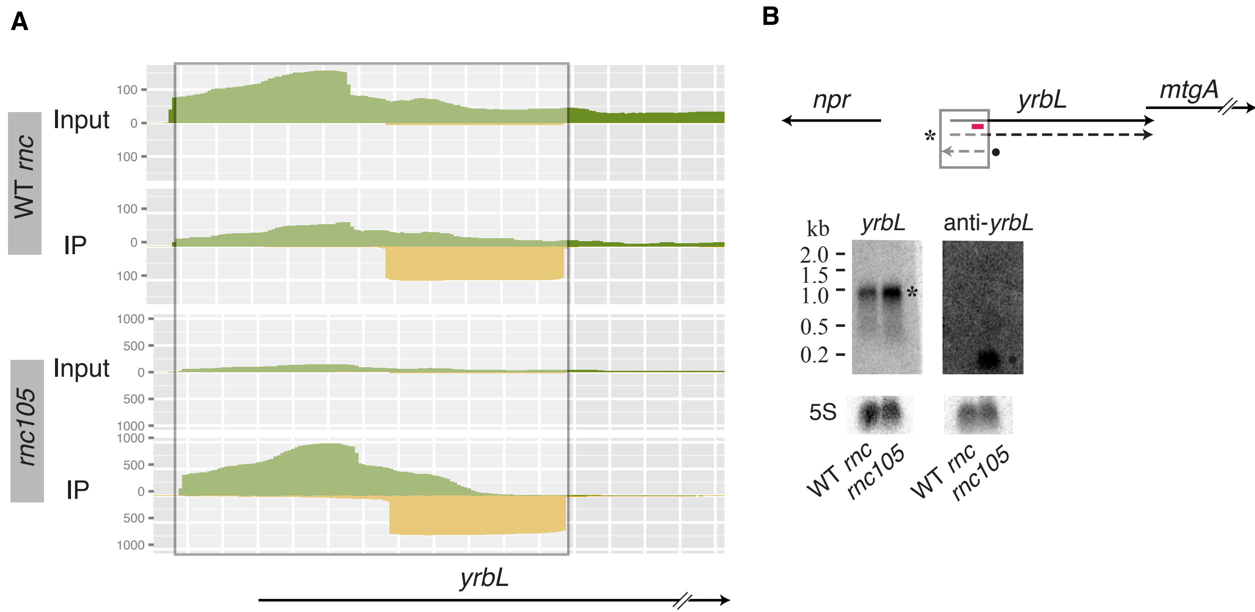


Figure S3. Northern blot analyses of IP-dsRNAs. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for *yrbL* (A and B). (A and B) as described in Fig. 4 legend.

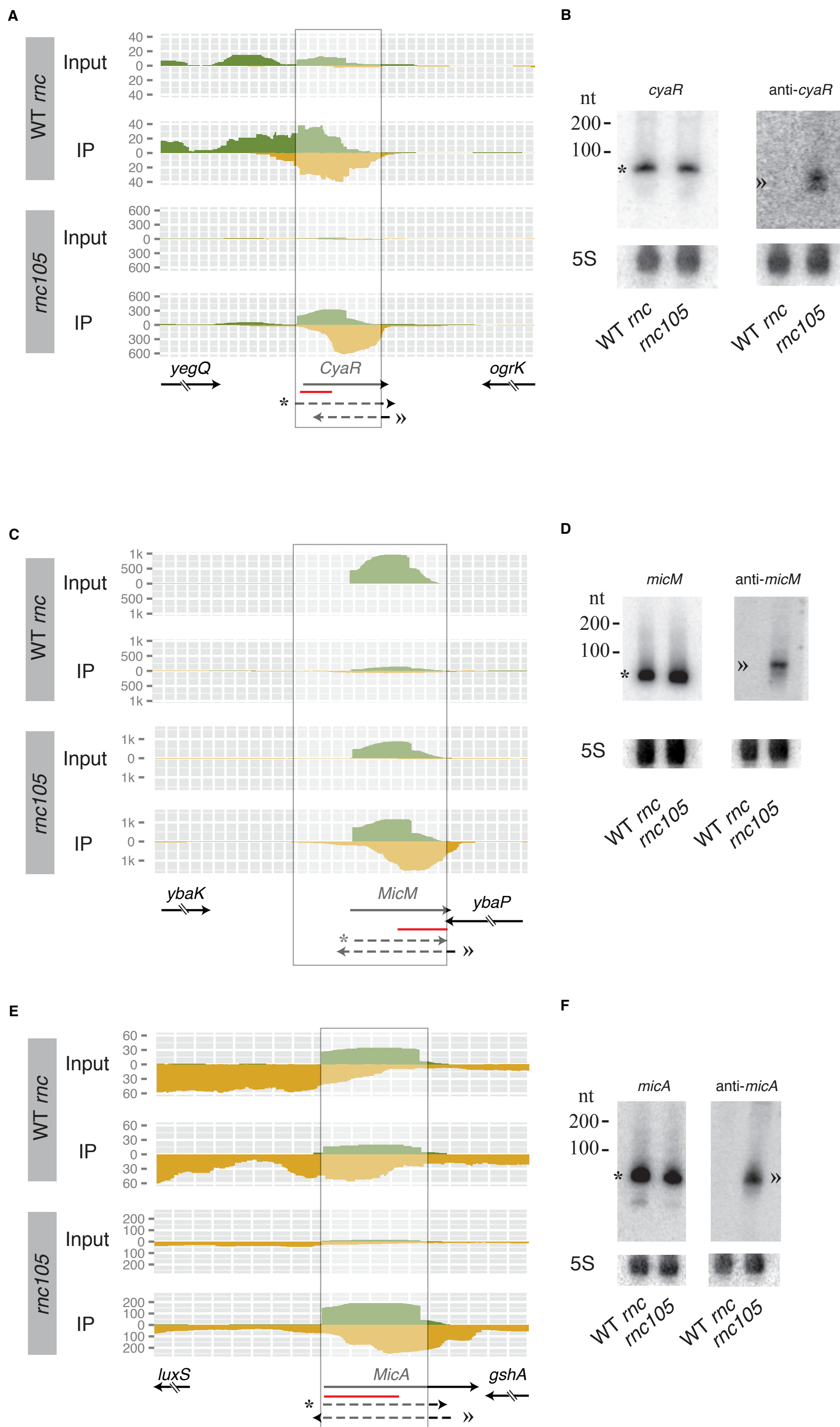


Figure S4. Verification of IP-dsRNAs transcripts at ncRNAs. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for *cyaR* (A and B), *micM/chiX* (C and D) and *micA* (E and F) genes. (A - F) as described in Fig. 4 legend.

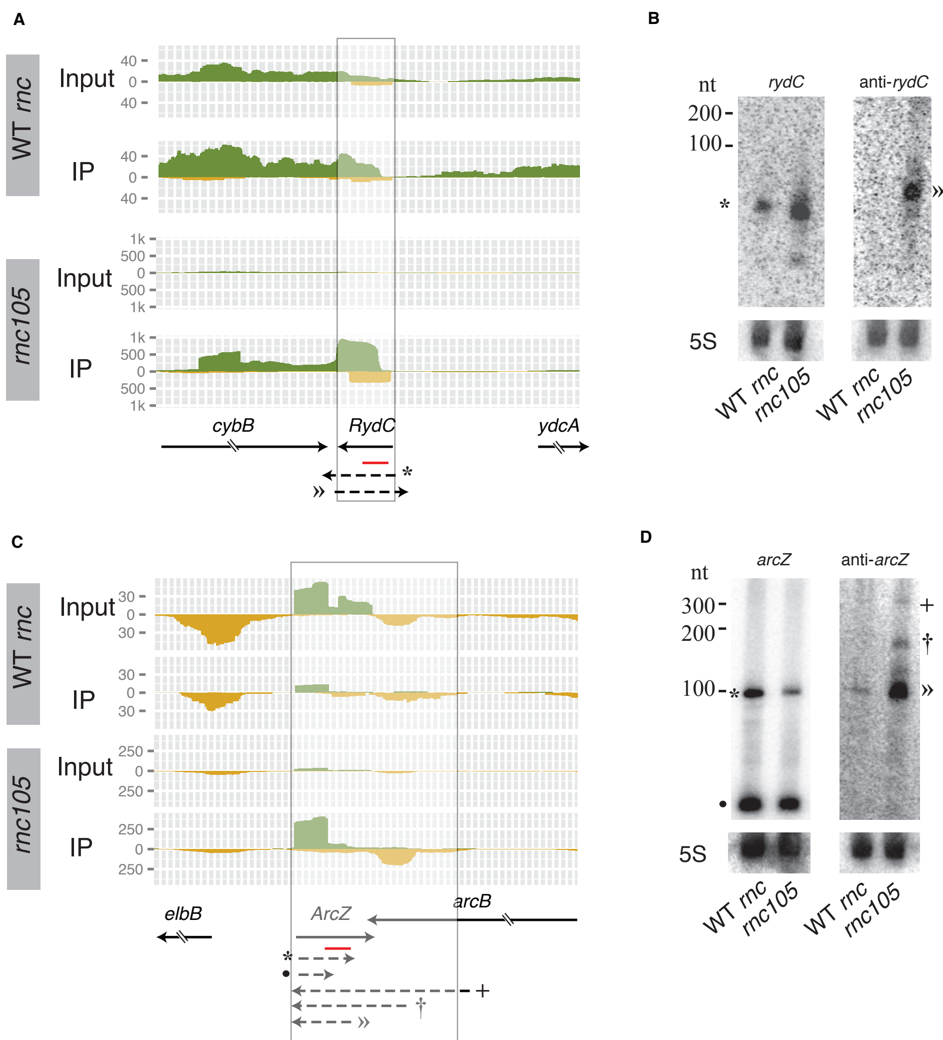


Figure S5. Verification of IP-dsRNAs transcripts at ncRNAs. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for *rydC* (A and B) and *arcZ* (C and D). (A - D) as described in Fig. 4 legend.

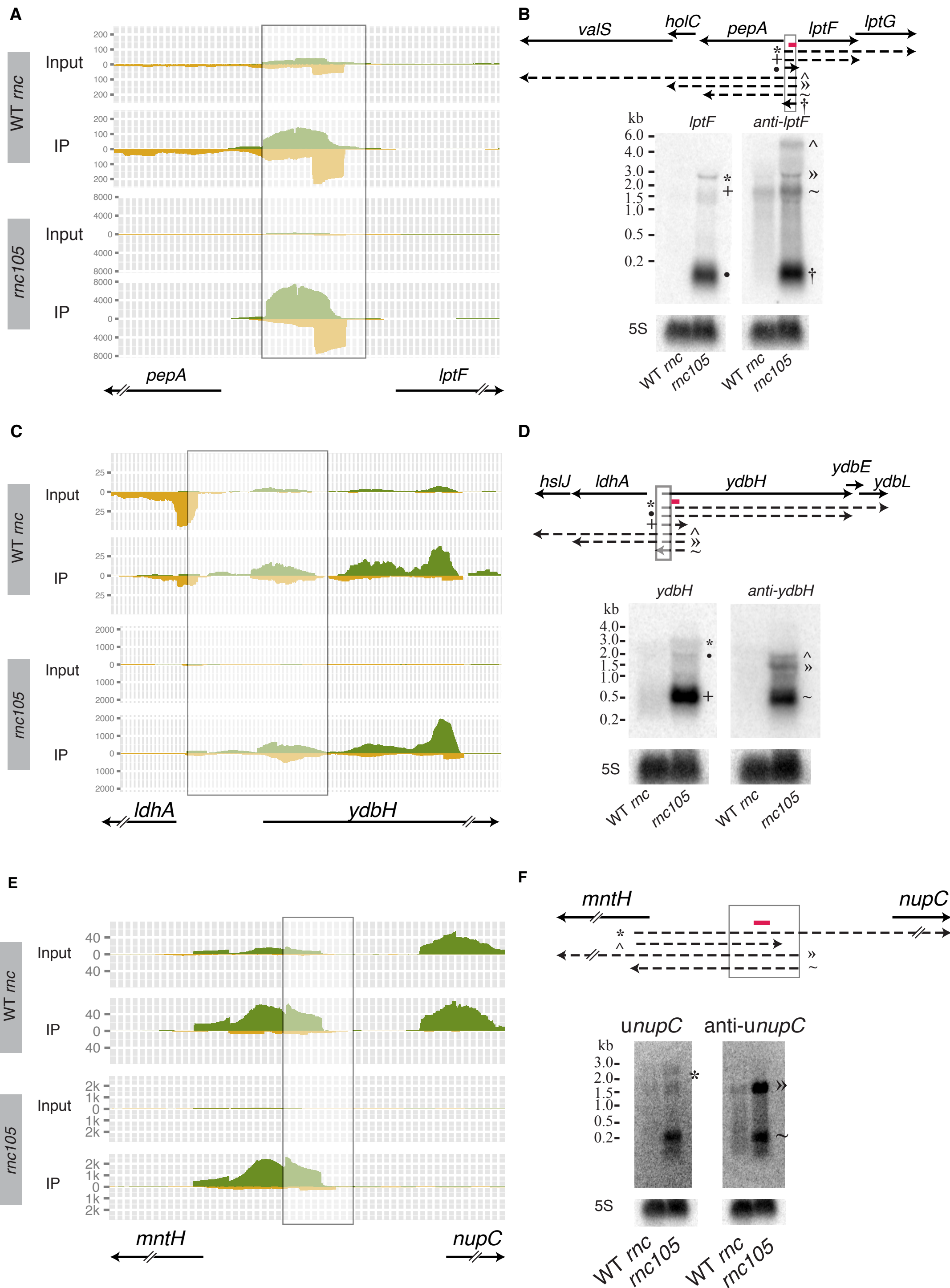


Figure S6. Verification of IP-dsRNAs formed by 5′ divergently transcribed mRNAs. The transcripts forming the IP-dsRNA were confirmed via Northern blot analyses for three different divergent gene pairs: *lptf/pepA* (A and B), *ydbH/ldhA* (C and D) and upstream (u) *nupC/mntH* (E and F). (A - D) as described in the Fig. 4 legend.

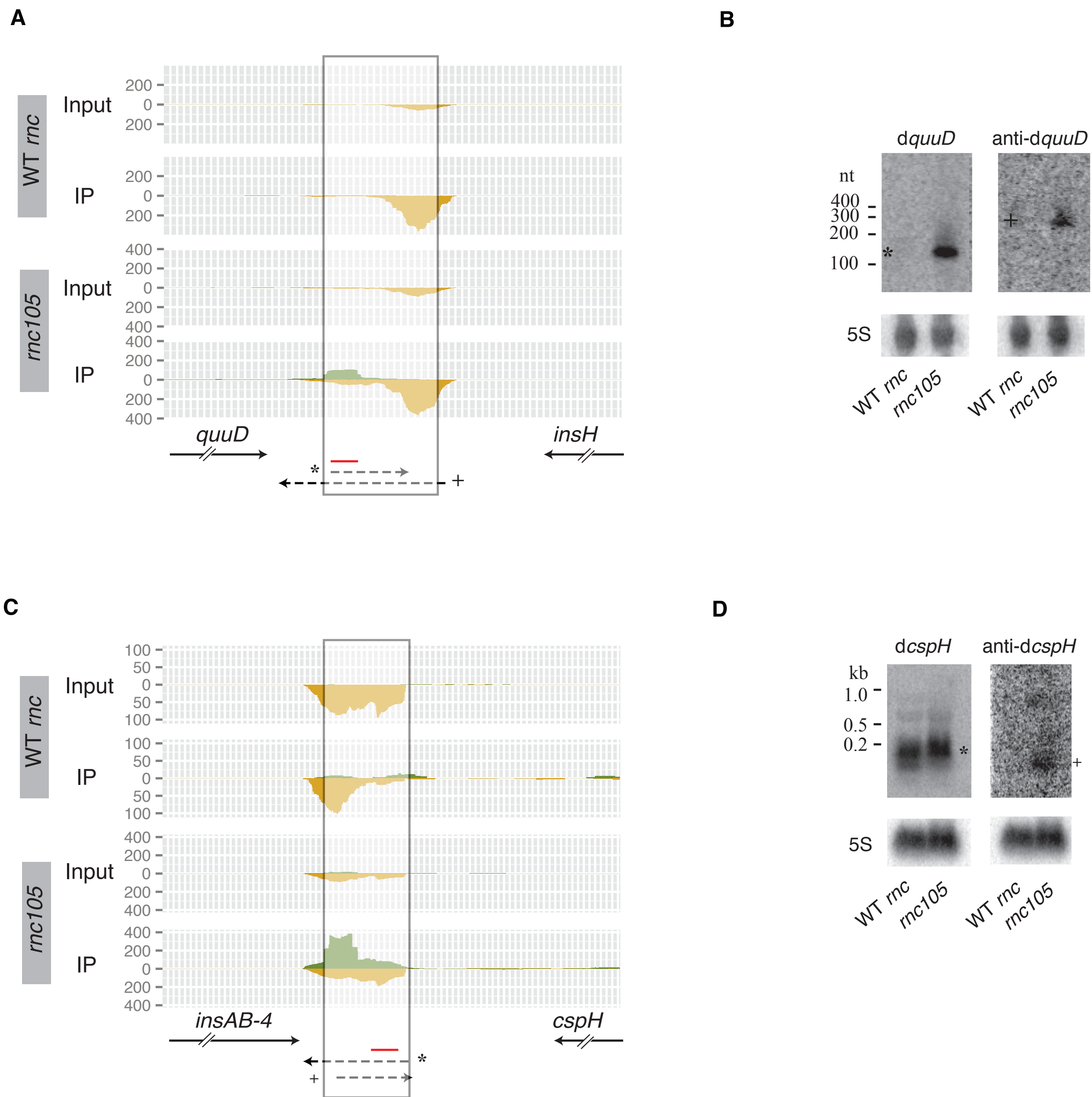


Figure S7. Northern blot analyses of IP-dsRNAs found downstream of transposase genes. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for regions downstream *quuD* (A and B) and *cspH* (C and D). (A - D) as described in Fig. 4 legend.

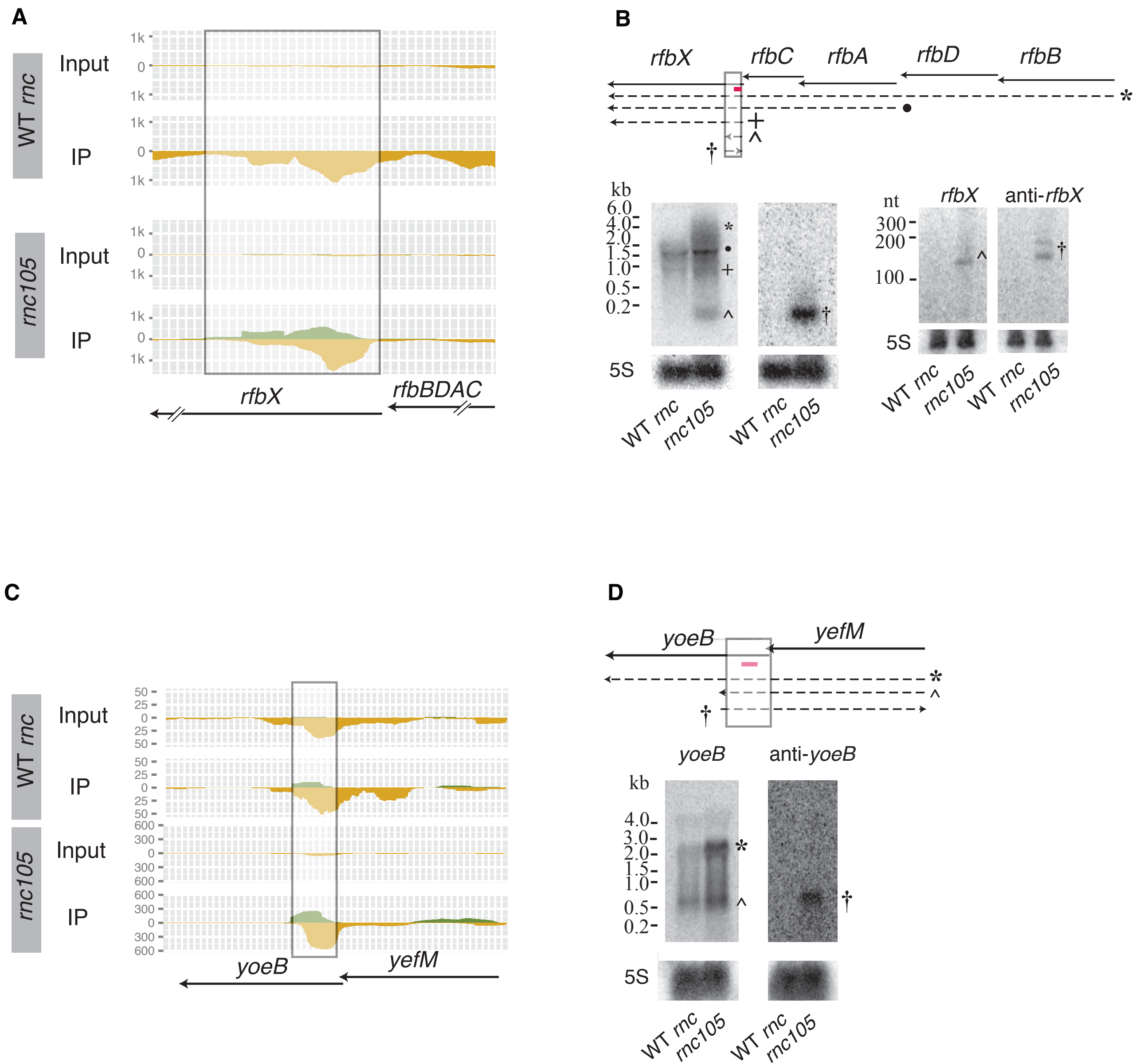


Figure S8. Northern blot analyses of IP-dsRNAs. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for *rfbX* (A and B) and *yoeB* (C and D). (A - D) as described in Fig. 4 legend.

Lybecker_Fig. S9

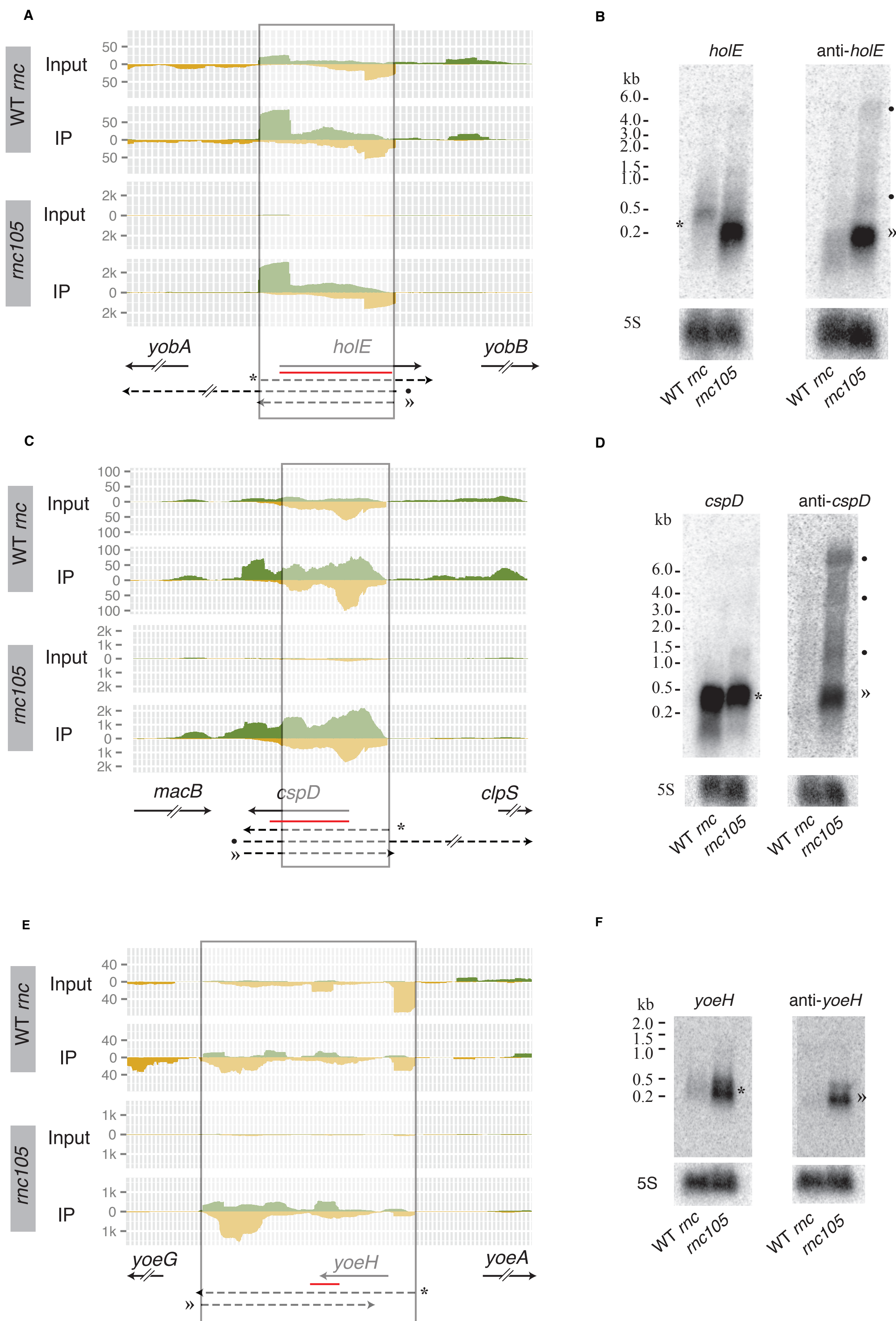
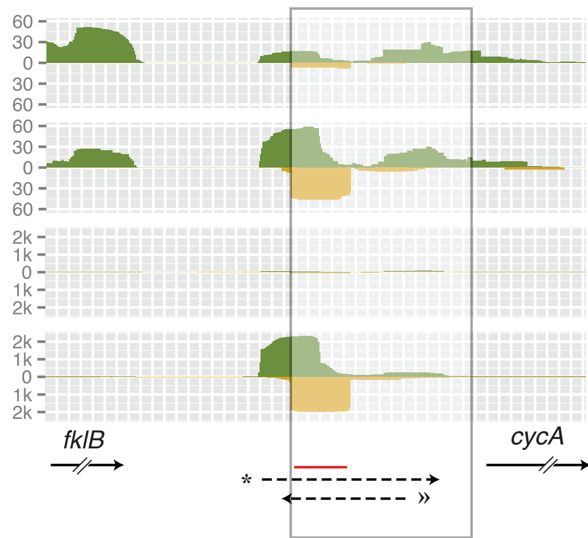


Figure S9.Verification of IP-dsRNAs that are categorized as Full ORF. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for *holE* (A and B), *cspD* (C and D) and *yoeH* (E and F). (A - F) as described in Fig. 4 legend.

A



B

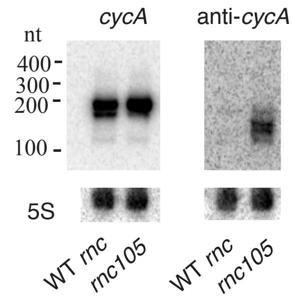


Figure S10. Northern blot analyses of IP-dsRNAs. The transcripts in IP-dsRNAs were confirmed via Northern blot analyses for *cycA* (A and B) as described in Fig. 4 legend.

Table S2. IP-dsRNAs confirmed by Northern blot

Category	Annotation of overlapping genes	Coordinates IP-dsRNAs	Seed IP-dsRNA Coordinates	Adjacent genes	Class	Score	Northern blot comments[*]		Shown in Fig.
							Sense	AS	
Divergent	*ydbH/ldhA*	1,440,854-1,441,185	1,441,030-1,441,185		II	77	+	+	S6
	lptF/pepA	4,484,030-4,484,238	4,484,030-4,484,157		II	15	+	∞	S6
	nupC/mntH	2,510,842-2,510,938	2,510,842-2,510,938		I	36	+	+	S6
5′/Intergenic	*yrbL*	3,346,446-3,346,554	3,346,446-3,346,554	*mtgA/npr*	I	71	+	∞	S3
	msbA	965,816-966,058	965,816-965,911	*ycaI/lpxK*	II	63	+	∞	4
5′ ORF	*yoeB*	2,087,410-2,087,484	2,087,410-2,087,484	*hisL/yefM*	I	40	+	∞	S8
	rfbX	2,107,251-2,107,598	2,107,451-2,107,598	*rfbC/glf*	II	67	+	∞	S8
Convergent	*cspH/insAB-4*	1,049,778-1,049,925	1,049,778-1,049,925		I	17	≈	∞	S7
	quuD/insH	573,638-573,805	573,638-573,805		I	11	∞	∞	S7
Junction	*mqsR/mqsA*	3,166,223-3,166,578	3,166,442-3,166,578		I	39	+	∞	5
Intergenic	*cycA*	4,427,722-4,427,870	4,427,723-4,427,772	*fklB/ytfE*	II	24	≈	∞	S10
Full ORF	*holE*	1,923,109-1,923,328	1,923,109-1,923,328	*yobA/yobB*	II	143	+	+	S9
	cspD	921,665-921,901	921,763-921,861	*macB/clpS*	I	83	+	+	S9
	yoeH	2,066,003-2,066,504	2,066,003-2,066,143	*yoeG/yoeA*	II	25	+	∞	S9
ncRNA	*cyaR*	2,165,136-2,165,219	2,165,136-2,165,219	*yegQ/ogrK*	I	50	≈	∞	S4
	spf	4,047,920-4,048,008	4,047,920-4,048,008	*polA/yihA*	I	200	-	∞	6
	micM (*chiX*)	506,377-506,506	506,426-506,506	*ybaK/ybaP*	I	167	≈	∞	S4
	mgrR	1,620,840-1,620,935	1,620,840-1,620,935	*ydeI/yneM*	I	35	≈	∞	6
	arcZ	3,348,596-3,348,835	3,348,596-3,348,835	*elbB/arcB*	II	13	-	+	S5
	micA	2,812,822-2,812,876	2,812,822-2,812,876	*luxS/gshA*	II	25	≈	∞	S4
	rydC	1,489,461-1,489,530	1,489,461-1,489,530	*cybB/ydcA*	II	42	+	∞	S5

[*] Steady state level of the transcript in the *rnc105* strain compared to the WT strain. (**+** increased, **-** decreased, ≈ approximately equal and **∞** only detected in the mutant strain)

Table S3. Oligonucleotides used.

	Sequence (5'-3')
DNA templates for *in vitro* transcription	
Single stranded Forward + T7	CCAACTAATACGACTCACTATAGGTAGAGCGGGGGAGGTGAAGCGACGTATCAGCGGG CGAGGCCAGGAACGCGTCAGAGGTCG
Single stranded Reverse + T7	CGACCTCTGACGCGTTCCTGGCCTCGCCCGCTGATACGTCGCTTCACCTCCCCCGCTCTA CCTATAGTGAGTCGTATTAGTTGG
Double stranded (+) Forward + T7	CCAACTAATACGACTCACTATAGGCACGACAGAAGAATTCTCAGTAACTTCCTTGTGTT GTGTGTATTCAACTCACAGAGTTGAACGTTCCTTT
Double stranded (+) Reverse + T7	AAAGGAACGTTCAACTCTGTGAGTTGAATACACACAACACAAGGAAGTTACTGAGAATT CTTCTGTCGTGCCTATAGTGAGTCGTATTAGTTGG
Double stranded (-) Forward + T7	CCAACTAATACGACTCACTATAGGAAAGGAACGTTCAACTCTGTGAGTTGAATACACACA ACACAAGGAAGTTACTGAGAATTCTTCTGTCGTG
Double stranded (-) Reverse + T7	CACGACAGAAGAATTCTCAGTAACTTCCTTGTGTTGTGTGTATTCAACTCACAGAGTTGA ACGTTCCTTTCCTATAGTGAGTCGTATTAGTTGG
Primers used for PCR amplification of template DNA in in vitro transcription	
Long gfp t7 fwd (+)	TAATACGACTCACTATAGGGCAACTACAAGACCCGC
Long gfp rev (+)	ACGAACTCCAGCAGGACC
Long gfp fwd (-)	GGCAACTACAAGACCCGC
Long gfp t7 rev (-)	TAATACGACTCACTATAGGGACGAACTCCAGCAGGACC
Northern blots oligonucleotide probes (target strand)	
ydbH Sense	ACGCGTAATGCGTGGGCTTTCATCTAATGCAATACGTGTCC
ydbH Antisense	GGACACGTATTGCATTAGATGAAAGCCCACGCATTACGCGT
lptF Sense	GTAATAACGTTAAGATTAACACGAAGTCATCGCAACAGCGGAC
lptF Antisense	GTCCGCTGTTGCGATGACTTCGTGTTAATCTTAACGTTATTAC
nupC Sense	ACATAACCATAACATTCAAGTCATCAATCTTTTCGTAGTC

nupC Antisense	GACTACGAAAAGATTGATGACTTGAATGTTATGGTTATGT
yrbL Sense	CGTTGGGCATCTTCCGGATGCGCATAACATTTCCGATGTCTT
yrbL Antisense	AAGACATCGGAAATGTTATGCGCATCCGGAAGATGCCCAACG
yoeH Sense	GTATCAGCAATTAACCGGACCTGATACTGATATGAGTCTTAC
yoeH Antisense	GTAAGACTCATATCAGTATCAGGTCCGGTTAATTGCTGATAC
cycA Sense	TACTAAGATTTAGATCTAATGTTAATTTTATGTTTAAAGT
cycA Antisense	ACTTTAAACATAAAATTAACATTAGATCTAAATCTTAGTA
yoeB Sense	ATCTTTTTAACAATTCGCTTATCTGTTTCCTGCCAGTACAGA
yoeB Antisense	TCTGTACTGGCAGGAAACAGATAAGCGAATTGTTAAAAAGAT
rfbX Sense	GCATAGTCGCTTGGCAAAAACCGAATATACCGAAATTTTC
rfbX Antisense	GAAAATTTCGGTATATTCGGTTTTTGCCAAGCGACTATGC
cyaR Sense	GGTTCCTGGTACAGCTAGCATTTTATGGGTTATGTTTTTC
cyaR Antisense	GAAAAACATAACCCATAAAATGCTAGCTGTACCAGGAACC
spf Sense	ATCCGATTACGTGAAGTAAAAGGTCTGAAAGATAGAACATC
spf Antisense	GATGTTCTATCTTTCAGACCTTTTACTTCACGTAATCGGAT
micM Sense	TGGCCAATATCGCTATTGGCCCGTCAAAGAGGAATTTCATT
micM Antisense	AATGAAATTCCTCTTTGACGGGCCAATAGCGATATTGGCCA
mgrR Sense	TTTGCTTTTACGCTAACAGGCATTTTCCTGCACTGATAACG
mgrR Antisense	CGTTATCAGTGCAGGAAAATGCCTGTTAGCGTAAAAGCAAA
msbA Sense	CCACAGTCGGCGGAATGTCTGCCACGTAGAGAGATCTTTGTC
msbA Antisense	GACAAAGATCTCTCTACGTGGCAGACATTCCGCCGACTGTGG
mqsR Sense	GACTCAAACGTGTATGTGGTGTGCGTTTTTCCATAACCCCCG
mqsR Antisense	CGGGGGTTATGGAAAAACGCACACCACATACACGTTTGAGTC
cspH Sense	ACTGCGGGTAGGAAAGGCGCGCAGAGGGAAATATAAGATTG
cspH Antisense	CAATCTTATATTTCCCTCTGCGCGCCTTTCCTACCCGCAGT
quuD Sense	CACTTCCTTTAGAAGCACAACCTTGCTTCTAACTATATA
quuD Antisense	TATATAGTTAGAAGCAAGGTTGTGCTTCTAAAGGAAGTG
micA Sense	TTCATCTCTGAATTCAGGGATGATGATAACAAATGCGCGTC
micA Antisense	GACGCGCATTTGTTATCATCATCCCTGAATTCAGAGATGAA

rydC Sense	CGCCTGTACTAAAACCGACCCGTGGTACAGGCGAAGAAT
rydC Antisense	ATTCTTCGCCTGTACCACGGGTCGGTTTTAGTACAGGCG
arcZ Sense	ACCCCGGCTAGACCGGGGTGCGCGAATACTGCGCCAACACCAGG
arcZ Antisense	CCTGGTGTTGGCGCAGTATTCGCGCACCCCGGTCTAGCCGGGGT
Northern blot oligos for RNA probes	
cspD Sense F	AAAGGGTTTGGTTTCATC
cspD Sense R +T7	TAATACGACTCACTATAGGGACGGGCACAATAACACTGGC
cspD Antisense F +T7	TAATACGACTCACTATAGGGAAAGGGTTTGGTTTCATC
cspD Antisense R	ACGGGCACAATAACACTGGC
holE Sense F	GCTGAAGAATCTGGCTAAAC
holE Sense R +T7	TAATACGACTCACTATAGGGAAACGGTGGGCAATAAGC
holE Antisense F +T7	TAATACGACTCACTATAGGGCTGAAGAATCTGGCTAAAC
holE Antisense R	AAACGGTGGGCAATAAGC
Primers used in qRT-PCR	
Short_GFP_FW3	CTCGCCGACCACTACCA
Short_GFP_RV3	TTTGCTCAGGGCGGACT

Bacterial strains and RNA isolation.

E. coli strains SDF204 (W3110*rnc*$^+$ TD1-17::Tn10) and SDF205 (W3110*rnc105* TD1-17::Tn10) were grown in LB medium to log phase (optical density at 600nm [OD_{600}] of ~ 0.5). For Northern blot analyses total RNA was isolated, unless otherwise stated, using a hot phenol protocol described by Jahn et al. (1). Briefly, cells were first harvested for RNA isolation by being added in a 10:1 ratio to a mixture of 95% ethanol and 5% saturated phenol and centrifuged at 3000 × *g* at 4ºC for 5 min to stabilize cellular RNA. Supernatant was decanted and the pellets were frozen in liquid N_2. Pellets were resuspended in a lysis buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mg/mL lysozyme) and 10% SDS (wt:vol) was added to a final concentration of 0.1%. The lysate was then incubated at 64ºC for 2 min. 1M NaOAc (pH 5.2) was added to the lysate to a final concentration of 0.1 M. Then an equal volume of water-saturated phenol was added, mixed by inversion and incubated at 64ºC for 6 min with inversion of the tubes approximately every 40 s. Samples were chilled on ice and centrifuged at 16,200 × *g* for 10 min at 4ºC. The aqueous layer was transferred to a Phase Lock Gel tube with an equal volume of chloroform/isoamyl alcohol (24:1), inverted several times and centrifuged for 5 min at 16,200 × *g*. The RNA was then ethanol precipitated from the aqueous layer using NaOAc. Total RNA isolated for immuno-dot blots and immunoprecipitations followed the hot phenol protocol described above with the following modifications to minimize dsRNA artifacts: the cell lysis was performed at 4ºC for 30 min and the RNA extraction was performed on ice with cold phenol. Total RNA was treated with DNase I (Roche) following the manufacturer’s protocol. RNA integrity was measured using the Agilent 2100 Bioanalyzer. RNA with a RNA Integrity Number (RIN) above a 9.0 was used for the immunoprecipitation assays.

In vitro transcription of double-stranded and single-stranded RNA substrates

For experiments shown in Fig. S1 the double-stranded DNA templates for *in vitro* transcription reactions were generated by annealing complementary oligonucleotides containing the T7 promoter (Table S2). The complementary oligonucleotides were annealed in 10 mM Tris-HCl, pH 8.0, and 10 mM $MgCl_2$ by heating to 95ºC for 3 min and then slow-cooling to 37ºC. The dsDNA templates were used for *in vitro* RNA transcription. Transcription was carried out in transcription buffer (40 mM Tris-HCl, pH 7.5, 26 mM $MgCl_2$, 3 mM spermidine, 5 mM DTT, 5 mM NTPs, 100 units T7 polymerase (New England Biolabs) and 20-40 units of RNasin (Promega)) shaking at 37°C overnight. Template DNA used for *in vitro* transcription reactions for experiments presented in Fig. S2 were generated using pGFP_LacZ (2) as a template for PCR with the T7 bacteriophage promoter added to the 5′ or 3′ end (Table 3). RNA was in vitro transcribed using the MEGAscript *in vitro* transcription kit from Ambion according to the manufacturer's instructions. The RNAs were gel-purified on an 8% denaturing (8 M urea) polyacrylamide gel in 1X TBE, excised from the gel, eluted in 1X elution buffer (10 mM Tris-HCl, pH 7.5, 0.1% SDS, 2 mM EDTA and 0.3 M NaOAc) by shaking at 4°C overnight and ethanol precipitated. Double-stranded RNA was constructed by annealing the *in vitro* transcribed + and – strands by heating the RNAs together to 65°C for 5 min and slow cooling to room temperature.

RNA immunoblots and dot-blots

SDS-PAGE RNA immunoblotting was performed with the J2 monoclonal antibody to dsRNA (English Scientific Consutling Bt.) The J2 monoclonal antibody

was reconstituted to 1.0 mg/ml concentration with RNase free water. Artificial *in vitro* transcribed RNAs were separated on a native 8% polyacrylamide gel and electroblotted to nylon + membrane (Amersham). Briefly, 1 μg of RNA was mixed with native RNA loading buffer (5X: 150 mM EDTA, pH 8.0, 2% SDS, 15% (wt:vol) sucrose, 12.5% (wt:vol) Ficoll and 0.2 mg/mL yeast tRNA), loaded onto an 8% native polyacrylamide gel and electrophoresed in 1X TBE for 1 h at 150V. The RNA was then electroblotted to nylon + membrane in 0.5X TBE at 12V for 1 h. The membranes were blocked overnight in 1X PBS, 0.5% Tween-20, 0.05 mg/ml sheared salmon sperm, and 5% (wt:vol) nonfat dried milk. The membranes were incubated at room temperature for 2 h with the primary antibody (J2 monoclonal dsRNA antibody) diluted 1:2000 in 1X PBS and 0.5% Tween-20. The membranes were washed 1× 5 min and 1× 30 min in wash buffer (1X PBS and 0.5% Tween-20) and then incubated at room temperature for 1 h with the secondary antibody (goat anti-mouse peroxidase conjugated (Jackson Immunology)) diluted 1:10,000 in wash buffer. The membranes were then washed 1× 5 min and 2× 10 min in wash buffer. The membranes were developed by chemiluminescence, using luminol per the manufacturer's protocol, with a Curix 60 table top processor (AGFA healthcare).

Dot-blotting was performed using the BioRAD Biodot Microfiltration apparatus following the manufacturer's protocol. RNA was applied in 50 μl reaction volumes in 0.5X TBE to a nylon + membrane using the dot blot apparatus. RNA was crosslinked at 30 mJ/cm^2. The membranes were immunoblotted as described above.

Immunoprecipitation assays

Total RNA and/or *in vitro* transcribed RNAs were incubated with the J2 monoclonal anti-double-stranded RNA antibodies at 4°C overnight in 1X PBS and

0.1% Tween 20 with 2 units of RNasin (Promega). Dynabeads® Protein A (Invitrogen) were used to immunoprecipitate the antibody:RNA complexes. The Dynabeads were prepared as suggested by the manufacturer. The antibody:RNA solutions were then added to the beads, gently mixed and incubated for 10 min at room temperature while rotating. The tubes were then moved to a magnetic stand and the supernatant was removed and the beads were washed 4X and then resuspended with 1X PBS and 0.1% Tween-20. The beads with the antibody:RNA complexes bound were then subjected to phenol/chloroform/isoamyl alcohol (25:24:1) extraction using a Phase Lock Gel tube. The aqueous phase was ethanol-precipitated and analyzed either by SDS-PAGE and SYBR® Safe staining or by the Agilent 2100 Bioanalyzer on a picoRNA chip. For the control experiment the *in vitro* transcribed pre-annealed dsRNA or ssRNAs were added to the cell pellet before lysis and the immunoprecipiation protocol was followed as described above.

RNA quantification by qRT-PCR

The amount of artificial transcripts immunoprecipitated in the control experiment were determined by absolute quantification using the standard curve method in one color detection system as described before (3). Briefly, to prepare the standards equal amounts of *in vitro* transcribed artificial ssRNA or dsRNA molecules were used in a qRT-PCR reaction. 1 ng of one-to-one mixed complementary ssRNAs or 2 ng of pre-formed artificial dsRNAs were added respectively with 299 ng or 298 ng of total RNA from *E.coli* for the reverse transcription (RT). The RT step was performed using SuperScript® III Reverse Transcriptase (Life Technologies) as suggested by the manufacturer. From the derived calculated molecular concentration (using the molecular mass of single stranded (ss) - 45244.5 g/mol and double stranded (ds)

artificial RNA – 90489 g/mol) accurate serial dilutions were made to obtain calibration curve. qPCR was performed using Eppendorf Mastercycler® RealPlex^2 and 5x HOT FIREPol® EvaGreen® qPCR Mix Plus (no ROX) from Medibena as suggested by the manufacturers.

Immunoprecipitated RNA from the control experiment, as well as the RNA inputs, were subjected to qRT-PCR. The absolute amounts of RNA molecules were calculated using the above mentioned calibration curve.

cDNA library preparation

Directional (strand-specific) RNA-seq cDNA libraries were constructed with the following protocol. The total and immunoprecipitated RNAs were first treated with DNase I (Roche) and depleted of ribosomal RNA using the Ribo-zero™ rRNA removal kit for Gram-negative bacteria (Epicenter). 250 ng of RNA were fragmented using the RNA fragmentation reagents (Ambion®) per the manufacturer's protocol. The total RNA libraries were fragmented at 70ºC for 3 min, while the immunoprecipitated RNA libraries were fragmented at 70ºC for 10 min. The fragmented RNAs were analyzed on an Agilent 2100 Bioanalyzer. 150 ng of the fragmented RNAs were treated sequentially with tobacco acid pyrophosphatase (Epicenter) and calf intestinal phosphatase (New England Biolabs) per the manufacturer's protocols to remove 5′ and 3′ phosphates. Finally, RNA was treated with polynucleotide kinase (T4 PNK; New England Biolabs), without ATP, to remove 2′-3′ cyclic-phosphates as described (4). A 3′ RNA adapter, based on the Illumina multiplexing adapter sequence (Oligonucleotide sequences © 2007-2009 Illumina, Inc. All rights reserved) blocked at the 3′ end with an inverted dT (5′-GAUCGGAAGAGCACACGUCU[idT]-3′), was phosphorylated at the 5′ end

using T4 PNK (New England Biolabs) per the manufacturer's protocol. The 3′ multiplex RNA adapter was ligated to the 3′ ends of the total and immunoprecipitated RNAs using T4 RNA ligase I (New England Biolabs). Briefly, RNA was incubated at 20ºC for 6 h in 1X T4 RNA ligase reaction buffer with 1 mM ATP, 20 μM 3′ RNA adapter, 1 μl DMSO, 5 units of T4 RNA ligase I, and 20-40 units of RNasin (Promega) in a 10 μl reaction. The RNA was then size selected (≈ 75–400 nucleotides) and purified on a denaturing 8M urea SDS polyacrylamide gel. Gel slices were incubated in RNA elution buffer at 4ºC overnight. The supernatant was subsequently ethanol precipitated using glycogen as a carrier molecule. The RNAs were phosphorylated at the 5′ ends, to allow for ligation to the 5′ RNA adapter, using T4 PNK (New England Biolabs) per the manufacturer's protocol. The Illumina small RNA 5′ adapter (5′-GUUCAGAGUUCUACAGUCCGACGAUC-3′) was ligated to the libraries using T4 RNA ligase I. The ligation reactions were size selected (≈ 100–450 nucleotides) and gel-purified on a denaturing 8M urea SDS polyacrylamide gel (as described above). The di-tagged RNA libraries were reverse transcribed with SuperscriptII reverse transcriptase (Invitrogen) using random nonamers per the manufacturer's protocol. Polymerase chain reaction (PCR) was carried out using Phusion High-fidelity polymerase. Illumina compatible PCR primers were used to add the appropriate sequences for Solexa sequencing. cDNA was amplified by 15 cycles of PCR and the products were analyzed on a Agilent 2100 Bioanalyzer. cDNA libraries were sequenced with Solexa technology on a HiSeq 2000 with single-end 50 base-pair reads at the CSF NGS unit http://csf.ac.at/.

Deep-sequencing analyses

The *E. coli* K12 genome was downloaded from Genbank (5). Reads were mapped using Bowtie2 in local alignment mode and default parameters (6).

Alignments with mapping quality values of less than 40 were discarded. This resulted in 44,364,068 of the 51,277,045 (~87%) high quality reads mapped to the genome. Coverage sequences, representing the number of reads mapped at each position of the genome (both strands), were determined using the samtools mpileup tool (7) The libraries were normalized relative to each other using a method previously described for the DESeq analysis package (8) with the modification that libraries were only reduced in size, and therefore $\hat{s}$ was 1 for the smallest library.

Scores for the *cis*-antisense and overlapping transcription models were determined by the function

$$S(c,d,a) = P(C > c, D > d, A > a)^{-1},$$

where P is the joint probability based on the background distributions of the two functions in C (the coverage term), D (the immunoprecipitated differential term), A (the immunoprecipitated ratio term). The coverage term is a function of the expression of the maximum covered strand in the IP library

$$F_C(R) = \log_2 \max_{s \in \{+,-\}} \beta_{IP_s}(R)$$

where the function $\beta_{P_s}(R)$ represents the average value of the coverage sequence over region R in library P on strand s.

In both models, we anticipate the ratio of per-strand coverage to be close to one, expressed in the function

$$F_A(R) = \left| \log_2 \frac{\beta_{IP_+}(R)}{\beta_{IP_-}(R)} \right|$$

The model-dependent term is a function of the increase in coverage intensity from the input to immunoprecipitated libraries. For Class I IP-dsRNAs, the notion of

an antisense strand is introduced, defined by the strand with the least coverage in the input library. Since we are measuring overall change in the antisense strand the term takes the form of

$$F_D(R \mid \text{Class I}) = \frac{\beta_{IP_{AS}}(R)}{\beta_{IN_{AS}}(R)}$$

where *AS* is the antisense strand and the beta function is as defined above.

For Class II IP-dsRNAs, the differential term is

$$F_D(R \mid \text{Class II}) = \log_2 \frac{\sum_{s \in \{+,-\}} \beta_{IP_s}(R)}{\sum_{s \in \{+,-\}} \beta_{IP_s}(R)}$$

in other words, the total log base 2 increase from input to immunoprecipitated library in both strands.

To calculate the background distribution, the three terms for each model were calculated on each of the regions of length 40 or more containing at least one read on both strands in both the input and IP libraries. Each distribution was computed for the wild-type and *rnc105* strains separately, for a total of four models. These were projected onto a three-dimensional matrix representing the cumulative distribution of values. The choice of joint probability allowed for the three characteristics to individually compensate for a weaker one and freed the model of weighting parameters.

IP-dsRNA candidates were extracted from the genome as any region for which one of the IP libraries had at least one read aligned on each strand continuously for at

least 38 bp. The threshold of 38 was chosen because of the requirement of 40 bp of the J2 monoclonal antibody and the fact that alignment errors on either end could cause the length to be reduced by a base. This resulted in 16,329 regions in the *rnc105* library. All possible subregions up to length 300 were scored with the above scheme given the background distribution, and all regions with a score of 10 or greater were reported as IP-dsRNAs. When overlapping IP-dsRNAs were found, the highest-scoring coordinates and model was reported. Since only optimal regions were reported, some dsRNA regions were fragmented due to sub-optimal intervals because of sequencing bias or sampling, so regions were clustered together which were within 100 bp of each other, reported as the highest scoring model.

The output of this algorithm resulted in optimal regions depending on the model. In order to extend these regions to the ends of the sequenced double stranded region and cluster nearby regions originating from the same double-stranded transcripts, we developed an heuristic algorithm based on the normalized coverage sequences determined above. First, if the end of the IP-dsRNA was covered by 20 or fewer reads, the region was trimmed until the coverage exceeded this threshold. In the IP library, a Gaussian filter with a window size of 50 (the median read length) was applied to the sum of both coverage strands G. For each region, a base level of enrichment e was determined by the mean coverage of the region. Each end was extended outward, and the current minimum value m was stored for the lowest value seen thus far in G. The current minimum was updated at each point of inflection in G. If the current value at any point of inflection was below e and above the m, or the coverage at any point dropped below 20 reads on both strands, the extension halted.

Then, the algorithm determined the end of the IP-dsRNA by determining the point of greatest change in coverage. To accomplish this, the first derivative of change

in coverage was computed for each strand and the sum of both strands. A Gaussian filter with a window size of 10 was applied to these sequences, to allow local accumulative changes to be detected. The end of the IP-dsRNA was determined by the highest value in the first-derivative sequence between the original start and the current position. For Class I IP-dsRNAs, this was determined in the antisense strand, as computed above. For Class II IP-dsRNAs, this was determined in the sum of both strands.

Finally, extended IP-dsRNAs were clustered together. If two IP-dsRNAs end coordinates were within 100 bp of each other, and at no point between the regions did the coverage drop below 20 reads, the regions were clustered together.

Gene annotations were downloaded from Genbank, accession number NC_000913 (9). Additionally, we considered the files included in the EcoCyc version 16.1 database (10), and in cases where a gene was not present or coordinates were not given in the Genbank annotation, the EcoCyc data were used.

The IP-dsRNA categorization was performed basically as shown in Figure 3*A*. If the IP-dsRNA overlapped an sRNA or a tRNA, it was always categorized as such. Next, IP-dsRNAs overlapping more than one feature were categorized as junction, convergent or divergent, depending on the orientation of the two transcripts. Following that, if the IP-dsRNA dominates a feature or overlaps more than 50% of it, the region is categorized as Full ORF, as only ORFs, sRNAs, tRNAs, tmRNA and rRNAs are considered. The remaining categories, 3’ ORF, 5’ ORF, 3’/Intergenic, 5’/Intergenic, and Convergent and Divergent IP-dsRNAs overlapping only one transcript, as illustrated in Figure 3*A*, are mutually exclusive and are therefore categorized without priority.

Northern blots

For Northern blot analysis 10 μg of DNase I (Roche) treated RNA was separated under denaturing conditions either by a 6-8% TBE-Urea (8M) polyacrylamide gels in 1X TBE (for small transcripts) or a 1% formaldehyde/MOPS agarose gels in 1X MOPS (for larger transcripts). RNA was initially denatured in 2X RNA load dye (Fermentas) and heated to 65°C for 15 min before loading on a gel. RNA was transferred to HybondXL membranes either by electroblotting at 12V for 1 h in 0.5X TBE (polyacrylamide TBE-Urea gels) or capillary action (formaldehyde-agarose gels). The membranes were UV cross-linked and probed with DNA oligonucleotide probes (Table S2). DNA oligonucleotide probes were end-labeled with γ32P [ATP] and T4 PNK (New England Biolabs) per the manufacturer's protocol.

Data access

The high-throughput read data is currently being deposited to SRA at NCBI. A complete list of IP-dsRNAs is provided in Supplementary Dataset 1. A genome browser containing the described genome annotation and visualized coverage maps can be found at http://alu.abc.univie.ac.at/fgb2_public/gbrowse/dsrna/ . All scripts used can be acquired from the authors upon request.

References :

1. Jahn CE, Charkowski AO, & Willis DK (2008) Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of microbiological methods* 75(2):318-324.
2. Han JS, Szak ST, & Boeke JD (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 429(6989):268-274.
3. Pfaffl MHM (2001) Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnology Letters* 23:275-282.
4. Schürer H, Lang K, Schuster J, & Mörl M (2002) A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic acids research* 30(12).

5. Benson D*, et al.* (2013) GenBank. *Nucleic acids research* 41(Database issue):42.
6. Langmead B & Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4):357-359.
7. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25(16):2078-2079.
8. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome biology* 11(10).
9. Riley M*, et al.* (2006) Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic acids research* 34(1):1-9.
10. Keseler I*, et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research* 41(Database issue):12.