Molecular Cell, Volume *51*

Supplemental Information

Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA

Ivan Olovnikov, Ken Chan, Ravi Sachidanandam, Dianne K. Newman, and Alexei A. Aravin

Inventory of Supplemental Information

Figure S1

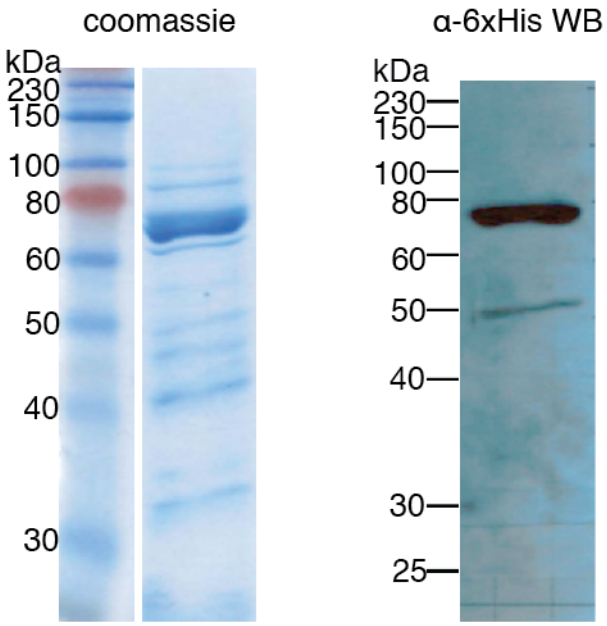

Figure S1. Isolation of 6xHis-RsAgo from *R. sphaeroides* strain 25, Related to Figure 1. Coomassie staining and western blot analysis of purified RsAgo using anti-His tag antibody. The apparent molecular weight of the protein is ~75 kDa (predicted 88 kDa).

Figure S2

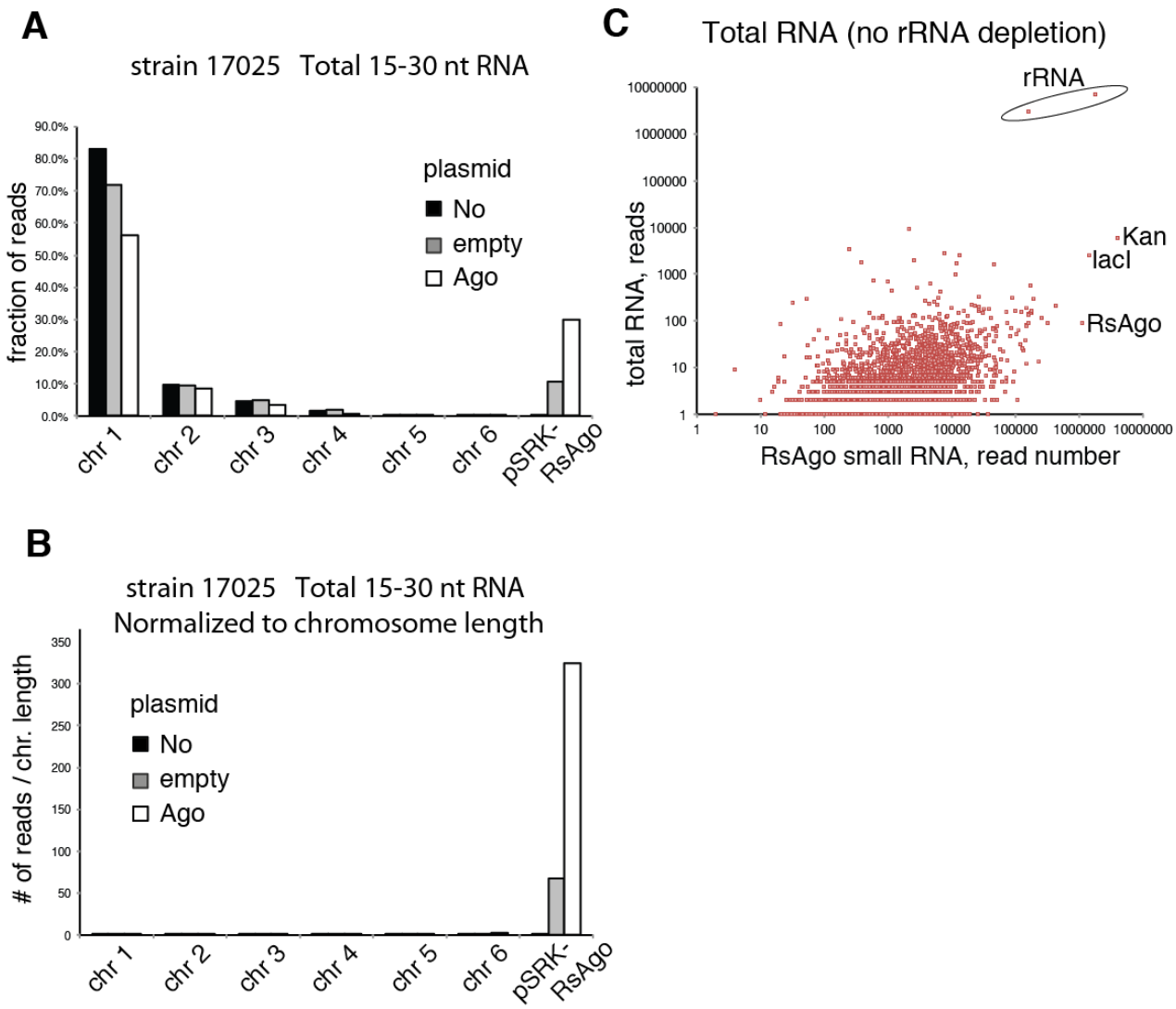

Figure S2. Distribution of total small RNA over *R. sphaeroides* chromosomes, Related to Figure 3. Small RNAs were cloned from total RNA of 13-30 nt size range (reads shorter than 15 nt were discarded from the analysis) from strain 25 without the expression plasmid and with a vector-only control or an RsAgo-containing plasmid. Small RNAs that mapped to chromosomes or plasmid are plotted as a fraction of raw read number **(A)** and read numbers normalized to chromosome length **(B)**. Sequences mapping to the RsAgo ORF are annotated as plasmid-specific. **(D)** Correlation between the amount of long and small RNA mapping to R. sphaeroides genes. The long RNA-Seq library was prepared from the same starting material as the library shown in Fig. 3D (strain 25 with pSRKKm-RsAgo plasmid), however rRNA depletion was omitted. Each dot represents a gene encoded in the genome or on the expression plasmid. Note the drastic depletion of rRNA genes in the small RNA population.

Figure S3

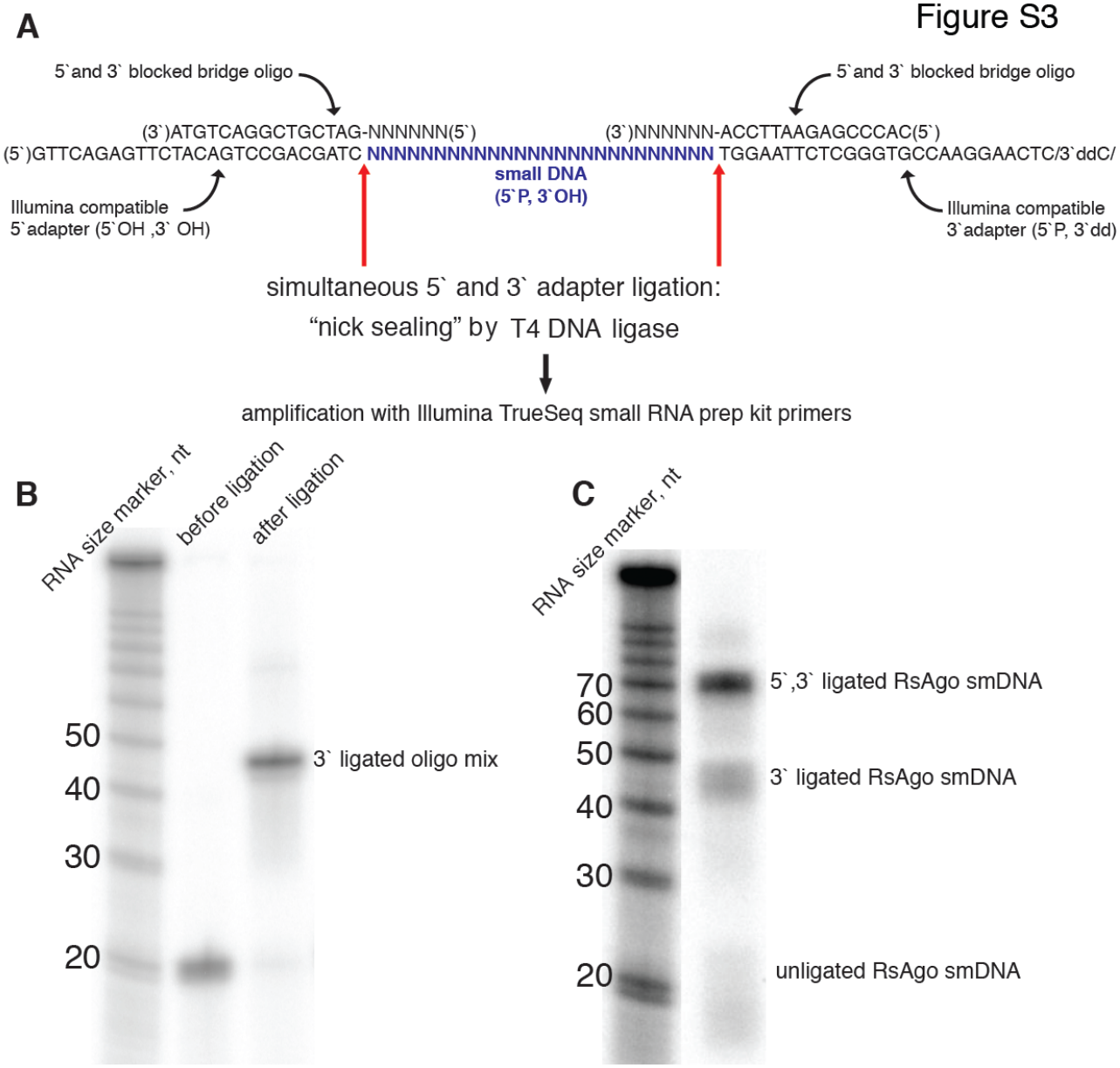

Figure S3. Cloning of small DNA library, Related to Figure 4 (A) A strategy for quick and efficient cloning of single-stranded small DNA for Illumina deep sequencing. The protocol requires the presence of 5'P and 3'OH residues in the small DNA. Ligations of 5' and 3' adapters were done using T4 DNA ligase: nicks to be sealed were created by "bridge" adapters that link the small DNA and adapters. Bridges anneal on any sequence because they contain six randomized nucleotide residues on their ends (5' or 3' depending on the bridge). All bridges were blocked on both ends to prevent their self-ligation. The 5' linker contains a 5'OH group so that it can be ligated only at its 3' end. The 3' linker contains a 5' phosphate and a 3'ddC to prevent 3' ligation. **(B)** 3' ligation of 5'-[32P]-labeled mixture of 20 different 20 nt DNA oligonucleotides using the method described in panel A. The reaction was incubated for 1 hour at room temperature under standard conditions (see Materials and methods) and resolved on a 15% PAGE. Only a small fraction of DNA oligonucleotides remains unligated. **(C)** Simultaneous 3' and 5' ligation of 5'-[32P]-labeled small DNA isolated from the RsAgo complex using the strategy shown in panel A. Most of the small DNA is ligated on both ends giving rise to a ~77 nt band (runs like the 70nt band of the single strand RNA ladder). A small fraction of DNA has the adapter ligated at the 3' end exclusively. Re-ligation of this fraction to 5' adapter generates the same ratio of ligated and non-ligated bands (not shown).

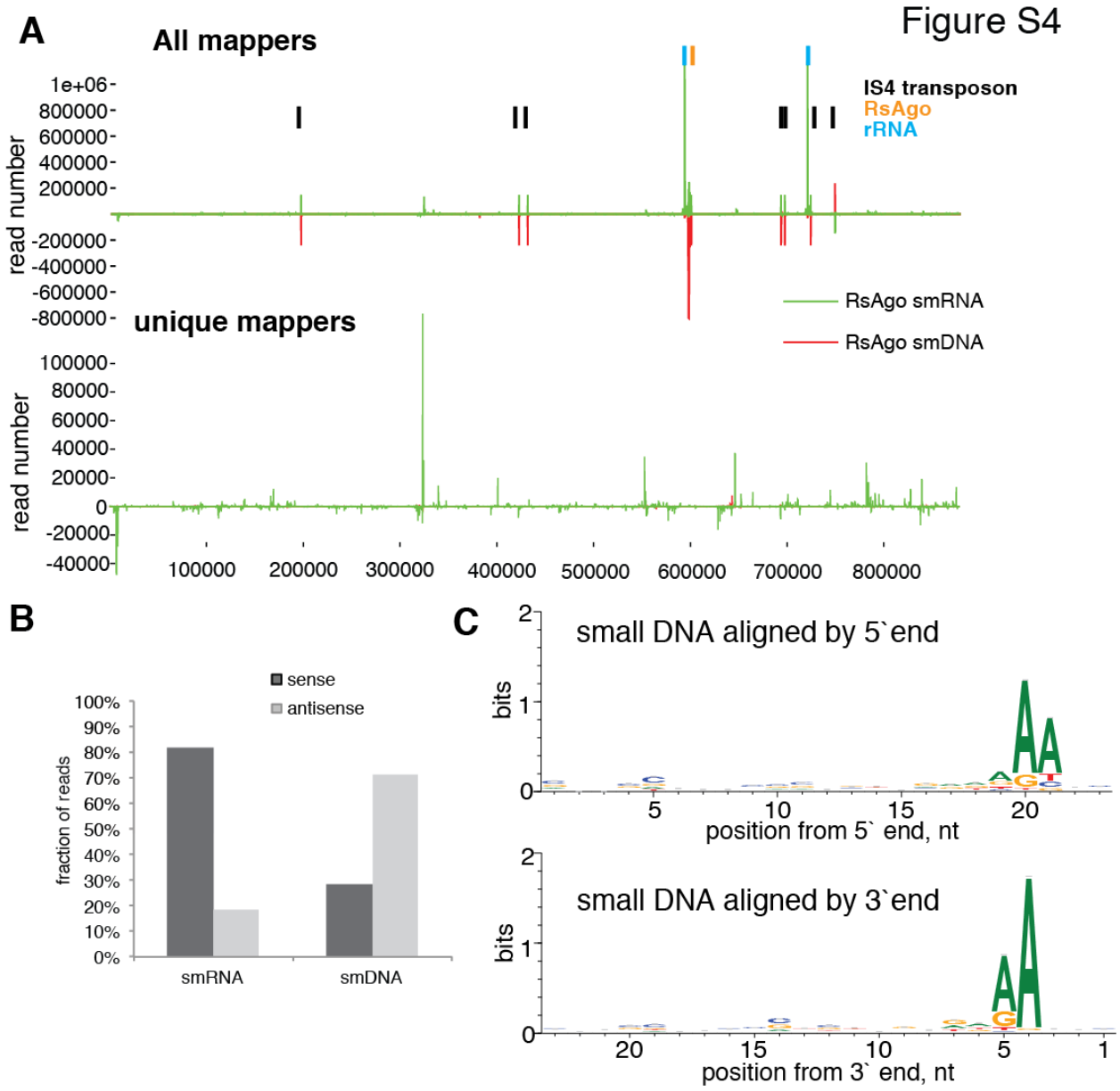

Figure S4. Origin and strand-bias of RsAgo-associated small RNA and small DNA, Related to Figure 4. (A) Mapping of RsAgo-associated small RNAs and small DNAs on chromosome 2 of *R. sphaeroides* strain 25. All sequences mapping to the chromosome are shown in the top graph. Uniquely- mapped sequences are shown in the bottom graph. Note that the high number of reads mapping to the RsAgo gene is likely derived from the expression plasmid. **(B)** Strand bias of RsAgo-associated small RNA and DNA mapped to all genes of *R. sphaeroides* strain 25. **(D)** Nucleotide bias in RsAgo-associated small DNA. 23-24 nt small DNA reads were aligned either by their 5'end (top panel) or by their 3' end (bottom, identical to Fig. 4F) and analyzed with WebLogo. Reads aligned by the 3' ends show a stronger bias for an adenine residue in position 20 and, unlike reads aligned by the 5' end, also exhibit purine enrichment in position 19, which matches the pyrimidine enrichment in position 2 in the small RNA (Fig. 2B). This result suggests that the 3' end of small DNA is fixed relative to the 5' end of small RNA and the distance between them equals 3 nucleotides (Fig. 4G). While the 3' overhang is almost invariably 3 nt long, the length of the 5' DNA overhang is slightly less precise as evidenced by higher correlation of the relative positions of the RNA 5' end and DNA 3' end compared to the RNA 3' end and DNA 5'end (Fig.4E).

Figure S5

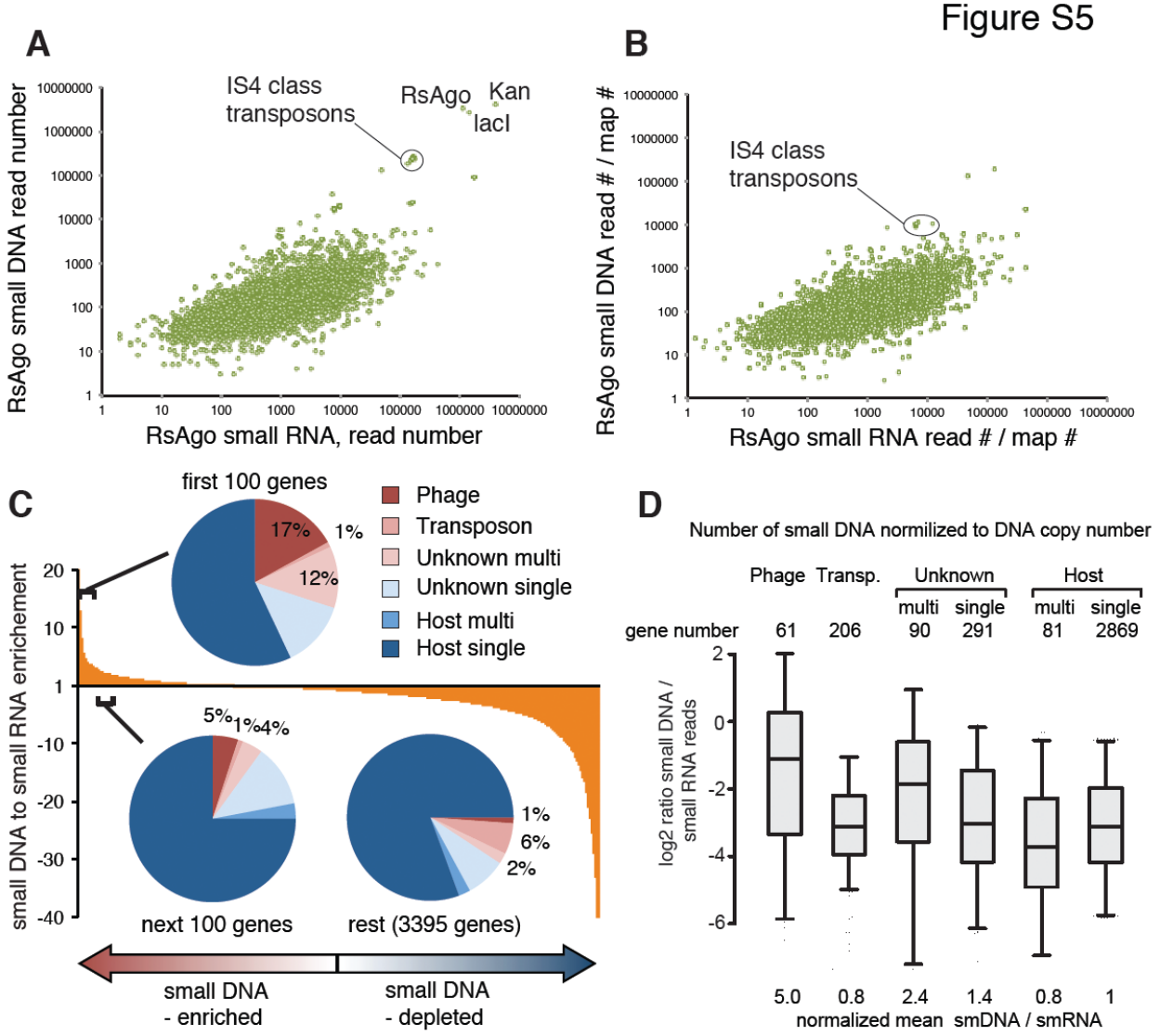

Figure S5. Correlation between the amount of small RNA and small DNA mapping to *R. sphaeroides* genes, Related to Figure 5. Shown are scatter plots of read numbers per gene **(A)** and of read numbers normalized to the number of genome mappings **(B)**. Each dot represents a gene encoded in the genome or on the expression plasmid. Genes from the expression plasmid (RsAgo, lacI and Kan) are eliminated from panel B as the precise copy number of the expression plasmid is unknown. **(C)** The distribution of the ratio of riDNA to diRNA for *R. sphaeroides* genes. Genes are sorted by the ratio of riDNA to diRNA reads. Small DNA, but not small RNA reads are normalized to copy number in the genome. Shown are fold enrichment (positive values) or depletion (negative values) of riDNA to diRNA normalized to the mean riDNA/diRNA ratio for all genes. A similar analysis without normalization to DNA copy number is shown in Fig. 4H. The frequencies of six different gene classes (single and multi-copy host genes, single- and multi-copy genes of unknown origin and phage- and transposon-related genes) were analyzed among the 100 most DNA-rich genes, the next 100 genes and the rest (3395 genes). **(D)** The ratio of riDNA to diRNA for different gene classes. Shown are box plots of riDNA to diRNA ratios for the same gene classes as on panel A. Number of genes in each class is shown above the plot. The box represents the 25th, 50th (the inner line) and the 75th percentiles of the distribution; whiskers are at the 5th and 95th percentile. The mean of the ratio of riDNA to diRNA was calculated for each gene class and normalized by that of host single-copy genes. DNA read numbers divided by the number of genome mappings (copy number) were used.

Figure S6

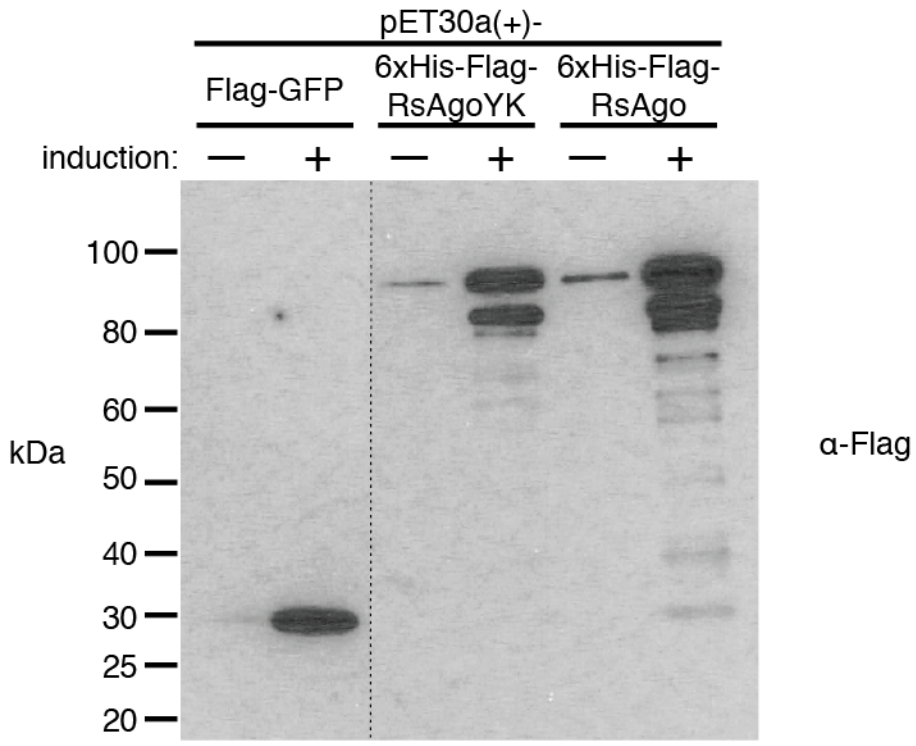

Figure S6. RsAgo expression in *E. coli* BL21(DE3), Related to Figure 6. Equal number of induced and not induced cells from experiment shown on Fig. 6C, D were lysed and loaded on 4-12% SDS gel followed by western blotting with anti-Flag antibody. The proteins are encoded on a high-copy plasmid pET30a(+) and have N-terminal Flag tag (GFP) or 6xHis-Flag tag (RsAgo and RsAgo-YK).

Figure S7

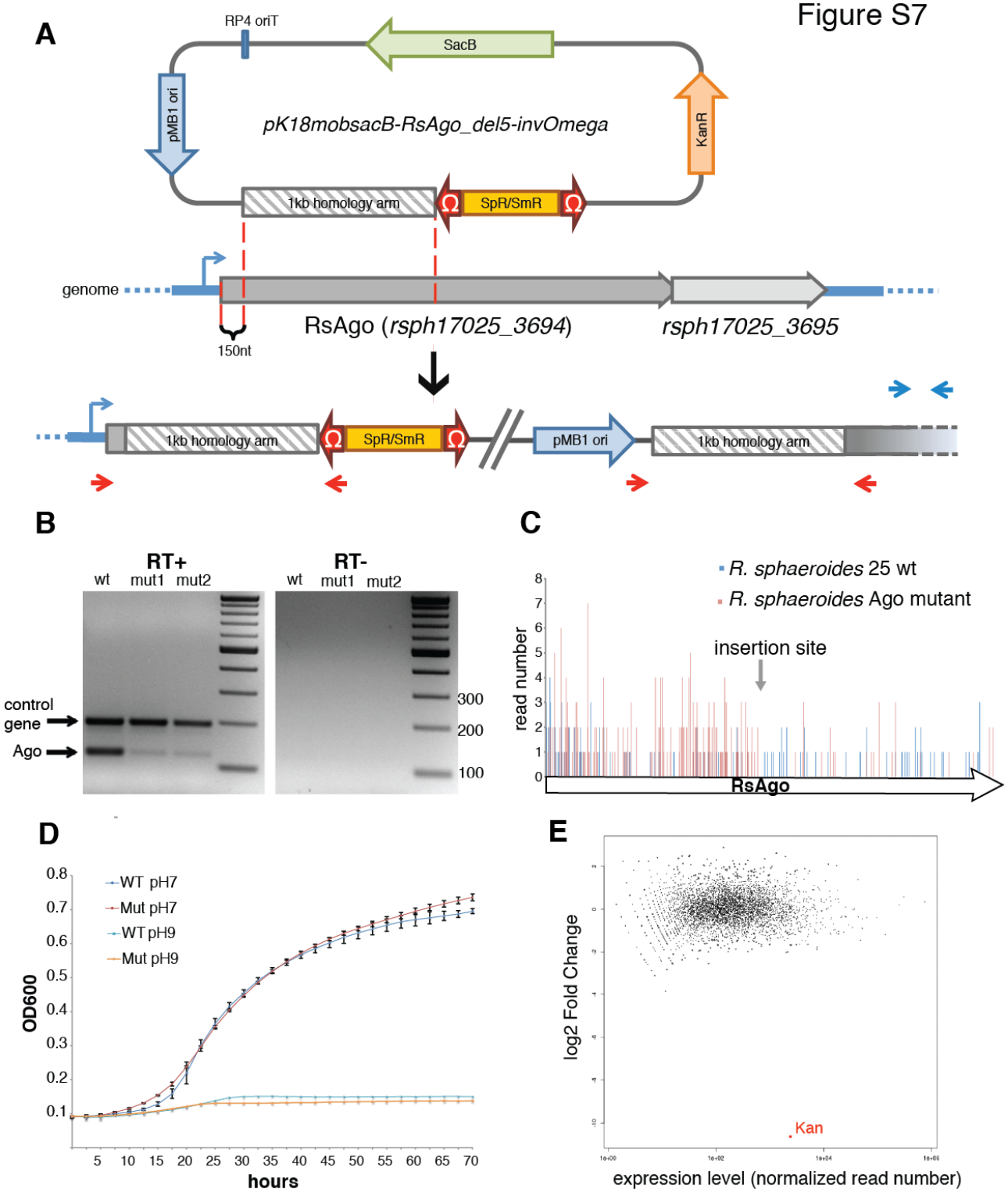

Figure S7. Construction and characterization of RsAgo mutant in *R. sphaeroides* strain 25, Related to Figure 7. (A) Scheme of the RsAgo mutagenesis construct. A suicide vector that cannot replicate in *R. sphaeroides* was integrated in the RsAgo gene using a 1 kb homology arm and successful recombinants were selected on Kanamycin. Correct integration was verified by primers that flank integration site (shown by red arrows) and by RT-PCR with primers, which reside in the region that follows the integration site (shown by blue arrows). **(B)** Duplex RT-PCR on RNA

extracted from *R. sphaeroides* parental wild type strain and two RsAgo mutant strains. The 200 bp product is amplified from a control gene and the 120 bp band from the piwi domain of RsAgo. **(C)** Deep sequencing of rRNA-depleted total RNA from R. sphaeroides 25 wild type and RsAgo mutant cells shows absence of RNA downstream of the integration site in the RsAgo mutant cells. **(D)** Growth dynamics (OD 600) of wild-type and RsAgo mutant cells in Minimal Sistrom`s Medium A with pH7 and 9. For each condition, four wells containing 250µl of medium were inoculated with equal number of cells and incubated with constant shaking at 30C for 72 hours. Measurement was performed every 2.5 hours. Shown are mean OD 600 values +/- SD. Similarly, these cultures showed only minimal differences in other media (not shown). **(E)** Transcriptome profiles of wild-type and RsAgo mutant cells. For each gene the normalized mean read number (RPM) is plotted on the X-axis and the fold-change in expression between the wild-type and the mutant is plotted on the Y-axis. The only gene that shows statistically significant change in expression in the mutant as determined by DESeq is Kanamycin, which is absent from the wild-type genome (marked in red).

Table S1. Small RNA and small DNA sequencing results, Related to Figure 2.

Library	6xHis-RsAgo smRNA in strain 25	6xHis-RsAgo smDNA in strain 25	Total small RNA (13-30nt) strain 25	Total small RNA (13-30nt) strain 25 with pSRKKm
Total sequences	1,732,773	1,126,436	3,136,540	2,704,899
Total reads	34,795,525	16,235,695	13,827,394	13,460,959
# of sequences mapped uniquely	534,775	559,699	1,148,140	974,316
# of reads mapped uniquely	24,585,582	12,367,355	6,339,034	5,505,602
# of sequences mapped not uniquely	42,716	45,524	173,449	154,130
# of reads mapped not uniquely	2,863,945	487,272	2,066,227	2,041,118
# of not mapped sequences	1,155,282	521,213	1,814,951	1,576,453
# of not mapped reads	7,345,998	3,381,068	5,422,133	5,914,239

Library	Total small RNA (13-30nt) strain 25 with pSRKKm-Ago	Total small RNA (13-30nt) strain 29 with pSRKKm	Total small RNA (13-30nt) strain 29 with pSRKKm-Ago	6xHis-RsAgo smRNA in *E. coli* BL21(DE3)	6xHis-RsAgo smDNA in *E. coli* BL21(DE3)
Total sequences	2,582,175	1,914,359	3,177,094	829,248	367,606
Total reads	15,989,596	14,651,848	16,244,565	20,370,560	6,914,163
# of sequences mapped uniquely	769,972	852,064	1,294,048	209,447	19,633
# of reads mapped uniquely	9,496,893	3,299,535	8,440,901	15,660,008	859,500
# of sequences mapped not uniquely	113,561	144,456	169,963	32,867	3,772
# of reads mapped not uniquely	1,296,442	7,773,480	3,111,081	2,429,801	135,968
# of not mapped sequences	1,698,642	917,839	1,713,083	586,934	344,201
# of not mapped reads	5,196,261	3,578,833	4,692,583	2,280,751	5,918,695

Table S2. Oligonucleotide sequences, Related to Experimental Procedures.

Primer	Sequence
1	TGACTCATATGATTCATCACCATCACCATCACGCCCCAGTGCAGGCTGC
2	TGACGGTACCTCATAGGAACCAGCGGCTCC
3	CGATCAGGATCCATCGAAAGTGAAGGAAGAGCG
4	AAAGCTTGCTCAATCAATCACCATTCTCCACTTTTCCTTGAGTG
5	GGTGATTGATTGAGCAAGCTTT
6	CTTCATCTGCAGGGTGATTGATTGAGCAAGCTTT
7	ACCAGGTCGAAGTGATTGTTC
8	CAAGCATAAAGCTTGCTCAATC
9	CAGGAAACAGCTATGAC
10	CGAGTAGTTCGAACCCATCC
11	GAAAACGATGCTGGCTACGT
12	GTCCATGACTGGCATTTTGC
13	CAGTTCCGCAAGATCTATGC
14	GTAGGAACCGATGTTCACG
15	TATTTCCATATGGCCGATGCTAAGAACATTAAG
16	ATCTATGCTAGCTTAGACGTTGATCCTGGCGC
17	CGCCTTTCTTAGCCTTGATC
18	AGGAGATCGTGGACTATGTG
19	ATGGCACCACGCTCAGAATA
20	TGATCATGAACAGCTCTGGG
21	GCTCCTTCTCCACCAGATGATA
22	GCAGCTGGTCAACTAAGTAG
23	ACCAAGAAGGTGAAGACTGC
24	AGGAATTCGTACATGCGGTC
25	GGTGGACAATGTGATGATGC
26	CCGATGGCGATGAAGATGAT
27	GACTGCCACTTTTACGCAAC
28	ATGCCGATTTCTCTGGACTG
29	ACGGCGGGATATAACATGAG
30	TGGTTGCCAACGATCAGATG
31	TTCGCGCACCATCTCCTATT
32	GCTTGATCGCCACATATTGC
33	AGCTGACCGAGACCAATTAC
34	TCCAGTACTTGTCGGTGAAG
35	CATTTTTCCGTGGAAGATGGGC
36	CAAGTCGCGCATTCTGCATT
37	GGTTTTCACATTCCGCCGAT
38	CATGGCTTCGCTTTCTCTCT
39	CTAGCATGCATATGCATCACCATCACCATCACGATTACAAGGATGACGATGAC
40	CTAGCATGCATATGCATCACCATCACCATCAC

41	CAGGGACCCGGTATGGATACCTGGGTTTC
42	GAAACCCAGGTATCCATACCGGGTCCCTG
43	GAAAACGATGCTGGCTACGT
44	CTAGCATGCATATGGATTACAAGGATGACGATGACAAGGTGAGCAAGGGCGAGGAG
45	GTTGACGGTACCTTACTTGTACAGCTCGTCCATG

Experimental Procedures

Bacterial strains

Rhodobacter sphaeroides strains ATCC17025 and ATCC17029 were kindly provided by Timothy Donohue (University of Wisconsin–Madison). Cells were grown on Sistrom's minimal medium A at 30C under aerobic conditions. Kanamycin was used at concentration 25 µg/ml for R. sphaeroides and 25 µg/ml for E. coli BL21(DE3), Tetracyclin at 1 µg/ml for R. sphaeroides.

RsAgo expression and purification in R. sphaeroides

N-6xHis-tagged ORF of RsAgo (*rsph17025_3694*) was amplified from genomic DNA of strain ATCC17025 (primers 1 and 2, see supplementary table 2 for all primer sequences) and cloned between the NdeI and KpnI sites of the pSRKKm broad-host-range expression vector (Khan et al., 2008). This plasmid was mobilized into *R. sphaeroides* by biparental mating with *E. coli* BW29427. Counter selection of donor cells was achieved by omitting diaminopimelic acid (DAP) from the medium, which is required for growth of this strain. Protein was isolated by 6xHis-tag using Talon beads (Clontech) from 1L of culture induced by 1mM IPTG at OD 1-2 for 5 to 10 hours under aerobic conditions. After induction cells were pelleted, washed with ice-cold PBS and frozen, then resuspended in 5ml of buffer A (50mM phosphate buffer, 300 mM NaCl, 5mM imidazole, pH 7.4) per 1 gram of cell pellet and disrupted on a French press at 20000 psi (two passes), clarified by centrifugation for 20min at 30,000 g and applied to the column. Resin was washed with 15 volumes of buffer A followed by 5 volumes of 50 mM Tris-HCl, 300 mM NaCl, pH 7.4. The protein was eluted with 50mM Tris-HCl, 300 mM NaCl, 300mM imidazole, pH 7.4.

Generation of RsAgo mutant strain ATCC17025

To create the mutagenesis vector a 1 kb homology arm corresponding to the genomic sequence pRSPA01:596308-597307 was amplified (using primers 3 and 4) and used in overlapping PCR (primers 3 and 6) with the omega transcription termination cassette amplified from vector pHP45-omega (Prentki and Krisch, 1984). The final PCR product was digested with BamHI and PstI and cloned into the suicide vector pK18*mobsacB* (Schafer et al., 1994). The mutagenesis vector was mobilized into *R. sphaeroides* and recombinants were selected on Kanamycin. Successful integration was verified with primer pairs 7-8 and 9-10 as shown on Fig. S7. To confirm disruption of RsAgo expression, RNA was extracted from wild type and mutant cells and RT-PCR was performed to amplify the Piwi domain of RsAgo (primers 11-12) and the control gene rpsD (control primers 13-14) as shown on Fig.S11B. RT-qPCR showed that expression of the N-terminally

truncated version of RsAgo in mutant cells is ~29 times lower than expression of full-length RsAgo in wt cells.

Plasmid expression in wild-type and RsAgo mutant strains

For expression of firefly luciferase (*Fluc*), CDS was amplified with primers 15-16 using vector pGL3(R2.1) (Promega) as a template, digested with NdeI and NheI and cloned into the corresponding sites of the shuttle vector pSRKTc (Khan et al., 2008). For firefly luciferase measurement wild-type or RsAgo mutant cells containing pSRKTc-Fluc were induced with 1mM IPTG for ~12 hours, equal number of cells (~1ml of culture at OD600=0.3) were pelleted and resuspended in passive lysis buffer (Promega) supplemented with 3 mg/ml of lysozyme (Sigma) and incubated for 15 min at room temperature. 40µl of lysate was mixed with equal volume of firefly luciferase substrate (Promega) and immediately measured for 10 seconds in a luminometer. Cells not induced with IPTG served as a control. Each measurement was performed three times for three independent experiments. For firefly luciferase and *lacI* mRNA RT-qPCR, RNA was extracted from the same cultures that were used in the luciferase assay and qPCR was performed using primers 17-18, 19-20 (*Fluc*), 21-22 (*lacI*) and primers to control genes 13-14 (*rpsD*), 23-24 (*rplE*), 25-26 (*rsph17025_2888*). For plasmid copy number comparison qPCR was performed using equal number of cells (same cultures that were used in the luciferase assay) using primers specific to pSRKTc-Fluc (17-18,19-20, 27-28, 29-30) and primers specific to genomic DNA (31-32, 33-34, 35-36, 37-38).

RsAgo expression in E. coli

For expression of RsAgo in *E. coli* vector pSRKKm-RsAgo was used. To achieve high level of RsAgo expression in *E. coli*, 6xHis-Flag-tagged RsAgo was cloned into vector pET30a(+) (Novagen) using PCR with primers 39-2 and vector pSRKKm-RsAgo as a template. The product was digested with NdeI, KpnI and cloned into vector pET30a(+). To introduce mutations Y463G and K467G in the RNA 5` end binding pocket of the RsAgo MID domain, PCR was performed using primers 40-41 and 42-43 and pET30a(+)-6xHis-Flag-RsAgo as a template, followed by overlapping PCR with primers 40-43. The final PCR product was digested with NdeI and SpeI and cloned into pET30a(+)-RsAgo linearized with NdeI and SpeI. eGFP CDS was amplified with primers 44-45 using pEGFP (Clontech) vector as template, PCR product was digested with NdeI and KpnI and ligated into pET-30a(+) linearized with NdeI and KpnI. For experiments shown on Fig. 6 C,D and Fig. S6, cells were grown in LB with kanamycin (50 µg/ml) overnight until stationary phase (OD(600) ~3.5). 0.5 ml of each culture was added to 5 ml of LB with kanamycin

supplemented with 1mM IPTG and incubated for 5 hours at 37°C. Expression of human Argonaute 1 in the same conditions did not cause any visible plasmid degradation. Plasmids were isolated using Plasmid Miniprep Kit (Zymo Research).

Small isolation and sequencing

Small RNA and DNA species were extracted from purified RsAgo complex using using proteinase K treatment followed by neutral phenol:chloroform extraction. Samples were dephosporylated using calf intestine phosphatase, phosphorylated in the presence of [γ-^{32}P]ATP by T4 polynucleotide kinase and differentially treated with DNase I or RNase A, respectively. Cloning of small RNA was done according to published protocols (Brennecke et al., 2007; Lau et al., 2001) using linkers and primers from the Illumina TrueSeq Small RNA Sample Prep kit. Efficiency of 3` ligations was very similar to that seen for *Drosophila* miRNA and piRNA, suggesting that majority of RsAgo-associated small RNA have 3` hydroxyl group. However cloning protocol does not allow us to make conclusions about the nature of 5` end. Small RNA libraries were barcoded, pooled and sequenced on the Illumina HiSeq2000 platform with a read length of 50 nt to an average depth of 15 million reads per library. For total small RNA cloning (13-30 nt range) RNA was isolated using the Amresco Phenol-Free Total RNA Purification Kit after fixation of the cell culture with equal volume of Ambion RNAlater reagent.

Long RNA sequencing and analysis

For regular RNA sequencing samples were processed according to the Illumina TrueSeq RNA prep kit. rRNA depletion was performed using the RiboZERO gram-negative bacterial rRNA depletion kit (EpiBio). To profile the transcriptomes of wild-type and RsAgo mutant strain 25, duplicate RNA-Seq libraries were prepared from rRNA-depleted RNA isolated from two independent experiments. The differential expression was analyzed using DESeq (Anders and Huber, 2010) via R statistical environment. Dispersion was modeled using the estimateDispersion function of DESeq library with fitType parameter set as ‘parametric’, method as ‘blind’ and sharing mode as ‘fit-only’. rRNA genes and genes that had zero read counts in any of the libraries were excluded from the analysis. Genes were considered to be differentially expressed if the multiple testing adjusted p-value was below 0.2.

Small DNA library preparation

Our approach to cloning the small DNA library and oligonucleotide sequences is shown in Fig. S3. To block bridge oligonucleotides (to prevent interference with ligation of small DNA) we used the 5' amino modifier C6 and the 3' amino modifier from Integrated DNA Technologies (IDT). The 3' linker was 5' phosphorylated and blocked on the 3' end by dideoxy cytidine (IDT). The 5' linker contained a hydroxyl group on both 5' and 3' ends. A simultaneous 5' and 3' linker ligation reaction was performed in 15 µl volume and contained varying amounts of small DNA, 100 pmoles of each linker and bridge oligonulceotides, 5% PEG8000, 1x T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl2,1 mM ATP, 10 mM DTT, pH 7.5) and 1 µl of T4 DNA ligase (NEB, 400,000 units/ml). The reaction was incubated at room temperature for 1 to 10 hours without significant increase in efficiency (~90%) after one hour, as seen from shifts of 5`-$^{32}$P labeled small DNA. This indicates that the majority of RsAgo-associated small DNA have 3` hydroxyl group, however does not allow us to make conclusions about the nature of 5` end. We found that pre-annealing of bridge oligonucleotides to linkers did not increase efficiency of cloning. Also, simultaneous ligation of 3' and 5' adapters was as efficient as consecutive ligations. Further library amplification was performed essentially as described for the small RNA libraries with the exception of omitting the reverse transcription step. Adapters shown in Figure S3 are compatible with the Illumina TrueSeq small RNA prep kit; Illumina primers were used to create indexed libraries.

Small RNA and DNA sequence analysis

For analysis of small RNA and DNA sequencing data, low quality reads were removed, adapter sequence clipped, sequences shorter than 15 nt were discarded. Mapping to the *R. sphaeroides* genome and pSRKKm plasmid was done using Bowtie (Langmead et al., 2009); only perfect matches were considered for further analysis. Annotations of the *R. sphaeroides* genome were taken from microbiological sequencing data repository http://www.microbesonline.org/. To separate genes into six classes (single and multi-copy host genes, single- and multi-copy genes of unknown origin and phage- and transposon-related genes) we analyzed all available gene annotations: gene description, COG, TIGR, GO and EC. A gene was classified as phage- or transposon-related if words 'phage' and 'transposase' were present in the annotation, respectively. Genes without clear annotation but with homology to genes classified in the first step were annotated accordingly. A few genes were classified as 'phage' even though they did not have clear annotation because they were located between two other phage genes. Other genes were classified as 'gene of unknown origin' if no available annotations or compelling contextual clues were found. A gene was classified as multi-copy if more than 30% of the small RNA and small DNA reads that mapped to it could also be mapped to other positions in the genome.

For sequence analysis of small RNA and DNA we used the small RNA dashboard server (Olson et al., 2008) and Galaxy tools (Blankenberg et al., 2010).

Analysis of the distances between ends of small RNA and DNA

To analyze correlations between specific ends (5' or 3') of the reads we asked how likely it is that one specific end (5’ or 3’) of one molecule (small RNA or small DNA) is at a particular distance from another molecule. For each read, we constructed pairs from all reads that mapped within a 31 nt window (-15 to +15 relative to the end of the read). Each pair was defined by the distance between the ends (pair-distance) and the abundances (read-counts) of the two reads. The product of the read-counts in a pair is the contribution to the correlation measure at the pair-distance. To make this a probability, this measure was divided by the total number of reads within the window. The sum over all reads of the contributions for each pair-distance gives a correlation function that is defined from -15 to +15 (the width of the window). In order to have an overall normalization, the function was divided by the sum of all reads that map to the chromosome.

Supplementary References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol *Chapter 19*, Unit 19 10 11-21.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Khan, S.R., Gaines, J., Roop, R.M., 2nd, and Farrand, S.K. (2008). Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. Appl Environ Microbiol *74*, 5053-5062.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Olson, A.J., Brennecke, J., Aravin, A.A., Hannon, G.J., and Sachidanandam, R. (2008). Analysis of large-scale sequencing of small RNAs. Pac Symp Biocomput, 126-136.

Prentki, P., and Krisch, H.M. (1984). In vitro insertional mutagenesis with a selectable DNA fragment. Gene *29*, 303-313.

Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G., and Puhler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. Gene *145*, 69-73.